Fall 2010

Methodology to Quantify Vertical Accelerations of Planing Craft in Irregular Waves

Jennifer Suzanne Grimsley
Old Dominion University

Follow this and additional works at: https://digitalcommons.odu.edu/mae_etds
Part of the Ocean Engineering Commons

Recommended Citation
Grimsley, Jennifer S. "Methodology to Quantify Vertical Accelerations of Planing Craft in Irregular Waves" (2010). Doctor of Philosophy (PhD), dissertation, Mechanical Engineering, Old Dominion University, DOI: 10.25777/n2n4-nm53
https://digitalcommons.odu.edu/mae_etds/136

This Dissertation is brought to you for free and open access by the Mechanical & Aerospace Engineering at ODU Digital Commons. It has been accepted for inclusion in Mechanical & Aerospace Engineering Theses & Dissertations by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.
METHODOLOGY TO QUANTIFY VERTICAL ACCELERATIONS
OF PLANING CRAFT IN IRREGULAR WAVES

by

Jennifer Suzanne Grimsley
B.S. June 1998, Webb Institute of Naval Architecture
M.S. June 1999, Massachusetts Institute of Technology

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirement for the Degree of

DOCTOR OF PHILOSOPHY
MECHANICAL ENGINEERING
OLD DOMINION UNIVERSITY
2010

Approved by:

Ayodeji Demuren (Director)
Gene Hou (Member)
Chet Grosch (Member)
ABSTRACT

METHODOLOGY TO QUANTIFY VERTICAL ACCELERATIONS OF PLANING CRAFT IN IRREGULAR WAVES

Jennifer Suzanne Grimsley
Old Dominion University, 2010
Director: Dr. Deji Demuren

Planing craft operating in waves at high speeds can experience high, repetitive vertical accelerations that are random and nonlinear in relation to the sea condition. A proper understanding of vertical accelerations is critical to meet statistically based criteria for structural design, habitability, and equipment selection. Historically, it was assumed that planing craft vertical accelerations fit the Exponential distribution, and design methods adopted this conclusion. However, several published papers have raised doubts regarding the accuracy and validity of this Exponential distribution assumption.

The statistical behavior of planing craft vertical accelerations are examined for the Parent and Peak data sets from twenty-eight (28) full-scale and model-scale tests of different hulls operating in irregular waves. Comparisons are made with Exponential, Rayleigh, Gumbel, and Lognormal distributions. Sensitivity studies regarding Peak Identification methods and threshold values are considered. Methods to extend legacy data, including the use of the Monte Carlo simulation technique and correlations between statistical parameters of Parent data sets and Peak data sets are examined.

The results of this research prove that the Exponential distribution is not appropriate for Peak or Parent vertical accelerations. For modern planing craft, the best fit for both the Peak and Parent vertical accelerations is the Gumbel distribution. The Monte Carlo method proved to be accurate in simulating the experimental data using the Gumbel distribution and only limited knowledge of the statistics of the experimental data. A strong linear correlation was found between statistical parameters of the Parent and Peak data sets and relationships are provided as guidance to planing craft designers. Additional statistical values, including Probability of Exceedance values, are included.
This dissertation is dedicated to my parents, Mike and Bobbi Grimsley, who instilled in me the courage and determination to pursue my ambitions, and provided me with unwavering support to achieve them.
I would like to acknowledge Mr. Donald Blount, whose professional dedication and achievements in advancing the state-of-the-art in planing craft design is the inspiration for this research. I would like to express my sincere appreciation to Dr. Ayodeji Demuren, Dr. Gene Hou, and Dr. Chet Grosch for their continual guidance and counsel throughout both this research program and my academic career. I would like to give special recognition to Mr. Yu Liu and Mr. David Pogorzelski for their untiring support and for their valuable suggestions in pursuit of this research. Finally, I am grateful for my family and friends who provided a wealth of encouragement and levity through this long, rewarding journey.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1. Seakeeping of Planing Craft</td>
<td>1</td>
</tr>
<tr>
<td>1.2. Planing Craft Design Criteria</td>
<td>8</td>
</tr>
<tr>
<td>1.3. Empirical Methods for Vertical Accelerations</td>
<td>11</td>
</tr>
<tr>
<td>1.4. Assumed Probability Distribution</td>
<td>13</td>
</tr>
<tr>
<td>1.5. Shortcomings of Earlier Work</td>
<td>15</td>
</tr>
<tr>
<td>1.6. Scope of the Dissertation</td>
<td>21</td>
</tr>
<tr>
<td>2. Statistical Distribution Analysis</td>
<td>26</td>
</tr>
<tr>
<td>2.1. Probability Distribution Functions</td>
<td>27</td>
</tr>
<tr>
<td>2.1.1. Exponential Distribution</td>
<td>29</td>
</tr>
<tr>
<td>2.1.2. Rayleigh Distribution</td>
<td>29</td>
</tr>
<tr>
<td>2.1.3. Gumbel Distribution</td>
<td>30</td>
</tr>
<tr>
<td>2.1.4. Lognormal Distribution</td>
<td>31</td>
</tr>
<tr>
<td>2.2. Goodness of Fit Testing</td>
<td>34</td>
</tr>
<tr>
<td>2.3. Peak Acceleration Distribution</td>
<td>36</td>
</tr>
<tr>
<td>2.3.1. Peak Identification Methods</td>
<td>38</td>
</tr>
<tr>
<td>2.3.2. Threshold Values</td>
<td>41</td>
</tr>
<tr>
<td>2.3.3. Example of Peak Distribution Method</td>
<td>42</td>
</tr>
<tr>
<td>2.4. Parent Acceleration Distribution</td>
<td>49</td>
</tr>
<tr>
<td>3. Statistical Parameters for Design</td>
<td>52</td>
</tr>
<tr>
<td>3.1. Statistical Parameters of the Peak Data Sets</td>
<td>53</td>
</tr>
<tr>
<td>3.1.1. Direct Method</td>
<td>53</td>
</tr>
<tr>
<td>3.1.2. Analytical Method</td>
<td>55</td>
</tr>
<tr>
<td>3.1.3. Monte Carlo Simulation of the Peak Data</td>
<td>64</td>
</tr>
<tr>
<td>3.2. Statistical Parameters of the Parent Data Sets</td>
<td>67</td>
</tr>
<tr>
<td>3.2.1. Direct Method</td>
<td>68</td>
</tr>
<tr>
<td>3.2.2. Monte Carlo Simulation of the Parent Data</td>
<td>68</td>
</tr>
<tr>
<td>3.3. Correlation Between Parent and Peak Parameters</td>
<td>74</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1.</td>
<td>10</td>
</tr>
<tr>
<td>1.2.</td>
<td>15</td>
</tr>
<tr>
<td>1.3.</td>
<td>18</td>
</tr>
<tr>
<td>1.4.</td>
<td>23</td>
</tr>
<tr>
<td>2.1.</td>
<td>44</td>
</tr>
<tr>
<td>3.1.</td>
<td>55</td>
</tr>
<tr>
<td>3.2.</td>
<td>70</td>
</tr>
<tr>
<td>3.3.</td>
<td>71</td>
</tr>
<tr>
<td>3.4</td>
<td>77</td>
</tr>
<tr>
<td>3.5</td>
<td>78</td>
</tr>
<tr>
<td>3.6</td>
<td>81</td>
</tr>
<tr>
<td>3.7</td>
<td>83</td>
</tr>
<tr>
<td>4.1.</td>
<td>95</td>
</tr>
<tr>
<td>4.2.</td>
<td>95</td>
</tr>
<tr>
<td>4.3.</td>
<td>99</td>
</tr>
<tr>
<td>4.4.</td>
<td>100</td>
</tr>
<tr>
<td>4.5.</td>
<td>101</td>
</tr>
</tbody>
</table>

1.1. Crew Habitability Guidelines

1.2. Statistical Multipliers for Exponential Distribution

1.3. Prediction of 1/Nth Peak Acceleration Statistics
    Using Exponential Distribution

1.4. Range of Parameters for Experimental Data Sets Examined

2.1. Estimating Distribution Parameters using the
    Method of Moments

3.1. The Direct Method to Calculate Average of 1/Nth Statistics

3.2. Accuracy of Estimating Parent Data Set Using Monte
    Carlo Simulation

3.3. Comparison of Statistical Parameters of Parent Data Set

3.4. Test Matrix for High L/B Ratio Planing Craft

3.5. Existing Analytical and Published Empirical Relationships

3.6. First Four Moments of Parent Data Set for Four Test Cases

3.7. Comparison of Probability of Exceedance Levels for
    Three Test Cases

4.1. Error Using Monte Carlo Simulation of Full-Scale
    Peak Data Sets

4.2. Error Using Monte Carlo Simulation of Model-Scale
    Peak Data Sets

4.3. Error Using Monte Carlo Simulation of Full-Scale
    Peak Data Sets

4.4. Error Using Monte Carlo Simulation of Model-Scale
    Peak Data Sets

4.5. Comparison of Full-Scale Experimental and Gumbel
    Analytical Cut-off Values for Given PoE
4.6. Comparison of Full-Scale Experimental and Gumbel Analytical Cut-off Values for Given PoE ......................... 102
4.7. Empirically Derived Relationships Between Parent and Peak Statistics ................................................. 105
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Various Planing Craft Designs</td>
<td>2</td>
</tr>
<tr>
<td>1.2</td>
<td>Planing Craft Operating in Rough Seas</td>
<td>3</td>
</tr>
<tr>
<td>1.3</td>
<td>Craft Vertical Acceleration and Wave Height Spectra</td>
<td>4</td>
</tr>
<tr>
<td>1.4</td>
<td>Parent Data Set, Time History of Vertical Acceleration Data</td>
<td>5</td>
</tr>
<tr>
<td>1.5</td>
<td>Peak Data Set, Selection of Peak Values from Parent Data Set</td>
<td>6</td>
</tr>
<tr>
<td>1.6</td>
<td>Average of 1/Nth Highest Statistics of Peak Data Set</td>
<td>7</td>
</tr>
<tr>
<td>1.7</td>
<td>Statistical Parameters and Distribution of Wave Heights</td>
<td>8</td>
</tr>
<tr>
<td>1.8</td>
<td>Peak Data Set, Sensitivity of Peak Identification and Threshold</td>
<td>20</td>
</tr>
<tr>
<td>2.1</td>
<td>PDF of Four Distributions with Equal Mean Value</td>
<td>33</td>
</tr>
<tr>
<td>2.2</td>
<td>CDF of Four Distributions with Equal Mean Value</td>
<td>33</td>
</tr>
<tr>
<td>2.3</td>
<td>PDF of Four Distributions with Equal Std Deviation Value</td>
<td>33</td>
</tr>
<tr>
<td>2.4</td>
<td>CDF of Four Distributions with Equal Std Deviation Value</td>
<td>34</td>
</tr>
<tr>
<td>2.5</td>
<td>Illustration of the K-S Test for Goodness of Fit</td>
<td>35</td>
</tr>
<tr>
<td>2.6</td>
<td>Peak Identification Methods</td>
<td>39</td>
</tr>
<tr>
<td>2.7</td>
<td>Visual Comparison of PDF of Peak Data Set</td>
<td>44</td>
</tr>
<tr>
<td>2.8</td>
<td>Visual Comparison of CDF of Peak Data Set</td>
<td>45</td>
</tr>
<tr>
<td>2.9</td>
<td>Peak Distribution Goodness of Fit (Buffer)</td>
<td>46</td>
</tr>
<tr>
<td>2.10</td>
<td>Peak Distribution Goodness of Fit (Vertical Threshold)</td>
<td>46</td>
</tr>
<tr>
<td>2.11</td>
<td>Goodness of Fit for Peak Distribution (Vertical Difference)</td>
<td>47</td>
</tr>
<tr>
<td>2.12</td>
<td>Goodness of Fit for Peak Distribution (Horizontal Threshold)</td>
<td>47</td>
</tr>
<tr>
<td>4.5</td>
<td>Goodness of Fit for Full-Scale Peak Data Sets Using Buffer Method</td>
<td>87</td>
</tr>
<tr>
<td>4.2</td>
<td>Trendline Results for Full-Scale Peak Data Sets Using Buffer Method</td>
<td>88</td>
</tr>
<tr>
<td>4.3</td>
<td>Goodness of Fit for Full-Scale Peak Data Sets Using HT Method</td>
<td>88</td>
</tr>
<tr>
<td>4.4</td>
<td>Trendline Results for Full-Scale Peak Value Data Sets Using HT Method</td>
<td>89</td>
</tr>
</tbody>
</table>
4.5. Goodness of Fit for Model-Scale Peak Value Data Sets Using Buffer Method

4.6. Goodness of Fit for Model-Scale Peak Value Data Sets Using HT Method

4.7. Trendline Results for Model-Scale Peak Value Data Sets, Buffer Method

4.8. Trendline for Model-Scale Peak Value Data Sets Using HT Method

4.9. Error in Mean Value for Monte Carlo Simulation of Full-Scale Peak Data

4.10. Error in Std Dev Value, Monte Carlo Simulation of Full-Scale Peak Data

4.11. Error in Avg 1/10th Value, Monte Carlo Simulation Full-Scale Peak Data

4.12. Error, Avg 1/100th Value, Monte Carlo Simulation of Full-Scale Peak Data

4.13. Goodness of Fit Results for Full-Scale Parent Data Sets

4.14. Trendline Results for Full-Scale Parent Data Sets

4.15. Goodness of Fit Results for Model-Scale Parent Data Sets

4.16. Trendline Results for Model-Scale Parent Data Sets

4.17. Correlation, Average of Peak Data Set and RMS of Parent Data Set

4.18. Correlation Average of 1/3rd Highest of Peak Data Set and RMS of Parent Data Set

4.19. Correlation Average of 1/10th Highest of Peak Data Set and RMS of Parent Data Set

4.20. Correlation Average of 1/100th Highest of Peak Data Set and RMS of Parent Data Set
NOTATION

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABS</td>
<td>American Bureau of Shipping</td>
</tr>
<tr>
<td>B</td>
<td>Craft beam, feet</td>
</tr>
<tr>
<td>$B_{PX}$</td>
<td>Maximum chine beam,</td>
</tr>
<tr>
<td>$B_w$</td>
<td>Maximum waterline beam, feet</td>
</tr>
<tr>
<td>$C_v$</td>
<td>$\sqrt{v/\sqrt{gb}}$</td>
</tr>
<tr>
<td>$C_A$</td>
<td>Load coefficient, $\Lambda/wB^3$</td>
</tr>
<tr>
<td>CDF</td>
<td>Cumulative distribution function</td>
</tr>
<tr>
<td>d</td>
<td>Full load static draft, feet</td>
</tr>
<tr>
<td>DRI</td>
<td>Dynamic Response Index</td>
</tr>
<tr>
<td>F</td>
<td>Longitudinal pressure distribution factor</td>
</tr>
<tr>
<td>$F_D$</td>
<td>Design area factor, ABS</td>
</tr>
<tr>
<td>$F_{NV}$</td>
<td>Volume Froude Number</td>
</tr>
<tr>
<td>$F_Y$</td>
<td>Vertical acceleration distribution factor, ABS</td>
</tr>
<tr>
<td>g</td>
<td>Acceleration due to gravity,</td>
</tr>
<tr>
<td>$H_{1/3}$</td>
<td>Significant wave height, feet</td>
</tr>
<tr>
<td>$K_D$</td>
<td>Pressure reduction coefficient</td>
</tr>
<tr>
<td>L</td>
<td>Craft length, feet</td>
</tr>
<tr>
<td>LCG</td>
<td>Longitudinal center of gravity</td>
</tr>
<tr>
<td>$L_p$</td>
<td>Projected chine length,</td>
</tr>
<tr>
<td>$L_w$</td>
<td>Craft length on the waterline, feet</td>
</tr>
<tr>
<td>MIF</td>
<td>Motion induced fatigue</td>
</tr>
<tr>
<td>MII</td>
<td>Motion induced interruption</td>
</tr>
<tr>
<td>MSI</td>
<td>Motion sickness incidence</td>
</tr>
<tr>
<td>$\bar{n}$</td>
<td>Statistical average</td>
</tr>
<tr>
<td>$\bar{n}_{1/N}$</td>
<td>Statistical average of the highest $1/N^{th}$ values</td>
</tr>
<tr>
<td>$n_{cg,1/100}$</td>
<td>Average of one-hundredth highest acceleration, ABS, g's</td>
</tr>
</tbody>
</table>
\( n_{cg} \) Average acceleration, Savitsky-Koelbel, g's

\( \left( \bar{\eta}_{1/10} \right)_{CG} \) Average of one-tenth highest acceleration, Hoggard-Jones, g's

\( N \) Number of data points in a data set

\( N_I \) Constant, ABS

\( N_Z \) Impact load factor

\( P_{b, cg} \) Impact Pressure, ABS Equation, psi

\( P_D \) Impact Pressure, Allen-Jones Equation, psi

PDF Probability Density Function

RMS Root mean square

\( v \) Craft speed, feet per second (fps)

\( V_K \) Craft speed, knots

\( V_K/\sqrt{L} \) Speed-length ratio

\( w \) Specific weight (fresh water 62.3 lb/ft^3)

WBV Whole body vibrations

\( \beta \) Deadrise angle, degrees

\( \Delta \) Displacement at design waterline, lbs

\( \tau \) Trim angle, degrees
CHAPTER 1

1. Introduction

The V-type hard chine monohull form, commonly referred to as a planing hull, is commonly utilized in the field of high-speed vessels for military, racing, and recreational use. Its main advantage is the dynamic lift it can achieve at high speed-to-length ratios due to its relatively flat bottom shape. However, when operating at high speeds in waves the planing hull can experience high magnitude, repetitive vertical accelerations due to the interaction of the hull with the water’s surface. High vertical accelerations have a negative influence on the overall performance of the craft including ride quality, structural integrity and weight, personnel comfort and safety, and equipment reliability. An accurate prediction of vertical acceleration is of first order importance beginning in the concept design phase (Hoggard and Jones, 1980).

Seakeeping of Planing Craft

Good seakeeping qualities in planing craft, including low vertical accelerations, are directly translatable into financial and human-factor benefits (Blount and Hankley, 1976). By reducing vertical accelerations the following benefits are achieved:

- Reduced hull structural material required during construction
- Increased payload fraction or reduced engine power and fuel for equal speed
- Reduced subsystem foundations and isolation requirements
- Increased time for crew to function without decreased proficiency
• Reduced probability of crew injury.

Vertical acceleration is a critical parameter affecting structural design, vessel classification, habitability, and personnel readiness and safety of planing craft. However, there are many challenges surrounding vertical accelerations for planing craft, including calculation, statistics, analysis and reporting. Koelbel (1995) examined the uncertainty in the structural design process for high-speed craft and concluded that vertical acceleration is the “single most pressing problem” facing planing hull designers. A proper understanding of the relationship of the geometry of the hull and the mass distribution with the craft’s vertical acceleration is vital to achieve a design with good seakeeping and habitability characteristics that meets all of the design criteria. Examples of planing craft monohull forms are shown in Figure 1.1. An example of a planing craft in waves is shown in Figure 1.2.

Figure 1.1: Various Planing Craft Designs (DLBA, 2010)
To quantify motions of planing craft, the designer could first look to adopt the seakeeping methodology that has historically been applied to slower moving displacement vessels operating at speed-length ratios \( \left( \frac{V_K}{\sqrt{L}} \right) \) less than 1.3. For displacement vessels, the theory of linear superposition can be applied, which assumes the motion of the vessel is linear with respect to wave height (Lewis, 1989). Fridsma (1969, 1971) first evaluated the validity of this approach for planing hulls. Fridsma conducted the earliest reported systematic study of seakeeping model tests on prismatic hulls in regular and irregular head seas. Irregular waves followed the Pierson-Moskowitz spectra, which is a one-parameter spectrum based on the average of the third highest wave height \( (H_{1/3}) \), also referred to as the significant wave height \( (H_s) \). Fridsma’s test program was the first of its kind, and in his reports Fridsma (1969, 1971) included discussions of model configuration, test techniques, data reduction, and statistical analysis of craft response in irregular waves. He showed that craft accelerations are random and highly nonlinear in relation to wave height as shown in Figure 1.3.
This nonlinear relationship between wave height and vertical acceleration was further confirmed in later test programs (Brown and Klosinski, 1980). As a result, linear superposition theories that are acceptable to predict seakeeping response of displacement hulls are not applicable for planing craft operating at speed-length ratios in the planing regime, $V_K/\sqrt{L}>2$. Currently, testing or empirically derived methods are generally used to characterize the motions and accelerations of planing craft. Since the vertical accelerations in irregular waves are random, statistics are used to characterize the planing hull response.

One challenge designers face early in the design process is that vertical acceleration criteria exist in a wide range of different statistical forms. Some criteria, such as structural integrity requirements, make reference to the statistics of the positive peak acceleration values, the Peak data set; other measures such as certain habitability limits consider the Root Mean Square (RMS) of the Parent data set, which is the entire data
signal. The terminology of Peak and Parent data sets will be used extensively throughout the remaining chapters. Figure 1.4 shows a representative Parent data set, a time history of the vertical accelerations collected from an accelerometer installed on the craft operated at planing speed in head seas.

![Figure 1.4: Parent Data Set, Time History of Vertical Acceleration Data](image)

Habitability requirements, such as the RMS as shown in Figure 1.4, are often based on Parent data set statistics, as the Parent data set relates to the overall exposure of the craft’s motions.

Structural criteria are often related to higher order statistics of the Peak data set in order to address extreme values and reliability-based design. The Peak data set is a subset of the Parent data set and is a collection of the positive peak acceleration values generated by analyzing the Parent data set according to a user-identified peak identification method and threshold value as shown in Figure 1.5. The peak identification method illustrated is the Vertical Threshold method, which will be discussed in further detail in subsequent
chapters. Any data point above the user-defined vertical threshold value of 1.25 g’s, such as the data points circled on Figure 1.5, will be collected and deposited into the Peak data set. Any points in the Parent data set that have a value less than 1.25 g’s will be discarded.

![Figure 1.5: Peak Data Set, Selection of Peak Values from Parent Data Set](image)

The current approach to designing planing craft structures requires the designer to have knowledge of extreme value statistics of the Peak data set, such as the average of the $1/N$ highest Peaks. For example, the average of the one-tenth highest Peaks represents the average of the highest 10% of the Peak data set. The average of the one-hundredth highest Peaks represents the average of the highest 1% of the Peak data set. These are illustrated in Figure 1.6.
Several researchers (Savitsky and Koelbel, 1978), (Zseleczky and McKee, 1989) point out that the average of the $1/N$th highest statistic was first discussed in the marine field in relation to wave heights. In ocean wave studies, the average of the one-third highest wave heights ($H_{1/3}$) was first characterized by Scripps Institute of Oceanography (Scripps, 1944). The idea was suggested to neglect very small waves and measure only the highest one-third of the remaining waves during a research program focused on characterizing ocean waves. It was determined based on comparison that the wave height estimated by observers typically corresponded with the average of the one-third highest recorded wave heights. This concept is illustrated in Figure 1.7 from Ainsworth (2010). The shaded area shows the highest one-third (33.3%) number of waves in the data set. The average height of waves in this shaded group is the average of the one-third highest, $H_{1/3}$. The average height of the highest 10% of waves ($H_{1/10}$) is also shown and is to the right of $H_{1/3}$. The average height of the highest 1% of waves ($H_{1/100}$) is not shown on the
graph but would be further to the right of \( H_{1/10} \). This statistical approach first used to quantify ocean waves was then adopted to quantify craft motions and accelerations.

![Graph showing statistical parameters and distribution of wave heights.](image)

**Figure 1.7: Statistical Parameters and Distribution of Wave Heights (Ainsworth, 2010)**

**Planing Craft Design Criteria**

Grimsley (1998), Koelbel (1995), and Silvia (1978) surveyed different design methods used for estimating the design pressures on planing hull bottom structure. Statistics of the Peak vertical acceleration is often an input required in most of the existing methods. A commonly used structural design method was described in Allen and Jones (1978). The Allen-Jones equation used to calculate impact design pressure, \( P_D \) (psi), in the structural design of planing craft is as follows:

\[
P_D = 4.44N_ZFK_Dd
\]  

In the equation above, \( N_Z \) is the impact load factor, \( F \) is the longitudinal pressure distribution factor, \( K_D \) is the pressure reduction coefficient, and \( d \) is full load static draft.
in feet. Allen and Jones (1978) concluded that the most difficult and most controversial input required in calculating the impact pressure for structural design is determining the impact load factor, $N_Z$, which correlates to the average of the one-tenth highest Peak vertical acceleration, in g’s.

For craft requiring licensing and classification, designers must comply with structural design standards specified by classification societies such as the American Bureau of Shipping (ABS), Det Norske Veritas (DNV), or Lloyd’s Register. Classification societies have impact design pressure equations that require the designer to input the average of the one-hundredth highest Peak vertical acceleration, in g’s. The following equation (1.2) is excerpted from the ABS High Speed Naval Craft Rules (ABS, 2003) for bottom-slamming pressure on craft less than 61 meters (200 feet) in length. As shown above in equation (1.1), the Allen-Jones formula requires the designer to input the average of the one-tenth highest Peak vertical acceleration. In contrast, ABS requires that the average of the one-hundredth highest Peak vertical acceleration, in g’s, be used in the calculation of impact pressure, $P_{b,cg}$ (psi) for planing craft bottom structure:

$$P_{b,cg} = \frac{N_f \Delta}{L_w B_w} \left[1 + \frac{n_{cg,1/100}}{F_D F_V} \right]$$  \hspace{1em} (1.2)

In the equation above, $n_{cg,1/100}$ is the average of one-hundredth highest Peak vertical accelerations in g’s at the craft’s center of gravity, $\Delta$ is the displacement at design waterline, $L_w$ is the craft length on the waterline at design displacement, $B_w$ is the maximum waterline beam, $F_D$ is the design area factor, $F_V$ is the vertical acceleration distribution factor, and $N_f$ is a constant.
These two equations for impact pressure have different parameters and will yield different values for impact pressure. In general terms, both equations take into account the displacement (weight) of the craft, the overall dimensions of the craft, the dimensions of the local structure being designed such as the unsupported span, and the location of the structure along the length of the craft which accounts for the longitudinal distribution of impact pressure which is part of the slamming phenomena. Further details regarding impact pressure predictions and an evaluation of these and other methods to predict impact pressures on seven (7) planing monohulls is included in Grimsley (1998).

Beyond structural criteria, the designer may also need to comply with a range of statistically based design criteria to ensure the safety of passengers and crew. Hubble (1980) published limits for habitability on military craft based on the crew's ability to perform military functions. These limits were based on an average of the one-tenth highest Peak vertical accelerations and are listed in Table 1.1.

<table>
<thead>
<tr>
<th>Duration</th>
<th>Limit on average of one-tenth highest vertical accelerations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2 hours</td>
<td>&lt; 1.5 g’s</td>
</tr>
<tr>
<td>&gt; 4 hours</td>
<td>&lt; 1.0 g’s</td>
</tr>
</tbody>
</table>

Other guidelines for habitability include those that analyze statistics of the Parent data set to determine the Root Mean Square (RMS), Motion Sickness Incidence (MSI), Motion
Induced Interruptions (MII), Motion Induced Fatigue (MIF), Whole Body Vibrations (WBV), and Dynamic Response Index (DRI). The reader is referred to the survey conducted by Schleicher, Bowles (2004) and the references cited therein for more details. In early stage design, how does the designer estimate these statistical parameters of the Peak and Parent data for input into the design criteria?

1.3 Empirical Methods for Vertical Accelerations

To date, vertical accelerations are characterized either through analysis of test data or through empirically derived methods. Fridsma (1971) presented a series of design charts and calculation procedures for predicting impact loads on hull structure at the craft’s bow and longitudinal center of gravity (LCG) based on the results of model testing prismatic hullforms in irregular head seas. Savitsky and Brown (1976) summarized the work of Fridsma and presented simplified expressions for the average Peak vertical acceleration at the craft’s bow and LCG. The Savitsky-Brown equation to predict the average Peak vertical acceleration at the LCG is:

\[
\bar{a}_{CG} = 0.0104 \left( \frac{H_{1/3}}{b} + 0.084 \right) \frac{L}{b} \left( 1 - \frac{5}{3} \frac{\beta}{\beta} \frac{V_k}{L} \right)^2 \frac{L}{b} C_{\Delta}
\]

In the equation above, \( H_{1/3} \) is the significant wave height, \( b \) is the craft beam, \( L \) is the craft length, \( V_k \) is the craft speed in knots, \( \tau \) is the planing trim angle, \( \beta \) is the deadrise angle, and \( C_{\Delta} \) is the load coefficient.
The regression method developed by Hoggard and Jones (1980) is another empirical method used by designers to estimate statistical parameters of Peak vertical acceleration. Hoggard and Jones developed an equation for the average of the one-tenth highest Peak acceleration at the craft’s LCG based on analyzing experimental data of planing craft Peak vertical acceleration collected both at model and full scale. The Hoggard-Jones equation to predict the average of the one-tenth highest Peak acceleration at the LCG is:

\[
(n_{1/10th})_{CG} = 7.0 \left[ \frac{H_{1/3}}{B_{PX}} \right] \left[ 1 + \frac{\tau}{2} \right]^{0.25} \left[ \frac{L}{B_{PX}} \right]^{-1.25} \left[ F_{NV} \right].
\]  

(1.4)

In the equation above, \(H_{1/3}\) is the significant wave height, \(B_{PX}\) is the maximum chine beam, \(L_P\) is the projected chine length, \(\tau\) is the planing trim angle, and \(F_{NV}\) is the Volume Froude Number.

The Savitsky and Brown (1971) and the Hoggard and Jones (1980) methods are used by planing craft designers in early stage design to predict statistical parameters of the Peak vertical acceleration for new planing hulls. However, these two methods calculate different statistical values. The Savitsky and Brown (1971) method calculates the average of the Peak acceleration data set. The Hoggard and Jones (1980) method calculates the average of the one-tenth highest of the Peak acceleration data set. As discussed earlier, some impact pressure equations require the designer to input the average of the one-tenth highest peak acceleration. Other impact pressure equations require the designer to input the average of the one-hundredth highest peak acceleration.
Further, when considering habitability requirements, the designer must consider statistics such as the RMS of the Parent data set. In the existing framework of planing craft design standards, how can the designer properly work between these various statistics? To do so, the designer must have an understanding of the probability distribution of vertical acceleration data.

1.4 Assumed Probability Distribution

Fridsma (1971) conducted a manual data analysis procedure to count the positive peak values of the vertical acceleration time history recorded from model test results of prismatic hulls in irregular seas. Positive peak data were manually collected by inspecting oscillograph records, grouping the peaks, and analyzing the resultant peak data sets. Fridsma reported that unlike wave height time histories, vertical acceleration data did not follow the Rayleigh distribution. Based on analysis of the data he collected, Fridsma concluded that the Peak data sets followed the Exponential distribution.

The Exponential distribution is a one-parameter class of the Weibull distribution under the Extreme Value Distribution Family. The Probability Density Function (PDF) of the Exponential Distribution is:

\[
\begin{align*}
    f_x(x) &= ve^{-x} \\
\end{align*}
\]  

(1.5)

In the equation above, the single parameter, \( v \), is often referred to as the occurrence parameter.
The Cumulative Distribution Function (CDF) of the Exponential Distribution is:

\[ F_X(x) = 1 - e^{-\nu x} \quad . \]  

(1.6)

The Mean of the Exponential Distribution Function is:

\[ E(X) = \frac{1}{\nu} \quad . \]  

(1.7)

The Variance of the Exponential Distribution Function is:

\[ Var(X) = \frac{1}{\nu^2} \quad . \]  

(1.8)

Because the Exponential distribution is a single-parameter distribution, the mean value of a sample set can be used to calculate the parameter \( \nu \) and recreate the entire set. Within the Exponential distribution, other statistical averages of the distribution can be calculated using the following equation, where \( N \) is the number of data points, \( \bar{a} \) is the average of the Peak vertical acceleration data, and \( \bar{a}_{1/N} \) is the average of the \( 1/N \)th highest of the Peak vertical acceleration data:

\[ \bar{a}_{1/N} = \bar{a} (1 + \ln N) \quad . \]  

(1.9)

Savitsky and Brown (1976), Hoggard and Jones (1980), and the classification standards implemented the conclusion made by Fridsma (1971) that the Peak acceleration data followed the Exponential distribution. Using Equation (1.9), they referenced the
multipliers shown in Table 1.2 to determine the statistical parameters of the Peak acceleration data required in the design process.

<table>
<thead>
<tr>
<th>Desired Peak Statistic</th>
<th>Peak Average Multiplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_{1/3}$</td>
<td>$2.1a$</td>
</tr>
<tr>
<td>$a_{1/10}$</td>
<td>$3.3a$</td>
</tr>
<tr>
<td>$a_{1/100}$</td>
<td>$5.6a$</td>
</tr>
</tbody>
</table>

1.5 **Shortcomings of Earlier Work**

Under the assumption that Peak vertical accelerations follow the Exponential distribution, the designer would only need to know the average of the Peak vertical acceleration data in order to determine the average of the one-tenth highest of the Peak accelerations, $\bar{a}_{1/10}$, or the average of the one-hundredth highest of the Peak acceleration, $\bar{a}_{1/100}$, for input into the impact pressure equations for structural design. The planing craft design industry has operated under this assumption for the past forty years. However, several references have questioned the use of the Exponential distribution for vertical accelerations. Brown and Klosinski (1980) carried out a model test program similar to Fridsma (1971) using higher length-to-beam (L/B) ratio planing hulls and attempted to fit an Exponential distribution to the Peak vertical accelerations data. They reported the data collected did not follow any known distribution. Data analysis from model testing of a high speed, double chine planing hull in irregular head seas suggested that the probability distribution for the vertical accelerations is not likely to follow the Exponential distribution (Blount,
Schleicher and Buescher, 2006). Some researchers have suggested that Fridsma's selection of the Exponential distribution of the data may have been the result of the limited data sampling rate and the data analysis methods available at that time and suggest that modern accelerometers with higher sampling rates and digital analysis techniques may yield different results (Schleicher and Bowles, 2004).

As a starting point to an in-depth examination of the statistical behavior of planing craft vertical accelerations, the author conducted an initial study to investigate the validity of the Exponential distribution for positive peak values of vertical acceleration before proceeding. Published planing hull model test data (Savitsky and Koelbel, 1978) were expanded around the average of the Peak accelerations using Equation (1.9) and the multipliers shown in Table 1.2.

The average of the Peak accelerations from the data set was reported as 0.12 g’s. Using the multiplier shown in Table 1.2, the average of the $1/3$ highest Peak acceleration, $\bar{a}_{1/3}$ would be calculated as follows:

$$\bar{a}_{1/3} = (0.12)(2.1) .$$  \hspace{1cm} (1.11)

The average of the one-tenth highest Peak acceleration, $\bar{a}_{1/10}$ would be calculated as follows:

$$\bar{a}_{1/10} = (0.12)(3.3) .$$  \hspace{1cm} (1.12)
In the measured data set, Savitsky and Koelbel (1978) reported that there were a total of 70 encounters. Equation (1.9) is used to calculate the average of the 1/70th highest Peak acceleration, $\bar{a}_{1/70}$, as follows:

$$\bar{a}_{1/70} = (0.12)(1 + \ln 70) \quad . \quad (1.13)$$

If the designer needed to class the vessel, the average of the one-hundredth highest peak acceleration, $\bar{a}_{1/100}$, would need to be estimated, since only 70 encounters were tested. Using the Exponential distribution assumption that has been assumed to date, the designer would use the multiplier shown in Table 1.2 as follows:

$$\bar{a}_{1/100} = (0.12)(5.6) \quad . \quad (1.14)$$

The results of Equations (1.11) to (1.13) are compared with the reported values and included in Table 1.3 using the calculated error as a percentage, shown below:

$$Error(\%) = \left( \frac{PublishedValue - PredictedValue}{PublishedValue} \right) \times 100\% \quad . \quad (1.15)$$

There was no test value to compare against Equation (1.14) since only 70 wave encounters were measured. Based on the calculated error, the Exponential distribution is not the correct distribution for this published planing hull vertical accelerations dataset (Savitsky, Koelbel, 1978).
Table 1.3: Prediction of 1/N<sup>th</sup> Peak Acceleration Statistics Using Exponential Distribution

<table>
<thead>
<tr>
<th>Statistical Parameter</th>
<th>Published Value</th>
<th>Predicted Value using Exponential distribution</th>
<th>Calculated Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>0.12 g's</td>
<td>Base</td>
<td>-</td>
</tr>
<tr>
<td>Average of the 1/3&lt;sup&gt;rd&lt;/sup&gt; Highest</td>
<td>0.20 g's</td>
<td>0.252 g's</td>
<td>26%</td>
</tr>
<tr>
<td>Average of the one-tenth Highest</td>
<td>0.23 g's</td>
<td>0.396 g's</td>
<td>72%</td>
</tr>
<tr>
<td>Extreme (70 Encounters)</td>
<td>0.30 g's</td>
<td>0.630 g's</td>
<td>110%</td>
</tr>
</tbody>
</table>

Of significance is that had the designer assumed the Peak vertical accelerations data followed the Exponential distribution and had expanded the data around the average of the Peak data as was described in Equations (1.11) to (1.14), the calculated Peak acceleration statistics would be 70% to well over 100% higher than the measured values. If the designer was using the Allen-Jones equation for impact pressure calculation, which requires the average of the \( \bar{a}_{1/10} \) as input, the predicted value using the Exponential distribution is 72% higher than the measured value. If the designer was using the ABS equation for impact pressure, which requires \( \bar{a}_{1/100} \) as input, it is likely that the predicted value using the Exponential distribution would be much more than double the actual acceleration level experienced on the craft. No measured value exists for comparison of the average of the one-hundredth highest Peak acceleration since only 70 encounters were made; however, the average of the 1/70<sup>th</sup> highest Peak acceleration predicted, \( \bar{a}_{1/70} \), using the Exponential distribution is 110% higher than the measured value. Higher impact pressures equate to larger structural members and higher structural weight. High
structural weight typically results in larger engines, propulsion system, and fuel required which drives up the overall weight of the craft significantly. Planing craft performance is extremely weight sensitive, and increases in the weight of the craft would necessarily have a detrimental affect on the craft performance (other weights equal) or the permissible payload capacity (in order to meet performance requirements).

Beyond the shortcomings that exist in the design methods for vertical accelerations, there are considerable variations among testing institutions regarding analysis and reporting of vertical acceleration data collected during model scale and full scale testing of planing hulls in irregular waves. In general, testing institutions do not analyze the measured Peak or Parent test data to determine the underlying distribution of actual data collected. Some institutions assume the Peak acceleration data set fits the Exponential distribution and only report out the average of the Peak accelerations, expecting the designer to recreate the distribution based on this single parameter and following a similar approach as was described above with results shown in Table 1.3. Other institutions calculate the RMS of the Parent data set and then assume the Rayleigh distribution to report statistical parameters of the Peak data set.

Peak identification methods and threshold values have also been regarded as a significant shortcoming regarding analysis of experimental Peak vertical acceleration data. It has been acknowledged in several studies (Savitsky and Koelbel, 1978 and 1992), (Zseleczky and McKee, 1989) that although the most common approach to data analysis is to identify positive peaks and report out statistics based on these peaks, this methodology
has shortcomings due to subjective input from the user and its affect on the resultant $1/N^{th}$ highest statistics of the Peak data set. Figure 1.5 illustrates the process of the peak identification method and setting a user-defined threshold value of 1.25 g’s. Figure 1.8 illustrates that by using the same peak identification method, but selecting a higher threshold value of 2.0 g’s, the number of peaks identified and the resulting statistics of the Peak data set would be significantly different. Shortcomings of peak identification methods and selection of threshold values will be discussed in further detail in Chapter 2.

Figure 1.8: Peak Data Set, Sensitivity of Peak Identification and Threshold

The reporting standards for vertical accelerations are incomplete. In general, the peak identification method and threshold values used to determine the Peak data set are not reported. The underlying distributions of the Peak or Parent data sets are not discussed or reported. The data sets are not reported in their entirety, only the resulting statistics. Further, the statistical parameters that are reported will vary based on the testing
institution, so a complete set of Parent and Peak data statistics from a single test program are not reported together.

These shortcomings in analysis and reporting of vertical accelerations data can lead to significant error in the design of planing craft and have made it impossible for designers to compare a data set from one source with that from another with any confidence. As a result the designer cannot evaluate a potential design against existing, proven hull forms. Overall, these deficiencies in planing craft testing and design can only be remedied through correct understanding of the probability distribution and statistical behavior of vertical acceleration data.

1.6 Scope of the Dissertation

Planing hull seakeeping issues can be grouped into three categories that must be improved (Zseleczky and McKee, 1989):

1. Better understanding of specific hydrodynamic effects that influence the performance of the craft to complete its mission;
2. Improved data analysis techniques;
3. Improved theoretical model of planing behavior that incorporates the two categories previously mentioned.

The scope of this research falls within the second category - that of expanding the knowledge of the statistical behavior of planing craft vertical acceleration data. As Zseleczky and McKee point out, the solution will not be deterministic as the response has
been proven to be nonlinear and only a general characterization of the sea state is to be expected. A probabilistic approach to planing hull vertical accelerations in irregular seas is needed to help achieve the much-needed advancements in the knowledge of high speed planing craft design.

An in-depth examination of the statistical behavior of planing craft vertical accelerations was carried out in this research program. The motivation of this research program was to develop a methodology to quantify vertical accelerations data that included Peak data sets, based on the doubt that has been raised regarding the Exponential distribution, as well as Parent data sets, which previously had not been addressed in the literature. The areas of focus for this research program include:

- Statistical Distribution of Peak and Parent Acceleration Data,
- Statistical Parameters of Peak and Parent Accelerations Data,
- Correlation between Parameters of Peak and Parent Accelerations Data,
- Supplementary Statistical Parameters of Parent Acceleration Data.

It was important to collect as large a test matrix as possible that represented modern planing hull dimensions and operating profiles, from both full-scale and model-scale testing institutions. The author analyzed vertical acceleration data collected from twenty-eight (28) different tests on planing hulls operating at speed in irregular head seas. The tests were conducted prior to this research effort, and the data files were provided to the author with the agreement that specifics of the hulls not be disclosed. In all test cases, vertical accelerations were measured at the craft’s LCG. Tests were run on planing hulls with differing hull form geometries, operating at a range of speeds, hull loadings, and sea
conditions. Nineteen (19) test cases were collected at full scale and nine (9) test cases were collected at model scale. Parameters for the data sets examined fell into the ranges shown in Table 1.4.

<table>
<thead>
<tr>
<th>Design Parameter</th>
<th>Data Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speed</td>
<td>$2.8 \leq F_{NV} \leq 4.5$</td>
</tr>
<tr>
<td>Sea Condition</td>
<td>$0.3 \leq \frac{H_{1/3}}{V^{1/3}} \leq 0.8$</td>
</tr>
<tr>
<td></td>
<td>$0.2 \leq \frac{H_{1/3}}{B} \leq 0.5$</td>
</tr>
<tr>
<td>Hull Geometry</td>
<td>$3.2 \leq \frac{L}{B} \leq 4.7$</td>
</tr>
<tr>
<td>Hull Loading</td>
<td>$2.5 \leq \frac{L}{V^{1/3}} \leq 5.8$</td>
</tr>
</tbody>
</table>

In the table above, $F_{NV}$ is the volumetric Froude number and is commonly used to non-dimensionalize planing hull speed as a function of the volume of displacement. Displacement ships operate at a much lower $F_{NV}$, typically below 1.0. The significant wave height, $H_{1/3}$ measured in the tests is commonly non-dimensionalized according to either the cube root of the volume of displacement $V^{1/3}$, or the craft beam (width), $B$, and is used to describe the sea condition relative to the size of the planing hull. Planing craft hull geometry is commonly related to the non-dimensional ratio of the craft length, $L$, to the craft beam, $B$. All test hulls were between 30 feet to 100 feet in length. Finally, as planing craft are dynamic-assist hullforms that rely on the lift generated by the hull bottom at high speeds, the hull loading is of interest to understand if the craft is “lightly-loaded” or “heavily loaded” and is commonly described by non-dimensionalizing the craft length, $L$, relative to $V^{1/3}$. 
The remaining chapters describe the technical approach to quantify Parent and Peak accelerations, the results of the analyses, and conclusions for the research program. Chapter 2 presents the details of the statistical distribution analysis that was carried out on the Peak data sets and the Parent data sets for each of the 28 test cases. The first focus of the statistical distribution analysis discussed in Chapter 2 was to examine the distribution of the Peak acceleration data sets to determine if the Exponential distribution is indeed the correct distribution for planing craft operating in waves as this has been the fundamental basis to current planing craft design standards regarding vertical accelerations. The second focus of the statistical distribution analysis discussed in Chapter 2 was to examine the distribution of the Parent acceleration data sets. To the author’s knowledge, the underlying statistical distribution of the Parent data has not been examined in previous research efforts regarding planing craft accelerations. Four distributions, including the Exponential distribution, were considered in this research program and are described in Chapter 2. The Kolmogorov-Smirnov (K-S) test was used to quantify the goodness of fit of each data set to the distributions and is described in Chapter 2. Sensitivity studies regarding peak identification methods and threshold values are also discussed in Chapter 2.

Chapter 3 examines the statistical parameters currently used for planing craft design that have been previously discussed in Chapter 1 such as the average, the standard deviation, and the average of the $1/N^{th}$ highest statistics for the Peak data as well as the RMS for the Parent data. Chapter 3 also presents three methods that can be used to determine these parameters: the direct method of analyzing the experimental data, the analytical method
based on a given distribution, and the Monte Carlo simulation method. The Monte Carlo simulation technique was used to simulate the Peak and Parent data and verify the findings of the K-S Test with regard to the statistical distribution of the data sets. The Monte Carlo technique was then evaluated to determine the accuracy of statistical parameters extracted from the simulated data sets in comparison to the experimental data sets. Chapter 3 also examines the relationship between statistical parameters of the Parent and Peak data sets and explores any correlations that exist which will benefit the designer in extending legacy data from previous test programs and in meeting current design criteria such as structural and habitability requirements. Finally, Chapter 3 discusses other statistical parameters of the Parent data set, not currently specified in data analysis or design methodologies of planing craft, which may benefit the designer in the future.

Chapter 4 presents the results of the research for all 28 test cases and discusses the findings for distribution of the Peak data sets, distribution of the Parent data sets, sensitivity studies into the peak identification and threshold methods, results and comparisons of the direct, analytical, and Monte Carlo simulation methods, correlations between statistical parameters of Peak and Parent data sets, and tables of supplementary statistical parameters of the Parent data set for future analyses.

Chapter 5 presents the impacts of this research program on the planing craft design and testing community, identifies areas for future work, and gives concluding remarks.
CHAPTER 2

2. Statistical Distribution Analysis

As was discussed in Chapter 1, planing craft vertical accelerations are random and nonlinear in relation to the sea condition; thus, linear methods implemented in displacement ship seakeeping analysis are not appropriate. Chapter 1 discussed the critical importance of vertical acceleration to the design of planing craft and the designer’s need to satisfy a range of statistically based design criteria for structural design, habitability, and equipment selection. The historical assumption reported by Fridsma (1971) has been that Peak vertical acceleration data fit the Exponential Distribution. Existing data reporting methods and regression analyses to predict vertical accelerations adopted this conclusion. No recommendations were provided in the literature regarding the statistical distribution of the Parent data set. The distribution of the Parent data set will also be explored in this research program.

Chapter 2 presents the details of the statistical distribution analysis that was carried out on the Peak data sets and the Parent data sets for each of the 28 test cases. In each test case, the distribution of the Peak acceleration data set was compared to four different distributions, including the Exponential distribution, using the Kolmogorov-Smirnov (K-S) goodness of fit testing. The distribution of each Parent acceleration data set was also evaluated according to the K-S test. An example of the methodology developed to determine the distribution of a data set is presented in Chapter 2. Sensitivity studies
regarding peak identification methods and threshold values are also discussed in Chapter 2.

**Probability Distribution Functions**

For this research, the following four (4) known distribution functions were considered:

1. Exponential distribution
2. Rayleigh distribution
3. Gumbel distribution
4. Lognormal distribution

The Exponential distribution was selected because it has been the assumed distribution for planing craft vertical accelerations for the past 40 years (Fridsma, 1971), (Savitsky and Brown, 1976), (Hoggard and Jones, 1985). The Rayleigh distribution was selected as it represents the distribution of the sea wave height (Scripps, 1944), and it is desirable to validate the earlier conclusions drawn that planing craft vertical accelerations have a non-linear response in relation to the sea wave height (Fridsma, 1969), (Brown and Klosinski, 1980). Both the Exponential distribution and the Rayleigh distribution are one-parameter variations under the two-parameter Weibull distribution. The Exponential distribution is presented in Section 2.1.1. The Rayleigh distribution is presented in Section 2.1.2.

The third distribution selected for the research was the Gumbel distribution. The Gumbel distribution is perhaps the most widely applied statistical distribution for problems in engineering. This two-parameter distribution is the Type 1 subset of the three-parameter
Generalized Extreme Value (GEV) distribution used in Extreme Value Theory (EVT). EVT was developed in the 1950s to address extreme values in the tail of the distribution and to assess the risk of unusual events (Gnedenko, 1943), (Gumbel, 1958). The most widely used form of the GEV is the two-parameter Gumbel distribution, which has been used extensively to model extreme events. Some of its recent application areas in engineering include flood frequency analysis, network engineering, nuclear engineering, offshore engineering, risk-based engineering, space engineering, software reliability engineering, structural engineering, and wind engineering. Kotz and Nadarajah (2000) describe this distribution and list over fifty applications ranging from accelerated life testing through earthquakes, floods, horse racing, rainfall, queues in supermarkets, sea currents, wind speeds, and track race records, to wave mechanics. Gumbel distribution has also been shown to provide good fits to the time series of the extreme dynamic pressures. The Gumbel distribution is presented in Section 2.1.3.

The fourth distribution selected was the Lognormal distribution. Schleicher (2008) non-dimensionalized the Parent data collected from one hull form tested at model scale in order to combine data collected under different loading, speed and sea conditions. Schleicher (2008) reported that the combined, non-dimensionalized Parent data showed some correlation with the Lognormal distribution but observed that there was only a loose correlation towards the “tail” of the distribution. The statistical behavior of the “tail” is of particular interest in design and reliability studies as this contains the higher magnitude, less frequent peak or extreme values that are necessary for design validation. For completeness, the author included the two-parameter Lognormal distribution to
evaluate how well the Lognormal distribution fits the acceleration data from the 28 test cases considered in this research. The Lognormal distribution is presented in Section 2.1.4.

**Exponential Distribution**

The PDF and CDF for the Exponential distribution have been discussed previously in Chapter 1. The reader is referred to Section 1.4 for an explanation of the Exponential distribution.

**Rayleigh Distribution**

The Rayleigh distribution is a one-parameter class of the Weibull distribution under the Extreme Value Distribution Family. The Probability Density Function (PDF) of the Rayleigh Distribution is:

\[
 f_X(x) = \frac{x}{\alpha^2} e^{-\frac{1}{2} \left(\frac{x}{\alpha}\right)^2} .
\]  

(2.1)

In the equation above, the single parameter, \(\alpha\), is often referred to as the modal value of the Rayleigh distribution.

The Cumulative Distribution Function (CDF) of the Rayleigh Distribution is:
The Mean of the Rayleigh Distribution Function is:

\[ E(X) = \sqrt{\frac{\pi}{2}} \alpha \]  \hspace{1cm} (2.3)

The Variance of the Rayleigh Distribution Function is:

\[ Var(X) = \left( 2 - \frac{\pi}{2} \right) \alpha^2 \]  \hspace{1cm} (2.4)

**Gumbel Distribution**

The Gumbel distribution is a two-parameter distribution, where \( \mu \) is the location parameter and \( \theta \) is the scale parameter, under the Extreme Value Distribution Family. The Probability Density Function (PDF) of the Gumbel Distribution is:

\[ f_X(x) = \frac{e^{-\frac{x-\mu}{\theta}} e^{-e^{-\frac{x-\mu}{\theta}}}}{\theta} \]  \hspace{1cm} (2.5)
The Cumulative Distribution Function (CDF) of the Gumbel Distribution is:

\[ F_X(x) = e^{-e^{-\frac{x-\mu}{\theta}}} \]  \hspace{1cm} (2.6)

The Mean of the Gumbel Distribution Function is:

\[ E(X) = \mu + \theta \gamma \]  \hspace{1cm} (2.7)

The Variance of the Gumbel Distribution Function is:

\[ Var(X) = \frac{\pi^2}{6} \theta^2 \]  \hspace{1cm} (2.8)

**Lognormal Distribution**

The Lognormal distribution is a two-parameter distribution, where \( \lambda \) and \( \zeta \) are the two parameters of the Lognormal distribution. The Probability Density Function (PDF) of the Lognormal Distribution is:

\[ f_X(x) = \frac{1}{x \zeta \sqrt{2 \pi}} e^{-\frac{(\ln x - \lambda)^2}{2 \zeta^2}} \]  \hspace{1cm} (2.9)

The Cumulative Distribution Function (CDF) of the Lognormal Distribution is:
\[ F_x(x) = \frac{1}{2} + \frac{1}{2} \text{erf} \left( \frac{\ln x - \lambda}{\zeta \sqrt{2}} \right). \quad (2.10) \]

The Mean of the Lognormal Distribution Function is:

\[ E(X) = e^{\lambda + \frac{1}{2} \zeta^2}. \quad (2.11) \]

The Variance of the Lognormal Distribution Function is:

\[ \text{Var}(X) = E(X^2) \left( e^{\zeta^2} - 1 \right). \quad (2.12) \]

For a visual comparison, the PDF and CDF of each of the four distributions described above are presented in Figures 2.1 to 2.4. Figures 2.1 and 2.2 show the PDF and CDF of the four distributions assuming the same mean value. Figures 2.3 and 2.4 show the PDF and CDF of the four distributions assuming the same standard deviation value.
Figure 2.1: PDF of Four Distributions With Equal Mean Value

Figure 2.2: CDF of Four Distributions With Equal Mean Value

Figure 2.3: PDF of Four Distributions With Equal Standard Deviation Value
Goodness of Fit Testing

Each of the 28 test cases was examined and both the Parent data set and the Peak data sets for each test case were compared with the distribution functions described in Section 2.1. For each data set, the PDF of the experimental data was visually compared to the PDF of the known distribution function. The PDF gives the researcher some understanding of the randomness of the peak vertical accelerations by the shape of the curve. Beyond visual inspection and comparison between the experimental data and the distribution functions, it was necessary to assess the correlation quantitatively by comparing the CDF of the experimental data and the CDF of the known distribution function.

Quantitative comparisons of each experimental data set to the four distributions described above were carried out using the Kolmogorov-Smirnov (K-S) Test for Goodness of Fit. The K-S Test is commonly used in statistical analysis to determine the underlying distribution of an experimental data set. This test was selected as it also permits
consideration of the $1/N^{th}$ statistics that are currently well embedded in planing craft structural design methods. The K-S test was run to determine the maximum of the absolute value of the difference ($D_n$) between the CDF of the known distribution $F(x)$ and the CDF of the empirical distribution $F_n(x)$ at each point $x$ along the two curves as follows:

$$\text{MAX } D_n(x) = \text{MAX } |F_n(x) - F(x)| . \quad (2.13)$$

To demonstrate the process, the graph shown in Figure 2.5 below is a plot of an empirical distribution function with a Normal cumulative distribution function for 100 normal random numbers. The K-S test is based on the maximum distance between these two curves.
Mathematically, $D_n$ is a random variable, and its distribution depends on the sample size, $n$. The CDF of $D_n$ can be related to the significance level $\alpha$ as shown in Equation (2.14):

$$P(D_n \leq D_n^\alpha) = 1 - \alpha.$$  

(2.14)

The $D_n^\alpha$ values at various significance levels $\alpha$ can be obtained from a standard mathematical table such as that found in Haldar and Mahadevan (2000). Thus, according to the K-S Test, if the $\text{MAX } D_n$ is less than or equal to the tabulated value $D_n^\alpha$, the assumed distribution is acceptable at significance level $\alpha$.

Peak Acceleration Distribution

The main focus of the analysis discussed in Section 2.3 was to examine the distribution of the Peak acceleration data sets to determine if the Exponential distribution was indeed the correct distribution for planing craft operating in irregular waves. As this has been the fundamental basis for current planing craft design standards regarding vertical accelerations, the author desired to explore this assertion before moving forward to study other aspects of the statistical behavior of planing craft accelerations. The overall approach follows the outline below:

Step 1: Collect vertical acceleration dataset for planing craft operating in irregular head waves (Parent Data Set).

Step 2: Identify Peak Values using a Peak Identification Method (Peak Data Set).
Step 3: Determine the underlying distribution of Peak data set by comparing to known distribution functions using the K-S Test.

Step 4: Simulate the Peak data set using Monte Carlo technique and compare the statistical parameters of the simulated data to experimental data.

Note, Steps 1 through 3 of the method to quantify peak accelerations outlined above will be discussed in Section 2.3. Step 4 will be discussed in Section 3.1.

For any planing craft test, whether it is conducted at full-scale in an open seaway or at model-scale in a towing basin, the seakeeping behavior of the vessel can be characterized by analyzing time histories of the vertical accelerations at known points along the length of the craft. A number of references (Savitsky and Koelbel, 1978), (Savitsky and Koelbel, 1992), and (Haupt, 2003) provide recommendations for test set-up including model size and construction, instrumentation selection, and data collection. This is beyond the scope of this research effort, and the reader is referred to the references cited above for guidance on proper test set-up.

As was discussed in Chapter 1, current structural design standards depend on the analysis and statistics of the Peak vertical accelerations, such as the average of the 1/Nth highest Peak acceleration. In order to determine these statistics, the Peak data set must be extracted from the Parent data set. The most commonly used data analysis approach is to scan the entire time series data of vertical accelerations using a peak identification method whereby the peaks and troughs are identified and sorted (Zseleczyk and McKee,
1989). The sorted peak acceleration values are grouped into a new data set, the Peak data set, which is not time dependent. As described in Chapter 1, the original data analysis procedure conducted by Fridsma (1971) was done by manually inspecting oscillograph records, counting the positive peak values of the vertical acceleration time history, and analyzing the data. Today, there are a number of peak identification methods available as experimenters have attempted to automate what was once done manually. Four methods considered in this research will be discussed in Section 2.3.1.

**Peak Identification Methods**

Chapter 1 identified one of the shortcomings of working with Peak acceleration data: the existing peak identification methods are subjective and require input from the user. There are a number of peak identification methods available to the experimenter and different facilities use different methods. While the approach in each method is slightly different, each requires the user to specify a buffer or threshold value, which is the criterion for sorting the peak values from the non-peak values. All of this is done in an effort to answer a seemingly easy but actually quite difficult problem of "what is a peak?" Selection of threshold values will be discussed in Section 2.3.2.

In the research undertaken herein, it was of interest to explore whether the peak identification method would affect the resultant best fit of the Peak data set distribution. In order to investigate the sensitivity of the Peak data set distribution, four different peak identification methods were considered. These four methods, illustrated in Figure 2.6
and briefly described below, were not developed by the author but instead were selected because they are in use today at various experimentation facilities throughout the world.

Figure 2.6: Peak Identification Methods

The buffer method scans the time history of the Parent data set and will recognize as a peak any maximum, whether above or below the zero level, provided that it exceeds the preceding and following minima by the buffer or threshold amount, T, selected by the user. Similarly, minima, either above or below the zero level, that exceed the buffer amount, T, will be recognized as troughs. Thus, the buffer method is independent of zero-crossings. Peaks and troughs are identified alternately. The buffer method is illustrated in Figure 2.6.

The vertical threshold method scans the time history of the Parent data set and will recognize as a peak any maximum that is above the zero level and is above the threshold amount, T, selected by the user. Similarly, any minimum that is below the zero level and
is below the threshold amount, T, selected by the user will be recognized as a trough. The vertical threshold method is illustrated in Figure 2.6.

The vertical difference method is similar to the vertical threshold method but also has a second criterion that must be satisfied in order for a data point to be considered a peak or trough. The vertical difference method scans the time history of the Parent data set and will recognize a maximum that is above the zero level and is above the threshold amount, T, selected by the user. The method continues to scan the time history looking for a minimum that is below the zero level and is below the threshold amount, T, selected by the user. If the difference in magnitude between the maximum and subsequent minimum is greater than the user defined difference amount, D, then the maximum is saved as a peak and the minimum is saved as a trough. The vertical difference method is illustrated in Figure 2.6.

The horizontal threshold method scans the time history of the Parent data set by using a moving window approach. The user defines a time-window, W, in which the maximum point above zero that falls within this window is identified as a peak. Similarly, the minimum point below zero that falls within this window is identified as a trough. The horizontal threshold method is illustrated in Figure 2.6.

The reader should note that it is not the purpose of this research effort to evaluate or rank the peak identification methods, as it is possible that refinements or improvements to these methods could be made or already exist. The intent instead is to examine if the
selection of a peak identification method affects whether the data follows a specific distribution function.

**Threshold Values**

In each of the peak identification methods described in Section 2.3.2 the user must set the threshold or buffer value that captures the peaks but avoids the small oscillations. This need for subjective input by the user has been discussed as a shortcoming of the peak identification methods. If the threshold value is set too low, the Peak data set is flooded with less significant magnitudes. If the threshold value is set too high, the Peak data set may not have sufficient data points to be considered statistically significant and there may be increased uncertainty in any conclusions drawn. Zseleczky and McKee (1989) pointed out that variations in the threshold value selected by the user could skew the statistical results as it alters the number of events in the Peak data set and thus alters the average of the $1/N$ statistics. Therefore, the user must take care that the solution is independent of the threshold or buffer value selected. There is no clear guidance from the literature on selecting the correct buffer or threshold value. Savitsky and Koelbel (1992) suggested a multiple of the RMS value of the Parent data be used. Zseleczky and McKee (1989) recommended a sensitivity study be performed as part of the analysis technique.
In order to better understand the sensitivity of the resultant distribution of the Peak data set to the threshold or buffer values, the values were varied in the analysis of each data set and for each peak identification method. Again, the emphasis was not to identify the optimal threshold value but to investigate whether the resultant best-fit distribution for the Peak data set is sensitive to the threshold or buffer value.

**Example of Peak Distribution Method**

To demonstrate the method to determine the statistical distribution of the Peak vertical acceleration data, a representative data set (Test Case 4) was selected. This section outlines the approach taken for each of the 28 test cases. This analysis was conducted using MATLAB®. The MATLAB® coding for this research program is included in Appendix A.

To develop the Peak data set from Test Case 4, each of the four (4) peak identification methods were applied to the Parent data set from Test Case 4. In each of the four peak identification methods, a range of threshold values was used. For the buffer and the vertical threshold methods, the threshold values considered were multipliers of the RMS of the Parent data set as follows: 0.25, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, and 4.0 times the RMS of the Parent data set. Nine Peak data sets were developed from each of the buffer and vertical threshold methods based on these threshold values. For the vertical difference method, the threshold values considered were 0.25, 0.5, 1.0, 1.5, 2.0, and the vertical difference was equal to 2 times the threshold value. Five Peak data sets were developed from the vertical difference method based on these threshold values. For the
horizontal threshold method, the threshold values considered were 0.125, 0.25, 0.5, 1.0 and 2.0 multiplied times the reported sampling frequency of the data. For example, if the data were sampled at 512 Hz, the threshold values were 64, 128, 256, 512, and 1024 Hz. Five Peak data sets were developed from the horizontal threshold method based on these threshold values.

The data in each of the Peak data sets were sorted in ascending order. The mean and standard deviation were calculated for each of the Peak data sets. The four known distributions discussed in Section 2.1 were developed for each of the Peak data sets based on the relationships of the distributions’ parameters to the experimental Peak data set’s mean and variance (square of the standard deviation) using the Method of Moments as described in Haldar, Mahadevan (2000). The basic concept of the Method of Moments is that all of the parameters of a distribution can be estimated using the information of its moments since the parameters of the distribution have a definite relationship with the moments of the random variable. For example, the Exponential and Rayleigh distributions are both one-parameter distributions; thus, only one moment, such as the first moment: the mean, is used to estimate the parameter. The Lognormal and Gumbel distributions are both two-parameter distributions, so the first two moments are used – the mean and the variance (square of the standard deviation) of the Peak data set.

The parameters for each of the four known distributions are shown in Table 2.1 where \( E(X) \) is the mean of the experimental Peak data set and \( Var(X) \) is the Variance of the experimental Peak data set.
Table 2.1: Estimating Distribution Parameters using the Method of Moments

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Parameters</th>
<th>Relation to mean and variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>$\nu$</td>
<td>$E(X) = \frac{1}{\nu}$</td>
</tr>
<tr>
<td>Rayleigh</td>
<td>$\alpha$</td>
<td>$E(X) = \sqrt{\frac{\pi}{2}} \alpha$</td>
</tr>
<tr>
<td>Lognormal</td>
<td>$\lambda$, $\zeta$</td>
<td>$E(X) = \exp\left(\lambda + \frac{1}{2}\zeta^2\right)$</td>
</tr>
<tr>
<td>Gumbel</td>
<td>$u$, $\alpha$</td>
<td>$E(X) = u + \frac{0.5772}{\alpha}$</td>
</tr>
</tbody>
</table>

Once the known distributions were developed, the PDF and CDF of the experimental Peak data set were plotted against the PDF and CDF of each of the known distribution functions for visual comparisons as shown in Figures 2.7 and 2.8 below.

Figure 2.7: Visual Comparison of PDF of Peak Data Set
The K-S test was run to calculate the $MAX D_n$ between the CDF of the experimental Peak data set and the CDF of the known distribution function. The results of the K-S test for each of the four peak identification methods and a range of threshold values for the representative data set are shown in Figures 2.9 to 2.12.
Figure 2.9: Peak Distribution Goodness of Fit (Buffer)

Figure 2.10: Peak Distribution Goodness of Fit (Vertical Threshold)
Figure 2.11: Goodness of Fit for Peak Distribution (Vertical Difference)

Figure 2.12: Goodness of Fit for Peak Distribution (Horizontal Threshold)
The results of the analysis of the statistical distribution of the Peak acceleration of this representative data set show that in all the Peak data sets developed from the four Peak identification methods and the range of threshold values considered, the Exponential distribution did not fit the data sets based on the results of the K-S test. Additionally, the Exponential distribution showed the most sensitivity to the peak identification method and threshold value selected. For the past forty years, Peak vertical acceleration data for planing craft has been assumed to fit the Exponential distribution. This assumption was not valid for the representative data set analyzed above. Instead, the Gumbel distribution was the best fit to the representative data set based on the K-S test with a confidence level of 99%. The Gumbel distribution was found to be the best fit for this representative data set for the four peak identification methods and was found to be the least sensitive to the range of threshold values considered.

The method to determine the statistical distribution of the Peak data set as outlined above was carried out on the Peak data sets from each of the 28 test cases. For each test case, the peak accelerations were identified and grouped into Peak data sets using the four different peak identification methods. Sensitivity studies on the threshold or buffer value for each method were also carried out. The CDF of each Peak data set was quantitatively compared to the CDF of each of four different distribution functions described in Section 2.1. The best fit for the CDF was determined using the K-S test described in Section 2.2. The results, including graphs and figures, of the statistical distribution analysis on the Peak data sets are presented in Section 4.1.2.
Parent Acceleration Distribution

Chapter 1 discussed a number of different statistics that are currently required in the design of planing craft. Many of the current structural design methods use statistics based on the Peak data set but others, for example habitability criteria, use statistics based on the Parent data set. There was no discussion or previous research discovered regarding the distribution of the Parent vertical acceleration data set based on the literature survey conducted. It was of interest to the author to investigate the distribution of the Parent data to have a better understanding of the statistical behavior of planing craft accelerations and to provide insight into correlations that may exist between the Parent and Peak data sets. This section discusses the approach taken to investigate the statistical distribution of the Parent data sets for each of the 28 test cases.

In the data reduction and analysis of vertical accelerations, or other time-dependent data, the common approach is to first carry out the process referred to as “de-meaning” the Parent data set. In previous literature on reduction and analysis of planing hull data including that of Fridsma (1971) and others, this method has been referred to as removing the Direct Current offset, or DC-offset. The “de-meaning” approach taken by some testing institutions is to calculate the mean, \( \bar{x} \), of the Parent data set \( X = (x_1, x_2, ..., x_n) \) and then to subtract the mean value from the data set, thus shifting the data set to a mean value of zero and resulting in a “de-meaned Parent data set” \( X_{\text{demean}} \).
For each of the 28 test cases, the CDF of the de-meaned Parent data set the mean and
standard deviation is calculated. Using the Method of Moments described in Table 2.1,
the parameters of the known distributions are calculated. For the analyses carried out on
the de-meaned Parent data sets, the Lognormal distribution was not used because the
Lognormal distribution cannot have a mean value equal to zero. The statistical behavior
of the Parent data sets were compared to the remaining three distribution functions
considered for the Peak data analyses, as shown below:

1. Exponential distribution,
2. Rayleigh distribution,

The CDF of the experimental Parent data set was compared to the CDF of each of the
three distributions listed above. Quantitative comparisons of each of the de-meaned
Parent data sets to the three distributions were carried out using the Kolmogorov-Smirnov
(K-S) Test for goodness of fit as described in Section 2.2 and Section 2.3.3. The K-S test
was run to determine the statistical distribution of the Parent data sets with at least 95%
confidence.

It should be noted here regarding the Peak data set analysis that was discussed in Section
2.3 for comparison the Peak Identification method was applied to Parent data sets which
were both demeaned and un-demeaned. The results of the Peak data set analyses were
found to be insensitive to whether or not the Parent data had been demeaned.
The results of the statistical distribution analysis for the Parent data sets, including graphs and figures, are presented in Section 4.2.1.
CHAPTER 3

3. Statistical Parameters for Design

Chapter 1 discussed a range of statistical parameters currently used in planing craft design such as the average, the standard deviation, and the average of the \( 1/N^{th} \) highest statistics for the Peak vertical acceleration data as well as the RMS for the Parent vertical acceleration data. There are three methods available to the designer to determine these statistical parameters. The three methods considered were the direct method, the analytical method, and the Monte Carlo simulation method. This chapter describes how to apply each of these three methods to the Parent and Peak data sets in order to determine the statistical parameters used for design. Additionally, the MATLAB® coding is included in Appendix A.

Chapter 2 investigated the statistical distribution of the Peak data sets and Parent data sets by applying the K-S Test. This chapter explores the Monte Carlo technique to simulate the Peak and Parent data set by assuming a distribution and thus validating the results of the K-S Test regarding the best fit distribution. The Monte Carlo was also used to estimate the statistical parameters of the Peak and Parent data sets and compared to the statistical parameters of the experimental data sets. This chapter also explores the correlations that exist between the statistical parameters of the Peak and Parent data sets and will discuss the benefits of these correlations for the experimenter and the designer. Finally, this chapter will discuss other statistical parameters of the Parent data set, not
currently specified in data analysis or design methodologies of planing craft, which may be of benefit to the experimenter and the designer in the future.

**Statistical Parameters for Peak Data Sets**

For a given test case, once the Peak data set is extracted from the Parent data set as discussed in Section 2.3.1, the experimenter can directly calculate the statistical parameters necessary in planing craft design, such as the average of the $1/3^{rd}$ highest, average of the one-tenth highest, and average of the one-hundredth highest Peak acceleration values. This is referred to as the direct method. However, if the designer does not have access to the full data set but has limited information about the data set and knowledge of statistical distribution of the data can the designer determine the necessary statistical parameters needed for design? This question is explored herein, using the results obtained from the direct method to assess the accuracy of using an analytical method and a Monte Carlo simulation method.

### 3.1.1 Direct Method

For each Peak data set extracted from the Parent data set using the peak identification methods discussed in Section 2.3.1, the Peak data set is organized in ascending order and the $1/N^{th}$ highest value is found directly. All data points above this cut-off value are averaged together to determine the average of the $1/N^{th}$ highest Peak acceleration values, the statistical parameters used in current planing craft structural design criteria. These
values are compared to the results of the analytical method and the Monte Carlo simulation method to determine the accuracy of the methods. An example of the direct method is described below and in Table 3.1.

Consider a data set containing 20 points, in random order as shown in the first column of Table 3.1. The second column shows the data points arranged in ascending order. The overall average of the twenty points is 2.025. If the designer were interested in the average of the 1/3\textsuperscript{rd} highest, the highest 7 data points would be averaged together as shown in the third column, with a value of 2.67. If the designer were interested in the average of the 1/4\textsuperscript{th} highest, the highest 5 data points would be averaged together as shown in the fourth column, with a value of 2.78. If the designer were interested in the average of the one-tenth highest, the highest 2 data points would be averaged together as shown in the fifth column, with a value of 3.05.
Table 3.1: The Direct Method to Calculate Average of the $1/N^{th}$ Statistics

<table>
<thead>
<tr>
<th>Data</th>
<th>Ascending Order</th>
<th>N=3</th>
<th>N=4</th>
<th>N=10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2</td>
<td>1.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.8</td>
<td>1.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td>1.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>1.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.4</td>
<td>1.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.1</td>
<td>1.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.4</td>
<td>1.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.6</td>
<td>1.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td>1.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.8</td>
<td>1.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.9</td>
<td>2.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>2.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.0</td>
<td>2.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>2.4</td>
<td>2.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td>2.4</td>
<td>2.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.7</td>
<td>2.5</td>
<td>2.5</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>2.4</td>
<td>2.6</td>
<td>2.6</td>
<td>2.6</td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>2.7</td>
<td>2.7</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td>2.7</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>1.6</td>
<td>3.1</td>
<td>3.1</td>
<td>3.1</td>
<td>3.1</td>
</tr>
</tbody>
</table>

Averages: 2.025 2.67 2.78 3.05

3.1.2 Analytical Method

The direct method discussed in Section 3.1.1 is the most accurate approach to determining the statistical parameters of the Peak data set if the designer has access to the experimental data. However, if the designer only has access to limited information about the data set, can additional statistical parameters be accurately estimated? As an example, the designer may have legacy data from previous test reports, where only limited results were reported. The reader will recall that this was the approach taken previously by Fridsma (1971) where he concluded that the Peak acceleration data followed the one-parameter Exponential distribution and the designer could regenerate
the data set by having access to the average of the Peaks, and could calculate the remaining statistical parameters. It was shown in Table 1.3 that an analytical approach could be applied to legacy data; however, the assumption of the Exponential distribution led to high errors when compared to the published results. With the knowledge gained regarding the distribution of the Peak data based on the K-S Test discussed in Section 2.3 and the results reported in 4.1.2, can more accurate estimates be made?

In the four distributions considered for the Peak data sets, analytical expressions exist to calculate the $1/N^{th}$ highest value, commonly referred to as the “cut-off” value, of each distribution. Analytical expressions also exist to integrate and average the $1/N^{th}$ highest values for the Exponential and Rayleigh distributions in order to obtain the desired parameters for the average of the $1/N^{th}$ highest Peak acceleration values. For the Gumbel and Lognormal distribution functions, the Monte Carlo simulation can be used to solve for the average the $1/N^{th}$ highest values based on the cut-off value from the analytical expression. These analytical expressions for each distribution are presented herein.

**Exponential Distribution**

The PDF of the Exponential distribution, Equation (1.5), can be expressed as follows:

$$ f_{y_n}(y_n) = e^{-\frac{y_n}{\mu_y}}, \quad y_n \geq 0 $$

$$ (3.1) $$

The corresponding CDF, Equation (1.6), can be expressed as:
Rayleigh Distribution

The PDF of the Rayleigh distribution, Equation (2.1), can be expressed as follows:

\[ f_{Y_n}(y_n) = \frac{y_n}{\alpha^2} \exp\left(-\frac{1}{2} \left(\frac{y_n}{\alpha}\right)^2\right), \quad y_n \geq 0 \] (3.3)

The corresponding CDF, Equation (2.2), can be expressed as:

\[ F_{Y_n}(y_n) = 1 - e^{-\frac{1}{2} \left(\frac{y_n}{\alpha}\right)^2}, \quad \text{where} \quad \mu_{Y_n} = \sqrt{\frac{\pi}{2}} \alpha \] (3.4)

Solving Exponential or Rayleigh for Average of the 1/Nth Highest

The average of the 1/N\textsuperscript{th} highest peaks for the Exponential or Rayleigh distribution is solved as follows:

\[ \overline{H}_{\frac{1}{N}} = \frac{1}{N} \int H_{\frac{1}{N}} f_{Y_n}(y_n) dy_n = n \int y_n f_{Y_n}(y_n) dy_n \] (3.5)
The value of the $1/N$ highest peaks, $H_1$, is given by:

$$
\frac{1}{N} = \int_{H_1/N}^{\infty} f_{y_n}(y_n) dy_n = P\left( \frac{H_1}{N} \leq y_n \right) = 1 - F_{y_n} \left( \frac{H_1}{N} \right) \quad . (3.6)
$$

Note that:

$$
\int_0^\infty x^n e^{-px} dx = \frac{n!}{p^{n+1}} , \quad p>0 \text{ and } n \text{ is an integer} \quad . (3.7)
$$

and

$$
\int_0^\infty xe^{-px} dx = \frac{1}{p^2} - \frac{1}{p^2} e^{-mu} (1 + pu) , \quad u>0 \quad . (3.8)
$$

Thus,

$$
\int_{H_1/N}^\infty y_n f_{y_n}(y_n) dy_n = \int_{H_1/N}^\infty y_n e^{-\frac{y_n}{\mu_{y_n}}} dy_n = \mu_{y_n} \int_{\frac{H_1}{\mu_{y_n}}}^\infty ue^{-u} du
$$

$$
= \mu_{y_n} \left( 1 + \left( \frac{\frac{H_1}{\mu_{y_n}}}{\mu_{y_n}} \right) \right) e^{-\left( \frac{\frac{H_1}{\mu_{y_n}}}{\mu_{y_n}} \right)} = \frac{1}{N} \left( \mu_{y_n} + \frac{H_1}{N} \right) \quad . (3.9)
$$

The cut-off value for the Exponential Distribution can be found as follows:
\[
\frac{1}{N} = 1 - F_{\gamma_n} \left( H_{\frac{1}{N}} \right) = 1 - \left( 1 - e^{-\frac{H_{\frac{1}{N}}}{\mu_{\gamma_n}}} \right) = e^{-\frac{H_{\frac{1}{N}}}{\mu_{\gamma_n}}} \sim, \quad (3.10)
\]

or

\[
H_{\frac{1}{N}} = \mu_{\gamma_n} \ln N \quad . \quad (3.11)
\]

Thus, the average of the \(1/N\) highest values for the Exponential distribution is:

\[
\bar{H}_{\frac{1}{N}} = \frac{1}{N} \int_{H_{\frac{1}{N}}}^{\infty} y_n \left[ e^{-\frac{y_n}{\mu_{\gamma_n}}} \right] dy_n = \left( H_{\frac{1}{N}} + \mu_{\gamma_n} \right) \quad . \quad (3.12)
\]

\[
= \mu_{\gamma_n} \left( 1 + \ln N \right)
\]

Note, Equation (3.12) is equivalent to Equation (1.9) and would yield the same results as shown in Table 1.3.

Rayleigh Distribution

\[
\frac{1}{N} = \int_{H_{\frac{1}{N}}}^{\infty} f_{\gamma_n}(y_n) dy_n = P \left( H_{\frac{1}{N}} \leq y_n \right)
\]

\[
= 1 - F_{\gamma_n} \left( H_{\frac{1}{N}} \right) = 1 - \left( 1 - e^{-\frac{\left( \frac{H_{\frac{1}{N}}}{\alpha} \right)^2}{2}} \right) = e^{-\frac{\left( \frac{H_{\frac{1}{N}}}{\alpha} \right)^2}{2}} \quad . \quad (3.13)
\]

Therefore, the cut-off value of the Rayleigh distribution is:
Thus, the average of the $1/N^{th}$ highest values for the Rayleigh Distribution is:

\[
\bar{H} = \frac{1}{N} \int_{y_n}^{\infty} y_n f_{y_n}(y_n) \, dy_n.
\]

Where the following equation is required:

\[
\int_{x}^{\infty} x^2 e^{-q(x)} \, dx = -\frac{\sqrt{\pi} \left[ \text{erf}(qu) - 1 \right] - 2qu e^{-q^2}}{4q^3}.
\]

There is a mistake in Equation (37) on P. 272 (Jeffrey and Dai, 2008). Therefore, the average of the $1/N^{th}$ highest is shown in Equation (3.17) below:

\[
\bar{H} = \frac{n}{\alpha^2} \int_{H_1/N}^{\infty} Y_n^2 e^{-\frac{Y_n^2}{2\alpha}} \, dY_n = -\frac{n}{\alpha^2} \left[ \sqrt{\pi} \left( \frac{H_1}{N} \right) - \frac{H_1}{\sqrt{2\alpha}} \right] - \frac{2\sqrt{2} \sqrt{\ln N}}{\alpha}.
\]
The above equation, Equation (3.17) is equivalent to Equation (4.105) in (Lewandowski, 2004).

Gumbel Distribution

The value of $H_{\frac{1}{N}}$, the cut-off value, is given by solving the following equation:

$$\frac{1}{N} = 1 - F_{\gamma}(H_{\frac{1}{N}}) = 1 - \exp \left[ - \frac{\beta_{\gamma}\ln \gamma}{\alpha_{\gamma}} \right]$$

(3.18)

or

$$1 - \frac{1}{N} = \exp \left[ - \frac{\beta_{\gamma}\ln \gamma}{\alpha_{\gamma}} \right]$$

(3.19)

Thus,

$$\ln \left( 1 - \frac{1}{N} \right) = -e^{-\frac{\beta_{\gamma}\ln \gamma}{\alpha_{\gamma}}}$$

(3.20)

and

$$\ln \left[ - \ln \left( 1 - \frac{1}{N} \right) \right] = -\alpha_{n} \left( \frac{H_{1} - \mu_{n}}{\alpha_{n}} \right)$$

(3.21)

Therefore, the $1/N$th highest peak for a Gumbel distribution is:

$$H_{\frac{1}{N}} = \mu_{n} - \frac{\ln \left[ - \ln \left( 1 - \frac{1}{N} \right) \right]}{\alpha_{n}}$$

(3.22)
This cut-off value can be applied to a simulated data set as will be discussed further in Section 3.1.3.

Lognormal Distribution

The $1/N^{th}$ highest peak of a Lognormal distribution is calculated by the following equation:

$$
\frac{1}{N} = 1 - \frac{1}{2} \left[ 1 + erf \left( \frac{\ln H_1 - \lambda_Y}{N} \frac{1}{\sqrt{2\sigma^2_Y}} \right) \right]
$$

or

$$
\frac{\ln H_1 - \lambda_Y}{N} \frac{1}{\sqrt{2\sigma^2_Y}} = erf^{-1} \left( \frac{N - 2}{N} \right)
$$

Since the error function and its inverse can be approximated by:

$$
erf(x) \approx \left[ 1 - \exp \left( -x^2 \frac{\pi}{1 + ax^2} \right) \right]^{\frac{1}{2}}
$$

and

$$
\text{erf}^{-1}(x) \approx \left[ -\frac{2}{\pi a} \frac{\ln(1-x^2)}{2} + \sqrt{\left( \frac{2}{\pi a} \frac{\ln(1-x^2)}{2} \right)^2 - \frac{1}{a^2 \ln(1-x^2)}} \right]^{\frac{1}{2}}
$$
where \( a = 0.147 \) will give relative errors of \( 1.3 \times 10^{-4} \) and \( 2 \times 10^{-3} \) uniformly for \( \text{erf}(x) \) and \( \text{erf}^{-1}(x) \), respectively for all real \( x \geq 0 \).

Thus, the \( 1/N \) highest peak of a Lognormal distribution is:

\[
\ln H_{1} = \frac{\gamma_{a}}{N} = \ln \left( \frac{4N-4}{N^{2}} \right) + \sqrt{2\gamma_{a}} \left[ \frac{2}{\pi \cdot a} + \left( \frac{2}{\pi \cdot a} \right)^2 - 2 \ln \left( \frac{4N-4}{N^{2}} \right) \right]^{\frac{1}{2}}
\]

or

\[
H_{1} = \exp \left( \lambda_{a} + \sqrt{2\gamma_{a}} \left[ \frac{2}{\pi \cdot a} + \left( \frac{2}{\pi \cdot a} \right)^2 - 2 \ln \left( \frac{4N-4}{N^{2}} \right) \right]^{\frac{1}{2}} \right)
\]

This cut-off value can be applied to a simulated data set as will be discussed further in Section 3.1.3.

The analytical method was considered for the four distributions considered in this research program as a way to expand the Peak data set around limited information to determine required statistical parameters of the Peak data sets. The analytical method described to calculate the average of the \( 1/N \) highest values for the Exponential distribution and the Rayleigh distribution is fairly straightforward. However, as was shown in Section 2.3, the Peak data sets did not follow the Exponential or Rayleigh distributions. Regarding the Gumbel and Lognormal distributions, the cut-off value,
HI/N, can be calculated using the analytical method; however, in order to determine the average of the 1/Nth highest values for either the Gumbel or Lognormal distributions, the designer would need to simulate the data set, for example by using the Monte Carlo simulation technique.

3.1.3 Monte Carlo Simulation of the Peak Data

If the designer only has access to limited information about the data set, can additional statistical parameters be accurately estimated? This question was presented in Section 3.1.2 and the analytical method was considered as a means to fill in the missing data so that a designer with legacy data could expand the database. The second method considered to aid the designer is the Monte Carlo simulation method.

Monte Carlo Methods are a class of computational algorithms that rely on repeated sampling to simulate the random occurrence of the data set. Monte Carlo methods are often used to simulate the random occurrence of the data set. Monte Carlo methods are often used in simulating physical and mathematical systems and are useful for modeling phenomena with significant uncertainty, such as the case of vertical accelerations on planing hulls running at high speed in irregular waves. Based on the literature survey conducted for this research program, no prior work was discovered where the Monte Carlo method had been applied in simulating motions data for planing craft.
The Monte Carlo simulation technique presented in Haldar, Mahadevan (2000) is summarized below. All computers have the capability to generate uniformly distributed random numbers between 0 and 1. By specifying an arbitrary “seed value,” the computer will generate the required number of uniform random numbers, $u_i$, between 0 and 1, referred to as *pseudo* random numbers. By varying the seed value, different sets of random numbers can be generated. Once generated, the pseudo random numbers of uniform distribution must be transformed to random numbers with the appropriate characteristics and distribution of interest. This is achieved using a process commonly known as inverse transformation technique or inverse CDF method, where the CDF of the random variable, $F_X(x)$, is equated to the random number, $u_i$, as shown below:

$$F_X(x_i) = u_i$$  \hspace{1cm} (3.29)

This method solves for the random variable $x_i$ using the equation below:

$$x_i = F_X^{-1}(u_i)$$  \hspace{1cm} (3.30)

In this research, the Monte Carlo simulation method was explored to evaluate its ability to accurately simulate the Peak vertical acceleration data set using limited information about the experimental data. It was assumed that the designer had access to only two statistical parameters from the legacy data set - the average of the Peak data set and the standard deviation of the Peak data set.
The Monte Carlo routine in MATLAB® was used to generate random numbers for the Peak vertical acceleration dataset according to the four different assumed distributions considered in this research: the Exponential, Rayleigh, Lognormal, and Gumbel distributions. The MATLAB® coding is included in Appendix A. It is necessary to define each distribution uniquely by evaluating its parameters. In this research effort, the Method of Moments (Haldar and Mahadevan, 2000) is used to estimate each distribution’s parameters as was discussed in Section 2.3.3 and shown in Table 2.1.

The Peak data set was simulated using the Monte Carlo simulation method as outlined below:

Step 1: Assume that a limited amount of test data is given (average and standard deviation of Peak data set).

Step 2: Assume the Peak data set followed a specific distribution (Exponential, Rayleigh, Gumbel or Lognormal).

Step 3: Estimate the parameters for each distribution using the Method of Moments (Table 3.2).

Step 4: Create the simulated Peak data set by using the Monte Carlo command in Matlab to generate random numbers that follow the assumed distribution from Step 2 with the specified parameters from Step 3.

Once the simulated Peak data set has been created, the statistical parameters of the data set can be obtained by directly analyzing the data points in the simulated data set as was previously described.
The Monte Carlo simulation was run a minimum of 65,000 times to calculate the statistical parameters of each Peak data set. The simulated results were compared against the statistical parameters of the experimental Peak data for each of the 28 test cases. The results showed that for the twenty-eight test cases the Monte Carlo simulation method showed very good accuracy between the simulated results for the average of the $1/N^{th}$ highest Peak values and the experimental results for these values when the experimental data set was assumed to follow the Gumbel distribution and the average and standard deviation of the Peak data set were used to estimate the parameters. Thus, the Monte Carlo method can be used to accurately simulate the Peak data set and determine the statistical design criteria regarding the Peak vertical accelerations when only limited data is known about the data and the Gumbel distribution is used. The results are presented in Section 4.1.3.

Statistical Parameters for Parent Data Sets

As was described previously, the Peak data set is a subset of the Parent data set. There are trade-offs between using statistics from the Parent data set or the Peak data set. Statistics from the Parent data consider the entire exposure for a given amount of time that data were collected, which is one of the reasons the Parent data set is used for habitability criteria. Additionally, statistics gleaned from the Parent data set will not be affected by the peak identification methods and their shortcomings described in Section 2.3.1. Certainly, the magnitudes of the overall statistics of the Parent data set can be
quite low and would not be considered adequate to ensure that the structure or installed equipment can withstand a large slamming event when the planing hull impacts a wave at high speed.

3.2.1 Direct Method

One statistical parameter of the Parent data set that is used for habitability and personnel exposure is the RMS value. This can be calculated directly from the Parent data set as follows:

\[ RMS = \sqrt{\frac{\Sigma x_i^2}{n}} \]  \hspace{1cm} (3.31)

3.2.2 Monte Carlo Simulation of Parent Data

The author desired to further the investigation into the use of the Monte Carlo simulation technique discussed in Section 3.1.3 that showed very good accuracy in simulating the Peak data sets. In Section 3.1.3, the Monte Carlo simulation technique was used to simulate the Peak data set based on the average and standard deviation of the Peaks and the assumption that the data set followed the Gumbel distribution. This section will explore whether the Monte Carlo simulation technique could be used to recreate the Parent data set.

In many published reports on planing craft tests, the published data is limited to certain statistics, and the raw data are not included. In some cases, the legacy data reports
include the RMS of the Parent data set. Additionally, as was discussed in Section 2.4.1, the experimenter typically de-means the Parent data sets; thus, the Mean value of the Parent data set is zero. Therefore, the RMS of the experimental data set and the zero mean were used in the Monte Carlo distribution. Note, for very large data sets ($n>>0$), the standard deviation of a demeaned data set approaches the Root Mean Square (RMS) of the data set. The approach described in Section 3.1.3 for the Monte Carlo method will now be applied to simulate the Parent data set as follows:

Step 1: Assume that a limited amount of test data is given (Zero mean and RMS of the Parent data set).

Step 2: Assume the Parent data set followed the Gumbel distribution.

Step 3: Estimate the parameters for each distribution using the Method of Moments (Table 2.1).

Step 4: Create the simulated Parent data set by using the Monte Carlo method to generate random numbers that follow the assumed distribution from Step 2 with the specified parameters from Step 3.

Quantitative comparisons between the experimental Parent datasets and Monte Carlo simulated Parent data set were made by comparing the statistical parameters of the distribution, $\mu$ and $\alpha$, and by comparing the mean and RMS of the experimental and simulated data sets.

The results for the first four test cases are shown below in Tables 3.2 and 3.3. Note the Monte Carlo simulation was run over 65,000 times for each of the four test cases. Table
3.2 shows the comparison for the two parameters, \( u \) and \( \alpha \), and the Max \( D_n \) from the K-S Test. Table 3.3 shows the comparison of the mean and standard deviation values.

Table 3.2: Accuracy of Estimating Parent Data Set Using Monte Carlo Simulation

<table>
<thead>
<tr>
<th>Test Case</th>
<th>Experimental Dataset</th>
<th>Monte Carlo Dataset</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( u )</td>
<td>( \alpha )</td>
<td>( u )</td>
</tr>
<tr>
<td>Case 1</td>
<td>2.7317</td>
<td>-0.2113</td>
<td>2.727</td>
</tr>
<tr>
<td>Case 2</td>
<td>3.4504</td>
<td>-0.1673</td>
<td>3.441</td>
</tr>
<tr>
<td>Case 3</td>
<td>2.5582</td>
<td>-0.2256</td>
<td>2.559</td>
</tr>
<tr>
<td>Case 4</td>
<td>2.1371</td>
<td>-0.2701</td>
<td>2.147</td>
</tr>
</tbody>
</table>
Table 3.3: Comparison of Statistical Parameters of Parent Data Set

<table>
<thead>
<tr>
<th>Test Case</th>
<th>Experimental Dataset</th>
<th>Monte Carlo Dataset</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>RMS</td>
<td>Mean</td>
</tr>
<tr>
<td>Case 1</td>
<td>2.3E-15</td>
<td>0.4695</td>
<td>-3.46E-4</td>
</tr>
<tr>
<td>Case 2</td>
<td>-1.37E-15</td>
<td>0.3717</td>
<td>-1.85E-4</td>
</tr>
<tr>
<td>Case 3</td>
<td>7.86E-15</td>
<td>0.5013</td>
<td>1.4E-3</td>
</tr>
<tr>
<td>Case 4</td>
<td>-4.09E-15</td>
<td>0.6001</td>
<td>3.34E-4</td>
</tr>
</tbody>
</table>

Based on the high accuracy for the two parameters and the statistics and the very low level of maximum difference, $D_n$, reported in Tables 3.2 and 3.3, the author has successfully demonstrated the Monte Carlo method can be used to simulate the Parent Dataset based on the zero mean and the RMS, with a very high level of confidence. The significance of this finding is that for legacy data where the dataset is not available but the RMS is reported, the raw dataset can be recreated using the Monte Carlo method and the two-parameter Gumbel distribution based on the zero mean and the RMS value reported.
The author was interested in investigating how the Peak Identification methods could be applied to the Monte Carlo simulated Parent data set to extract the Peak data set from the simulated Parent data set as discussed in Section 2.3.1. For comparison purposes, the buffer method was used to build the Parent data set from both the experimental Parent data set and the simulated Parent data set. Statistics, including the mean, standard deviation and average of the $1/N^{th}$ statistics were calculated and compared for both the experimental and simulated Peak data sets.

The results of four test cases showed a wide range of variation in accuracy when comparing the statistics of the experimental Peak data set to the statistics of the simulated Peak data set. Upon further analysis, the reason for the wide variation in accuracy is due to the fact that the existing peak identification methods rely heavily on the time-dependent sequence of the Parent data set, where there is a trough that follows a peak, a requirement for a zero crossing, or a designation of time window. These are some of the known shortcomings of the existing peak identification methods as described in Section 2.3.1. The Monte Carlo Method that was implemented in this research effort can indeed simulate a dataset that has the same statistical characteristics as the experimental dataset in terms of the size and shape functions of the dataset with high accuracy for the mean, RMS, and Max $D_n$ as was shown in Tables 3.2 and 3.3 above. However, the Monte Carlo method generates random data points and does not replicate the time-dependent sequence relationship that exists within the experimental Parent dataset and is required if the existing peak identification methods are to be applied. Note, this characteristic of the Monte Carlo method does not affect its ability to accurately simulate the Peak data set as
was discussed in Section 3.1.3 because the Peak data set is resultant from the peak selection method and the time-dependency has been eliminated.

In the future, if a variation of the Monte Carlo Method is implemented that allows the random points in the simulated dataset to be generated in such a way that the transient or sequence relationship of the Parent data set is not lost then this approach may be applicable to help meet the near-term needs of the designer looking to extract Peak statistics from simulated Parent data sets based on legacy Parent data statistics. One possible approach is to model the Parent data set as a joined distribution in terms of both the magnitude of the vertical acceleration and its peak encounter frequency. This investigation was beyond the scope of this research but could be considered in the future. More than likely, however, given the shortcomings of peak identification methods, alternative statistics based solely on the Parent data set similar to the aerospace industry will likely be preferred for future design requirements. These are discussed in Section 3.4.

This marks the current limit to the use of the Monte Carlo method that was implemented to simulate vertical accelerations. The author has successfully demonstrated the Monte Carlo method can be used to simulate the Peak data set based on limited legacy data from the Peak data set and the two-parameter Gumbel distribution with a very high level of confidence. The author has also successfully demonstrated the Monte Carlo method can be used to simulate the Parent Dataset based on the zero mean and the RMS with a very
high level of confidence. Thus, the designer can implement the Monte Carlo simulation technique to extend the design database by using legacy data reported from previous test programs. Results of these studies are included in Section 4.2.2.

**Correlation Between Parent and Peak Parameters**

Chapter 2 investigated the underlying distribution for the Peak acceleration data sets and the Parent acceleration data sets in the 28 test cases examined. In addition to studying the statistical behavior of the Peak data sets and the Parent data sets independently, it is of interest to examine the correlation between these two data sets, as the Peak data set is a subset of the Parent data set. As was discussed in Chapter 1, certain planing craft design criteria are based on analysis of the Parent data set while other design criteria are based on analysis of the Peak data set. Legacy data reports may report out the RMS of the Parent data set, which is unaffected by the subjective Peak Identification Methods discussed in Section 2.3.1. For purposes of design and risk analysis of high-speed craft, it is critical to consider the tail or Peak values of the data set. However, data analysts typically do not report out the peak identification method or threshold values used in their analysis. As a result, a designer cannot be sure how the data were analyzed nor can a comparison be made between two data sets taken from the same test facility or between two different test facilities. It is of interest to investigate a relationship between the Parent and Peak data sets to discern what correlations or dependences exist that would benefit the designer. For the case of planing craft design, insight and knowledge of the correlations between these two data sets will help planing craft designers who must work within the current design requirements where statistics of both the Parent data set and the
Extreme value data set are necessary. It may also help designers better utilize legacy data where only specific statistics of the data set are reported out.

**Analytical Correlations**

For certain distribution functions, such as the Exponential or Rayleigh distributions, analytical relationships for the Parent and Peak data sets can be derived between the RMS and the average of the 1/N\textsuperscript{th} statistics.

The following relationships exist between the Parent and Peak data sets for the Exponential distribution:

\[ n = (1) \times RMS \]  
\[ n_{1/3} = (2.1) \times RMS \]  
\[ n_{1/10} = (3.3) \times RMS \]  
\[ n_{1/100} = (5.6) \times RMS \]

The following relationships exist between the Parent and Peak data sets for the Rayleigh distribution:

\[ n = (1.25) \times RMS \]
\[ \bar{n}_{1/3} = (2.0) \times RMS \]  (3.37)

\[ \bar{n}_{1/10} = (2.55) \times RMS \]  (3.38)

\[ \bar{n}_{1/100} = (3.34) \times RMS \]  (3.39)

There are no known analytical relationships between the Parent and Peak data sets for either the Lognormal or the Gumbel distributions.

**Empirical Correlations**

Some researchers, including Brown and Klosinski (1980) and Blount et al. (2006), suggested empirical relationships exist between the RMS of the Parent data set and the average of the 1/N\(^{th}\) statistics of the Peak data set based on results of their testing. Brown and Klosinski (1980) conducted tests on high length-to-beam planing craft in the range of test craft as shown in Table 3.4.
Table 3.4: Test Matrix for High L/B Ratio Planing Craft (Brown and Klosinski, 1980)

<table>
<thead>
<tr>
<th></th>
<th>L/B = 5</th>
<th>L/B = 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_v$</td>
<td>2.66, 4.00, 5.32</td>
<td>2.66, 4.00, 5.32</td>
</tr>
<tr>
<td>$V_k/L^{1/2}$</td>
<td>4, 6, 8</td>
<td>3.38, 5.07, 6.76</td>
</tr>
<tr>
<td>$C_\Delta$</td>
<td>0.60, 0.72</td>
<td>0.73, 0.96, 1.20</td>
</tr>
<tr>
<td>$LCG/B$</td>
<td>1.77, 2.06, 2.22</td>
<td>2.00, 2.29, 2.79</td>
</tr>
<tr>
<td>$H_{1/3}/B$</td>
<td>0.44, 0.67</td>
<td>0.22, 0.44, 0.67</td>
</tr>
</tbody>
</table>

In Table 3.4 above, $C_v$ is the speed coefficient, $\sqrt{\frac{v}{\sqrt{gb^2}}}$, $v$ is boat speed feet per second (fps), $g$ is acceleration due to gravity 32.2 fps$^2$, $B$ is the boat’s chine beam, $L$ is the boat length (feet), $C_\Delta$ is the load coefficient, $\frac{\Delta}{wB^3}$, $w$ is the specific weight of fresh water. $LCG$ is the longitudinal position of the center of gravity measured from the transom, and $H_{1/3}$ is the significant wave height (feet).


A summary of the analytical and published empirical relationships is shown in Table 3.5.
Table 3.5: Existing Analytical and Published Empirical Relationships

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{n}{\text{RMS}} )</td>
<td>1.0</td>
<td>1.25</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>( \frac{n_{10}}{\text{RMS}} )</td>
<td>2.1</td>
<td>2.0</td>
<td>NA</td>
<td>1.1</td>
</tr>
<tr>
<td>( \frac{n_{100}}{\text{RMS}} )</td>
<td>3.3</td>
<td>2.55</td>
<td>6.0</td>
<td>2.2</td>
</tr>
<tr>
<td>( \frac{n_{1000}}{\text{RMS}} )</td>
<td>5.6</td>
<td>3.34</td>
<td>NA</td>
<td>4.2</td>
</tr>
</tbody>
</table>

For the 28 test cases included in this research program, the author conducted a linear regression analysis to investigate any empirically-derived relationships between the RMS of the Parent data set and the average of \( 1/N^{\text{th}} \) statistics of the Peak data set and analyzed the results for trends. The results were also compared against previously reported relationships. For the sake of comparison and to determine if any sensitivity existed, both the buffer method and the horizontal threshold method were used to identify the Peak data sets. These two methods were selected as their results for peak selection showed to be the most robust against threshold sensitivity based on the studies described in Section 2.3.
For each regression, the correlation factor, $R$ was computed. The correlation factor, $R$, measures the strength and the direction of a linear relationship between two variables and is computed as shown below:

$$
R = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n \sum x^2 - (\sum x)^2} \cdot \sqrt{n \sum y^2 - (\sum y)^2}}
$$

In general, a correlation factor, $R$, which has a value greater than 0.8 is considered to be a strong correlation, whereas a value less than 0.5 is considered to be a weak correlation. Additionally, the square of this term, $R^2$, is known as the Coefficient of Determination. The Coefficient of Determination gives the proportion of the variance between two variables and determines the certainty in making the prediction. A value of $R^2$ close to 1.0 indicates that the predictability of the regression is quite high. Based on this investigation into correlations between statistical parameters of the Parent and Peak data sets of the 28 test cases examined, there was a very strong linear relationship between the RMS of the Parent data set and the average of the $1/N^{th}$ statistics of the Peak data set. These relationships can be considered to extend legacy test reports where only limited data is available. Furthermore, these relationships can be considered to estimate the Peak data set statistical parameters required for design based only on the Parent data set parameters. The results, including graphs and charts, are presented in Section 4.3.
Supplementary Statistical Parameters of Parent Data

In addition to the empirical relationships between the Parent and Peak data sets discussed in Chapter 3.3, it is of interest to consider other statistical values of the Parent data sets that are currently used in the design process for other industries as were examined by Zseleczky and McKee (1989). The authors suggest that the design community consider moving away from subjective Peak Identification methods and instead to put focus on gleaning more information from the Parent data set. They propose reporting additional information in test reports to provide more insight into the statistical behavior of vertical accelerations data. They are in general support of a probabilistic approach to the problem, while recognizing that the current design approach still requires reporting of the 1/Nth statistics of the Peak data set for structural design methods and the RMS of the Parent data set for personnel exposure and habitability criteria. Additional statistics are discussed below.

Moments

Researchers (Zseleczky and McKee, 1989) suggest that future test reports include the first four moments of the distribution of the Parent data set – the Mean (M₁), Variance (M₂), Skewness (M₃) and Kurtosis (M₄) – in order to build a better understanding of the distribution of the Parent data and to begin to formulate multi-parameter families similar to the approach taken to characterize the wave spectrum data in oceanographic studies.
The moments of a distribution can be gained from the following equation, where \( k = 1, 2, 3, \) or 4 in the case of Mean, Variance, Skewness, or Kurtosis respectively:

\[
\frac{1}{n} \sum_{i=1}^{n} X_i^k .
\]  \hspace{1cm} (3.41)

The first four moments for the first four test cases are shown in Table 3.6 below. The results of the remaining test cases are included in Appendix B.

<table>
<thead>
<tr>
<th>Table 3.6: First Four Moments of Parent Data Set for Four Test Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>M2</td>
</tr>
<tr>
<td>M3</td>
</tr>
<tr>
<td>M4</td>
</tr>
</tbody>
</table>

**Probability of Exceedance**

As an alternative to using the Peak Identification methods to extract the Peak data sets from the Parent data sets and then calculate the average of the \( 1/N^{th} \) statistics, future test reports could include the Probability of Exceedance values at 10\%, 1\% and 0.2\% based on the Parent data sets directly. These levels represent the observed value that is exceeded by no more than \( P\% \) of the collected data. These are no more than a selected...
subset of the distribution function for the observed response. The difference between these values and the current design approach of using the average of the $1/N$th Peak values is that the P% values are more relevant to the issue of characterizing the tail of the response distribution, i.e. the Peak or Extreme Values, without confusing or misrepresenting the process with subjectivity introduced through different peak identification methods and threshold values selections. As noted in the literature, experimentally obtained probability levels have been used in aircraft work to assess the impact of stress from wind gusts and landing loads on the life expectancy of the aircraft. This approach could be considered for planing craft design (Zseleczky and McKee, 1989).

For the data sets considered herein, the 10%, 1% and 0.2% probability values for the experimental Parent data set will be examined by organizing the data in ascending order and the value selected based on the desired probability level and the number’s rank in the order. To further investigate the utility and accuracy of the Monte Carlo method, the Probability of Exceedance values can be obtained using the Parent data set simulated using the Monte Carlo simulation and assuming the Gumbel distribution. As is the procedure for the experimental Parent data set, the simulated Parent data set is organized in ascending order and then the value selected based on the desired probability level and the number’s rank in the order. Additionally, the Probability of Exceedance value can be obtained analytically using the equation below and the Gumbel distribution equation for cut-off value, $H_{1/N}$, discussed in Section 3.1.2.
The results of the Probability of Exceedance values for the three data sets considered are shown in Table 3.7. In all three test cases evaluated, the Probability of Exceedance comparisons between the Monte Carlo simulation and the Gumbel analytical calculation were over 99% accurate. This is expected as the analytical equation is derived based upon the Gumbel distribution. In comparing the Probability of Exceedance values between the experimental data and the Monte Carlo simulation, the Monte Carlo simulation was at least 85% accurate for PoE(10%), at least 70% accurate for PoE(1%), and at least 85% accurate for PoE(0.2%). The PoE values for the remaining data sets are presented in Section 4.4. The purpose of including these values is to give the experimenter and designer information on the test cases considered for benefit in future research.

Table 3.7: Comparison of Probability of Exceedance Levels for Three Test Cases (Experimental; M.C Simulation; Gumbel Analytical)

<table>
<thead>
<tr>
<th></th>
<th>Test Case 1</th>
<th></th>
<th>Test Case 2</th>
<th></th>
<th>Test Case 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>PoE: 10%</td>
<td>0.577</td>
<td>0.6088</td>
<td>0.6125</td>
<td>0.4223</td>
<td>0.4832</td>
<td>0.4849</td>
</tr>
<tr>
<td>PoE: 1%</td>
<td>1.2515</td>
<td>1.4596</td>
<td>1.4727</td>
<td>0.9064</td>
<td>1.1722</td>
<td>1.1659</td>
</tr>
<tr>
<td>PoE: 0.2%</td>
<td>1.8537</td>
<td>2.0758</td>
<td>2.0633</td>
<td>1.91</td>
<td>1.6221</td>
<td>1.6336</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4. Results and Discussion

The research discussed herein examined the statistical behavior and distribution of vertical accelerations data for the design of planing craft. This study examined 28 data sets collected during full-scale and model scale testing of different hulls. The data sets consist of vertical accelerations measured at the Longitudinal Center of Gravity (LCG) of planing craft operating at planing speed in random, head sea waves. These test cases represented a range of planing hull forms, volumetric Froude numbers in the fully planing regime, and significant wave heights as discussed in Section 1.6. For each test case, the statistical distribution and statistical parameters of the Parent data sets and the Peak data sets were examined. As the Peak data sets were obtained by applying a peak identification method to the Parent data set, sensitivity studies were carried out to consider four different peak identification methods and a range of threshold value.

Qualitative and quantitative comparisons are made between the experimental data and known distribution functions including the Exponential, Rayleigh, Gumbel, and Lognormal distribution functions for the Peak data sets. For studies in the Parent data sets, three of the four distributions are included, but the Lognormal distribution is excluded, as it is not applicable to negative data or data with a zero mean.

Explicit and Monte Carlo simulation methods are implemented to determine statistical design parameters and compare against experimental data. Correlations between Parent
and Peak data sets are investigated. Other statistical parameters based only on the Parent data sets are also presented for consideration in future research.

Peak Acceleration Studies

The results of the statistical studies on the Parent data sets are presented herein.

Statistical Distribution of Peak Data Sets

For each of the 28 test cases the Peak data sets were extracted from the Parent data sets using four peak identification methods and a range of threshold values as described in Section 2.3.3. In the examination of all of these Peak vertical acceleration data sets collected from tests conducted on different planing hulls operating in irregular waves, the Exponential distribution did not fit any of the data sets.

Sensitivity studies were carried out to consider four different peak identification methods and a range of threshold or buffer values. In all cases, the Exponential distribution proved to be an inappropriate fit to the experimental data.

Thus, the earlier findings of Fridsma (1971) that were adopted by Savitsky and Brown (1976) and Hoggard and Jones (1980) and some of the classification societies regarding the validity of the Exponential distribution for Peak acceleration data have been disproved.
The results of the analysis showed that the Rayleigh distribution was also a poor fit to Peak acceleration data sets. This finding was in agreement with Fridsma (1971) and Brown and Klosinski (1980) regarding the nonlinearity of vertical accelerations in relation to wave height.

The Gumbel distribution was a very good fit for all peak identification methods and was the least sensitive peak identification methods or threshold/buffer values. Overall, the Gumbel distribution was the best fit to the full-scale test cases. The Gumbel distribution was also the best fit in the model-scale test cases.

The Lognormal distribution showed sensitivity to threshold values for each of the peak identification methods. The Lognormal distribution performed slightly better than the Gumbel distribution in two of the full-scale test cases using the buffer method but by less than 0.002 difference between the maximum $D_n$ value for the two distributions. In these two test cases, the Exponential distribution had a maximum difference of six times the maximum difference of the either the Gumbel or Lognormal distribution.

A summary of the results are presented herein in Figures 4.1 to 4.4 for the Peak data sets of the full-scale test cases using two different peak identification methods, the buffer method and the HT method. The results for the other two peak identification methods
showed the same trends — the Exponential distribution had the worst fit to the data sets and Gumbel distribution had the best fit to the data sets. The results for the Peak data sets of the model-scale test cases are shown in Figures 4.5 to 4.8.

Figure 4.1: Goodness of Fit for Full-Scale Peak Data Sets Using Buffer Method
Figure 4.2: Trendline Results for Full-Scale Peak Data Sets Using Buffer Method

Figure 4.3: Goodness of Fit for Full-Scale Peak Data Sets Using HT Method
Figure 4.4: Trendline Results for Full-Scale Peak Value Data Sets Using HT Method

Figure 4.5: Goodness of Fit for Model-Scale Peak Value Data Sets, Buffer Method
Figure 4.6: Goodness of Fit for Model-Scale Peak Value Data Sets Using HT Method

Figure 4.7: Trendline Results for Model-Scale Peak Value Data Sets, Buffer Method
The Monte Carlo simulation was used to estimate statistical parameters of the Peak data sets for each of the four distribution functions and compare against the full-scale and model-scale experimental data. This was performed for each of the four peak identification methods and for a range of threshold values. In all cases, the Exponential and Rayleigh estimates had wide ranges of error when comparing the mean, standard deviation, and averages of the one-third, one-tenth and one-hundredth values of the Peak Value data sets. The Gumbel distribution showed very high accuracy for mean, standard deviation, and average of the $1/N$ statistics predictions. Figures 4.9 to 4.12 show the
results of the Monte Carlo simulation of Peak value statistics from the full-scale test cases using the HT method for peak identification. The other peak identification methods had very similar results, as did the model-scale test cases. The calculated errors for the full-scale test cases using the Gumbel distribution are shown in Table 4.1. The calculated errors for the model-scale test cases using the Gumbel distribution are shown in Table 4.2.

Figure 4.9: Error in Mean Value for Monte Carlo Simulation of Full-Scale Peak Data
Figure 4.10: Error in Std Dev Value for Monte Carlo Simulation of Full-Scale Peak Data

Figure 4.11: Error in Avg 1/10th Value for Monte Carlo Simulation, Full-Scale Peak Data
Figure 4.12: Error in Avg one-hundredth Value, Monte Carlo Simulation, Full-Scale Peak Data
Table 4.1: Calculated Error Comparison of Statistical Parameters of Experimental and Simulated Peak Data Sets for Full Scale Test Cases Using the Gumbel Distribution

<table>
<thead>
<tr>
<th>test case</th>
<th>mean</th>
<th>std dev</th>
<th>avg 1/3rd</th>
<th>avg one-tenth</th>
<th>avg one-hundredth</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0%</td>
<td>0%</td>
<td>-3%</td>
<td>-4%</td>
<td>7%</td>
</tr>
<tr>
<td>2</td>
<td>0%</td>
<td>0%</td>
<td>-3%</td>
<td>5%</td>
<td>16%</td>
</tr>
<tr>
<td>3</td>
<td>0%</td>
<td>0%</td>
<td>-4%</td>
<td>2%</td>
<td>16%</td>
</tr>
<tr>
<td>4</td>
<td>0%</td>
<td>0%</td>
<td>-3%</td>
<td>5%</td>
<td>8%</td>
</tr>
<tr>
<td>5</td>
<td>0%</td>
<td>0%</td>
<td>1%</td>
<td>-2%</td>
<td>-12%</td>
</tr>
<tr>
<td>6</td>
<td>0%</td>
<td>0%</td>
<td>-5%</td>
<td>-4%</td>
<td>17%</td>
</tr>
<tr>
<td>7</td>
<td>0%</td>
<td>0%</td>
<td>-2%</td>
<td>3%</td>
<td>10%</td>
</tr>
<tr>
<td>8</td>
<td>0%</td>
<td>0%</td>
<td>-1%</td>
<td>4%</td>
<td>3%</td>
</tr>
<tr>
<td>9</td>
<td>0%</td>
<td>0%</td>
<td>-2%</td>
<td>3%</td>
<td>9%</td>
</tr>
<tr>
<td>10</td>
<td>0%</td>
<td>0%</td>
<td>-4%</td>
<td>2%</td>
<td>13%</td>
</tr>
<tr>
<td>11</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>1%</td>
<td>-2%</td>
</tr>
<tr>
<td>12</td>
<td>0%</td>
<td>0%</td>
<td>-1%</td>
<td>0%</td>
<td>1%</td>
</tr>
<tr>
<td>13</td>
<td>0%</td>
<td>0%</td>
<td>-1%</td>
<td>0%</td>
<td>-8%</td>
</tr>
<tr>
<td>14</td>
<td>0%</td>
<td>0%</td>
<td>-2%</td>
<td>4%</td>
<td>9%</td>
</tr>
<tr>
<td>15</td>
<td>0%</td>
<td>0%</td>
<td>-2%</td>
<td>3%</td>
<td>9%</td>
</tr>
<tr>
<td>16</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>2%</td>
<td>-3%</td>
</tr>
<tr>
<td>17</td>
<td>0%</td>
<td>0%</td>
<td>-1%</td>
<td>1%</td>
<td>3%</td>
</tr>
<tr>
<td>18</td>
<td>0%</td>
<td>0%</td>
<td>-6%</td>
<td>4%</td>
<td>17%</td>
</tr>
<tr>
<td>19</td>
<td>0%</td>
<td>0%</td>
<td>-2%</td>
<td>-7%</td>
<td>-2%</td>
</tr>
</tbody>
</table>

Table 4.2: Calculated Error Comparison of Statistical Parameters of Experimental and Simulated Peak Data Sets for Model Scale Test Cases Using the Gumbel Distribution

<table>
<thead>
<tr>
<th>test case</th>
<th>mean</th>
<th>std dev</th>
<th>avg 1/3rd</th>
<th>avg one-tenth</th>
<th>avg one-hundredth</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0%</td>
<td>0%</td>
<td>-14%</td>
<td>2%</td>
<td>21%</td>
</tr>
<tr>
<td>2</td>
<td>0%</td>
<td>0%</td>
<td>-6%</td>
<td>6%</td>
<td>16%</td>
</tr>
<tr>
<td>3</td>
<td>0%</td>
<td>0%</td>
<td>-4%</td>
<td>8%</td>
<td>11%</td>
</tr>
<tr>
<td>4</td>
<td>0%</td>
<td>0%</td>
<td>-8%</td>
<td>6%</td>
<td>14%</td>
</tr>
<tr>
<td>5</td>
<td>0%</td>
<td>0%</td>
<td>-7%</td>
<td>6%</td>
<td>18%</td>
</tr>
<tr>
<td>6</td>
<td>0%</td>
<td>0%</td>
<td>-10%</td>
<td>3%</td>
<td>22%</td>
</tr>
<tr>
<td>7</td>
<td>0%</td>
<td>0%</td>
<td>-4%</td>
<td>-3%</td>
<td>-18%</td>
</tr>
<tr>
<td>8</td>
<td>0%</td>
<td>0%</td>
<td>-5%</td>
<td>-2%</td>
<td>-5%</td>
</tr>
<tr>
<td>9</td>
<td>0%</td>
<td>0%</td>
<td>-7%</td>
<td>1%</td>
<td>15%</td>
</tr>
</tbody>
</table>
Parent Acceleration Studies

The results of the statistical studies on the Parent data sets are presented herein.

Statistical Distribution of Parent Data Sets

This research study examined the statistical distribution of the Parent data sets. Based on the literature survey conducted, the statistical distribution of the Parent vertical acceleration data had not been considered previously. The results of this research show that in all cases for both full-scale and model-scale testing, the Exponential and Rayleigh distributions were the worst fit to the Parent data sets. The Gumbel was the best fit to the Parent data sets with at least 95% confidence. The Lognormal distribution is not applicable for data that can have a value of less than one or where the mean is zero. The results of the distribution study for the Parent data sets are shown in Figures 4.13 and 4.14 for the full-scale test cases and Figures 4.15 and 4.16 for the model-scale test cases.
Figure 4.13: Goodness of Fit Results for Full-Scale Parent Data Sets

Figure 4.14: Trendline Results for Full-Scale Parent Data Sets
Figure 4.15: Goodness of Fit Results for Model-Scale Parent Data Sets

Figure 4.16: Trendline Results for Model-Scale Parent Data Sets
Statistical Parameters for Parent Data Sets

The Monte Carlo method was used to simulate the Parent data sets. This approach was carried out using the Root Mean Square (RMS) and the Zero Mean of the Parent data set. This information is typically reported out in legacy data reports even if the raw data is not available. The results of the Monte Carlo method showed that using the Monte Carlo method and assuming a Gumbel distribution was highly accurate in generating the Parent data sets for these test cases. The results of this analysis are shown in Table 4.3 which depicts the error in the mean and RMS for the full-scale test cases and Table 4.4 depicts the error in mean and RMS for the model-scale test cases.

Table 4.3: Error Using Monte Carlo Simulation of Full-Scale Parent Data Sets

<table>
<thead>
<tr>
<th></th>
<th>Exponential Distribution</th>
<th>Rayleigh Distribution</th>
<th>Gumbel Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>RMS</td>
<td>Mean</td>
</tr>
<tr>
<td>Test Case 1</td>
<td>43%</td>
<td>-43%</td>
<td>12%</td>
</tr>
<tr>
<td>Test Case 2</td>
<td>45%</td>
<td>-45%</td>
<td>13%</td>
</tr>
<tr>
<td>Test Case 3</td>
<td>36%</td>
<td>-36%</td>
<td>6%</td>
</tr>
<tr>
<td>Test Case 4</td>
<td>23%</td>
<td>-23%</td>
<td>-9%</td>
</tr>
<tr>
<td>Test Case 5</td>
<td>-100%</td>
<td>100%</td>
<td>200%</td>
</tr>
<tr>
<td>Test Case 6</td>
<td>60%</td>
<td>-60%</td>
<td>26%</td>
</tr>
<tr>
<td>Test Case 7</td>
<td>-142%</td>
<td>58%</td>
<td>607%</td>
</tr>
<tr>
<td>Test Case 8</td>
<td>-100%</td>
<td>100%</td>
<td>200%</td>
</tr>
<tr>
<td>Test Case 9</td>
<td>-100%</td>
<td>100%</td>
<td>200%</td>
</tr>
<tr>
<td>Test Case 10</td>
<td>-100%</td>
<td>100%</td>
<td>200%</td>
</tr>
<tr>
<td>Test Case 11</td>
<td>-100%</td>
<td>100%</td>
<td>200%</td>
</tr>
<tr>
<td>Test Case 12</td>
<td>45%</td>
<td>-45%</td>
<td>14%</td>
</tr>
<tr>
<td>Test Case 13</td>
<td>42%</td>
<td>-42%</td>
<td>11%</td>
</tr>
<tr>
<td>Test Case 14</td>
<td>50%</td>
<td>-50%</td>
<td>18%</td>
</tr>
<tr>
<td>Test Case 15</td>
<td>59%</td>
<td>-59%</td>
<td>25%</td>
</tr>
<tr>
<td>Test Case 16</td>
<td>31%</td>
<td>-31%</td>
<td>-1%</td>
</tr>
<tr>
<td>Test Case 17</td>
<td>32%</td>
<td>-32%</td>
<td>0%</td>
</tr>
<tr>
<td>Test Case 18</td>
<td>-128%</td>
<td>72%</td>
<td>822%</td>
</tr>
<tr>
<td>Test Case 19</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>
Table 4.4: Error Using Monte Carlo Simulation of Model-Scale Parent Data Sets

<table>
<thead>
<tr>
<th>Test Case</th>
<th>Exponential Distribution</th>
<th>Rayleigh Distribution</th>
<th>Gumbel Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>RMS</td>
<td>Mean</td>
</tr>
<tr>
<td>Test Case 1</td>
<td>5.2%</td>
<td>-5.1%</td>
<td>-36.3%</td>
</tr>
<tr>
<td>Test Case 2</td>
<td>-100.5%</td>
<td>99.8%</td>
<td>200.6%</td>
</tr>
<tr>
<td>Test Case 3</td>
<td>-101.8%</td>
<td>98.0%</td>
<td>197.5%</td>
</tr>
<tr>
<td>Test Case 4</td>
<td>-116.3%</td>
<td>83.5%</td>
<td>130.7%</td>
</tr>
<tr>
<td>Test Case 5</td>
<td>-109.9%</td>
<td>89.9%</td>
<td>203.9%</td>
</tr>
<tr>
<td>Test Case 6</td>
<td>-107.3%</td>
<td>92.6%</td>
<td>271.5%</td>
</tr>
<tr>
<td>Test Case 7</td>
<td>-97.0%</td>
<td>97.2%</td>
<td>-658.5%</td>
</tr>
<tr>
<td>Test Case 8</td>
<td>-101.3%</td>
<td>98.8%</td>
<td>163.7%</td>
</tr>
<tr>
<td>Test Case 9</td>
<td>-101.4%</td>
<td>98.8%</td>
<td>160.7%</td>
</tr>
</tbody>
</table>

Supplementary Statistical Parameters of Parent Data Set

New statistics based only on the Parent data sets, such as the first four Statistical Moments, or the Probability of Exceedance were calculated for each of the 28 test cases. A table of the first four moments for the Parent data sets is included in Appendix B. The Probability of Exceedance values, similar to those used in the aerospace industry may in the future help the designer in test planning, data analysis and design of new planning craft and are discussed herein.

The experimenter could use either the Monte Carlo method or the Gumbel analytical equation for future analyses. The Probability of Exceedance values of the experimental Parent data are compared to the Probability of Exceedance values of the Monte Carlo simulated data set assuming the Gumbel distribution and estimating the parameters based on the Zero mean and the RMS. The results are shown for the full-scale Parent data sets in
the Zero mean and the RMS. The results are shown for the full-scale Parent data sets in Table 4.5 with an average error of 8% with a maximum error of 33% for PoE (10%), an average error of 6% with a maximum error of 23% for PoE (1%), and an average error of 3% with a maximum error of 27% for PoE (0.2%). For the model-scale data sets shown in Table 4.6 the Gumbel analytical method had an average error of 26% with a maximum error of 37% for PoE (10%), an average error of 9% with a maximum error of 20% for PoE (1%), and an average error of 26% with a maximum error of 42% for PoE (0.2%).

Table 4.5: Comparison of Full-Scale Experimental and Gumbel Analytical Cut-off Values for Given Probability of Exceedance (PoE) Levels; values shown in g’s

<table>
<thead>
<tr>
<th>Test Case</th>
<th>Cut-off Value for 10% PoE</th>
<th>Cut-off Value for 1% PoE</th>
<th>Cut-off Value for 0.2% PoE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Experimental</td>
<td>M.C., Gumbel</td>
<td>% Error</td>
</tr>
<tr>
<td>1</td>
<td>1.76</td>
<td>1.80</td>
<td>-2%</td>
</tr>
<tr>
<td>2</td>
<td>1.39</td>
<td>1.46</td>
<td>-4%</td>
</tr>
<tr>
<td>3</td>
<td>1.66</td>
<td>1.73</td>
<td>-5%</td>
</tr>
<tr>
<td>4</td>
<td>1.47</td>
<td>1.75</td>
<td>-19%</td>
</tr>
<tr>
<td>5</td>
<td>0.68</td>
<td>0.90</td>
<td>-33%</td>
</tr>
<tr>
<td>6</td>
<td>1.41</td>
<td>1.42</td>
<td>-1%</td>
</tr>
<tr>
<td>7</td>
<td>6.76</td>
<td>7.33</td>
<td>-8%</td>
</tr>
<tr>
<td>8</td>
<td>0.79</td>
<td>0.89</td>
<td>-13%</td>
</tr>
<tr>
<td>9</td>
<td>0.83</td>
<td>0.89</td>
<td>-7%</td>
</tr>
<tr>
<td>10</td>
<td>0.67</td>
<td>0.76</td>
<td>-14%</td>
</tr>
<tr>
<td>11</td>
<td>0.67</td>
<td>0.71</td>
<td>-7%</td>
</tr>
<tr>
<td>12</td>
<td>1.61</td>
<td>1.63</td>
<td>-1%</td>
</tr>
<tr>
<td>13</td>
<td>1.63</td>
<td>1.66</td>
<td>-2%</td>
</tr>
<tr>
<td>14</td>
<td>1.44</td>
<td>1.53</td>
<td>-6%</td>
</tr>
<tr>
<td>15</td>
<td>1.36</td>
<td>1.44</td>
<td>-6%</td>
</tr>
<tr>
<td>16</td>
<td>1.61</td>
<td>1.65</td>
<td>-3%</td>
</tr>
<tr>
<td>17</td>
<td>1.62</td>
<td>1.64</td>
<td>-1%</td>
</tr>
<tr>
<td>18</td>
<td>1.57</td>
<td>1.90</td>
<td>-21%</td>
</tr>
<tr>
<td>19</td>
<td>1.37</td>
<td>1.38</td>
<td>-1%</td>
</tr>
</tbody>
</table>
Table 4.6: Comparison of Model-Scale Experimental and Gumbel Analytical Cut-off Values for Given Probability of Exceedance Levels; values shown in g's

<table>
<thead>
<tr>
<th>Test Case</th>
<th>Cut-off Value for 10% PoE</th>
<th>Cut-off Value for 1% PoE</th>
<th>Cut-off Value for 0.2% PoE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Experimental</td>
<td>M.C., Gumbel</td>
<td>% Error</td>
</tr>
<tr>
<td>1</td>
<td>0.56</td>
<td>0.75</td>
<td>-34%</td>
</tr>
<tr>
<td>2</td>
<td>0.74</td>
<td>0.96</td>
<td>-31%</td>
</tr>
<tr>
<td>3</td>
<td>1.29</td>
<td>1.77</td>
<td>-37%</td>
</tr>
<tr>
<td>4</td>
<td>0.42</td>
<td>0.51</td>
<td>-21%</td>
</tr>
<tr>
<td>5</td>
<td>0.55</td>
<td>0.74</td>
<td>-35%</td>
</tr>
<tr>
<td>6</td>
<td>1.17</td>
<td>1.54</td>
<td>-31%</td>
</tr>
<tr>
<td>7</td>
<td>0.53</td>
<td>0.59</td>
<td>-11%</td>
</tr>
<tr>
<td>8</td>
<td>0.53</td>
<td>0.60</td>
<td>-13%</td>
</tr>
<tr>
<td>9</td>
<td>0.86</td>
<td>1.02</td>
<td>-18%</td>
</tr>
</tbody>
</table>

**Correlation Between Parameters of Peak and Parent Data**

Correlation between RMS of the Parent data set and average of 1/N<sup>th</sup> statistics of the Peak value data set was examined for each of the 28 data sets. This study was carried out using two different Peak Identification methods for comparison. The results presented in Figures 4.17 to 4.20 show that there is a strong linear relationship between the statistical parameters of each Parent data set and Peak data set, both of which follow the Gumbel distribution.
Figure 4.17: Correlation Between Average of Peak Data Set and RMS of Parent Data Set

Figure 4.18: Correlation Between Average of 1/3rd Highest Peak Data Set and RMS of Parent Data Set
Figure 4.19: Correlation Between Average of one-tenth Highest Peak Data Set and RMS of Parent Data Set

Figure 4.20: Correlation Between Average of one-hundredth Highest Peak Data Set and RMS of Parent Data Set
Based on these findings, empirically derived linear relationships between the Parent data statistical parameters and the Peak data statistical parameters are proposed in Table 4.2, along with the square of the Correlation Coefficient, $R$, and the Coefficient of Determination, $R^2$, of this equation to the twenty-eight cases considered. These relationships are provided as an aide to the designer working with the current range of design requirements, which include Parent and Peak value statistics and the troublesome peak identification methods. These relationships can also be used to extend legacy test reports, where only limited information about the data is presented and it is unknown how the peak data sets were determined (what method and threshold were used).

Additionally, these relationships may benefit the designer when evaluating various design requirements such as the average of the $1/N^{th}$ values for structural design criteria and RMS for human exposure criteria.

Table 4.7: Empirically Derived Relationships Between Parent and Peak Statistics

<table>
<thead>
<tr>
<th>Peak Data Statistic</th>
<th>Relationship to Parent Data</th>
<th>Correlation Coefficient, $R$</th>
<th>Coefficient of Determination, $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>$2.39 * RMS - 0.19$</td>
<td>0.96</td>
<td>0.92</td>
</tr>
<tr>
<td>Average of $1/3^{rd}$ Highest</td>
<td>$4.58 * RMS - 0.40$</td>
<td>0.97</td>
<td>0.93</td>
</tr>
<tr>
<td>Average of one-tenth Highest</td>
<td>$7.11 * RMS - 0.67$</td>
<td>0.96</td>
<td>0.94</td>
</tr>
<tr>
<td>Average of one-hundredth Highest</td>
<td>$9.58 * RMS - 0.77$</td>
<td>0.93</td>
<td>0.87</td>
</tr>
</tbody>
</table>
Discussion

The results of this comprehensive analysis of vertical accelerations from a range of test craft, test speeds, and sea conditions conclude that the Exponential distribution is not appropriate for vertical accelerations of planing craft operating in irregular head seas. This conclusion was based on the K-S Test and proved true regardless of the peak identification method or the threshold value sensitivity. These findings disprove the previously accepted method based on the work of Fridsma (1971) and Savitsky and Brown (1976), and extended to Hoggard and Jones (1980), that stated the Exponential distribution should be used for Peak vertical acceleration data of planing craft. Based on the results of the analysis conducted herein, the author concludes that for modern planing craft, the best fit of distribution for the Peak data sets is the Gumbel distribution, part of the General Peak Value Distribution Family.

The author’s initial focus for this research was in investigating the correct method of distribution for peak values of vertical accelerations. Once this was determined to be the Gumbel distribution, it was of interest to analyze the Parent data sets to understand the underlying distribution of the entire data signal as a planing craft operates at speed in a seaway. Based on the results of the K-S Test conducted herein, the author concludes that for modern planing craft, the best fit of distribution for the Parent data sets is the Gumbel distribution.
The results of this study show that the Gumbel distribution is the most robust fit for Parent and Peak data sets over a range of hull forms, speeds, and sea conditions. Peak identification method, while troublesome and subjective, did not influence the best-fit distribution of the data sets based on the results of the sensitivity studies. Scale effects that may occur when testing either at full scale or model scale did not appear to influence the result away from the Gumbel distribution. The Gumbel distribution was also the best fit in the test case of a novel, non-planing, high-speed hull form operating in irregular seas; thus, this distribution may extend to non-traditional high speed hullforms. This could open up new research areas for future studies.

This research effort examined the ability to simulate the data sets using the Monte Carlo simulation method. The Monte Carlo method was used to simulate with high accuracy the Parent data sets using only the Root Mean Square (RMS) of the data set, mean value of zero, and the Gumbel distribution function. The significance of this finding is that for legacy data where the data set is not available but the RMS is reported, the raw data set can be recreated using the Monte Carlo method and the two-parameter Gumbel distribution and estimating the two parameters based on the zero mean and the RMS value reported by using the Method of Moments.

The Monte Carlo simulation method also showed good correlation for predicting statistical values of the Peak data sets, such as the average of the $1/N^{th}$ highest value,
using the mean and standard deviation of the Peak data set and the two-parameter Gumbel distribution by using the Method of Moments.

This research effort examined the relationship between the statistical parameters of the Parent and Peak value data sets. A linear regression analysis of the relationship between the average of the $1/N^{th}$ highest values of the Peak data set and the RMS of the Parent data set showed good correlation based on $R$ and $R^2$ coefficients. Formulas for predicting the average of the $1/N^{th}$ highest Peak data based on the RMS of the Parent data are included as a guide for planing craft designers that are designing craft to meet the current design standards for average of the $1/N^{th}$ values for structural design criteria and RMS for human exposure criteria.

New statistical values based on the Parent data sets, including the Probability of Exceedance as well as the first four statistical moments, $M_1$ through $M_4$, are included for consideration in future research. The Probability of Exceedance levels are well known in the area of aerospace design and should be further investigated for its use in the design of planing craft structures. By including the first four statistical moments, the Mean, Standard Deviation, Skewness, and Kurtosis, for the twenty-eight test cases examined in this research, it is possible to continue to add to this database if future experimenters include these values in their data reporting. The comparison of the PoE levels of the experimental Parent data sets to the PoE levels of the Monte Carlo simulated Parent data
sets showed very good accuracy; thus, the Monte Carlo method could be used to estimate the PoE values of legacy data where limited information is published.
CHAPTER 5

5. Conclusions

This concludes the research effort to investigate the statistical behavior of planing craft vertical accelerations at speed in irregular head seas. The research considered the distribution, statistical parameters and trends of Parent data and Peak data for 28 data sets collected from different model-scale and full-scale tests of modern planing craft operating at speed in irregular head seas. Sensitivity studies were carried out into the subjective nature of Peak Identification Methods and selected threshold values. This research investigated the use and accuracy of the Monte Carlo simulation method to recreate data sets when only limited statistical information is known. Empirical correlations between Parent and Peak data sets are included as an aid to planing craft designers, based on the 28 data sets included in this research program. Suggestions for future test programs to include examination of other statistics based only on the Parent data sets are also included. Over time, these parameters may shed new light on the statistical behavior of the vertical accelerations of planing craft and thus influence new approaches for structural design and human exposure levels.

Benefits

This research expands the knowledge base of planing craft accelerations to better equip designers in early stage design and to better equip experimenters in data analysis and reporting.
1. The Statistical distribution of Peak data set has been clarified. The Exponential distribution is not appropriate for vertical accelerations of planing craft either the Parent data sets or Peak Value data sets. This contradicts the previously accepted recommendations for planing craft (Fridsma, 1971), (Savitsky and Brown, 1980) and (Hoggard and Brown, 1985).

2. The vertical accelerations of planing craft did not follow the Rayleigh distribution, thus validating Fridsma’s suggestion that vertical acceleration behaves nonlinearly with respect to the sea condition (Fridsma, 1969) and (Fridsma, 1971).

3. For the 28 data sets examined, the Gumbel Distribution was the closest fit to the Peak data and was the least sensitive to Peak Identification method or threshold value.

4. The Gumbel distribution was the closest fit to the Parent data sets, which had not been considered in previous research.

5. The Monte Carlo method was used to accurately simulate the Peak data sets using the Gumbel distribution and input parameters of the average and standard deviation of the Peak values.

6. The Monte Carlo method was used to accurately simulate the Parent data sets using the Gumbel distribution and two parameters – the zero mean and the RMS.

7. Correlations between the RMS of the Parent data sets and the average of the 1/Nth values of the Peak data sets are presented as a guide to planing craft designers and a way to unlock legacy data.

8. The researcher recommends future data analysis and reporting of the Parent data for vertical accelerations to include Probability of Exceedance levels. Suggested levels include 10%, 1% and 0.2%.

9. The researcher includes tables of the first four statistical moments in the Appendix for future researchers to access as further gains are made in understanding the statistical behavior of planing craft.
Concluding Remarks

This was a very worthwhile research endeavor for the author, and it is hoped that this research will inspire renewed interest in advanced methodologies for predicting motions and loads on high performance planing craft both for recreational and military applications. Higher fidelity of understanding for craft motions and impact loads is fundamental in order to advance the design community in areas of hull form development and use of more sophisticated materials, such as composites technologies.

Future Work

Several areas of research are discussed below as areas to extend this research program.

The author suggests extending this methodology to other dynamic-lift hull forms beyond the planing monohull form considered herein. A single test case included in this research program indicates that this methodology does in fact extend to other high-speed hull forms. This should be further quantified.

In this research program, only head sea conditions were considered as this typically yields the highest vertical accelerations for design purposes. However, it may be of interest to the designer to use this same methodology in the investigation of other headings to the waves, such as bow quarter or stern sea conditions.
In addition to vertical accelerations, another critical aspect in the design of planing craft is the understanding and prediction of the impact pressure loading on the hull bottom structure. The author suggests extending this methodology to impact pressure data studies.

This methodology can likely be adapted to and support data analysis and validation of numerical time-domain tools for seakeeping studies of high-speed craft such as planing monohulls and novel hull forms.

The Probability of Exceedance levels are well known in the area of aerospace design and should be further investigated for its use in the design and reliability of planing craft structures.

This research effort focused on the vertical acceleration as a random variable. An extension of this research effort could be to implement the knowledge gained about the statistical behavior of this variable and to initiate a study into a 2-variable relationship that includes the average vertical acceleration and the mean recurring period. Based on studies included in this research effort, these two random variables are statistically independent. Knowledge gained in researching this relationship could be implemented with the Monte Carlo method to improve the accuracy of the simulated Parent data set.
LITERATURE CITED


APPENDIX A: MATLAB CODING

A.1. Parent Analysis
A.2. Peak Analysis with Buffer
A.3. Peak Analysis with Vertical Threshold
A.4. Peak Analysis with Vertical Difference
A.5. Peak Analysis with Horizontal Threshold
MATLAB Code

A.1. Parent Analysis

clear all
clc
clf
% format long
load BoatN.txt
Dr2=BoatN;
Data_Len=length(test)
A_mean=mean(test)
A_std=std(test)
test_sort=sort(test);

%%% find the average of 1/3, 1/10, 1/100, 0.2/100 highest and their cut values
H_nth=[ 3 10 100 500 ];
for jj=1:length(H_nth)
    if length(test)<H_nth(jj)
        Raw_mean(jj)=0;
        Raw_cut(jj)=0;
    else
        Raw_mean(jj)=mean(test_sort(ceil((1-1./H_nth(jj))*length(test)):length(test)));
        Raw_cut(jj)=test_sort(ceil((1-1./H_nth(jj))*length(test)));
    end
end

%%% PDF Plot for the Original data
% Parameters for Lognormal distribution
stdll=sqrt(log(1+(A_std/A_mean)^2));
meanll=log(abs(A_mean))-stdll^2/2;
% Parameters for Rayleigh distribution
alphal=(A_mean/sqrt(pi/2)+A_std/sqrt(2-pi/2))/2;
% Parameters for Exponential distribution
lamdal=(1/A_mean+1/A_std)/2;
% Parameters for Gumbel distribution
all=pi/A_std/sqrt(6);
mul=A_mean-0.5772/all;
% Generate the value according to input
xl=linspace(min(test),max(test));
% Lognormal pdf
y11=lognpdf(xl,meanll,stdll);
% Rayleigh pdf
y12=raylpdf(xl,alphal);
% Exponential pdf
y13=lamdal*exp(-lamdal*xl);
% Gumbel distribution
y14=all*exp(-all*(xl-mul)).*exp(-exp(-all*(xl-mul)));
% Normal pdf
y15=normpdf(xl,A_mean,A_std);
% CDF information for different distribution
y24=exp(-exp(-all*(xl-mul)));
%

figure(1)
ksdensity(test)
hold on
```matlab
plot(xl,yl2,'-bo','markersize',4)
hold on
plot(xl,yl3,'-ko','markersize',4)
hold on
hold on
plot(xl,yl1,'-bo','markersize',4)
plot(xl,yl4,'-ro','markersize',4)
legend('Observed','Rayleigh','Exponential','Lognormal','Gumbel')
title('PDF comparison for the Original data')
figure(2)
ksdensity(test)
hold on
plot(xl,yl1,'-bo','markersize',4)
hold on
plot(xl,yl4,'-ro','markersize',4)
legend('Observed','Lognormal','Gumbel')
title('PDF comparison for the Original data')
figure(3)
ecdf(test)
hold on
plot(xl,yll,'-bo','markersize',4)
hold on
plot(xl,yl4,'-ro','markersize',4)
legend('Observed','Gumbel cdf')
title('CDF comparison for original Data')

%% KS test for the original data
% cdf of lognormal distribution according to the inputdata
y31=logncdf(sort(test),meanll,stdll);
% cdf of Rayleigh distribution according to the inputdata
y32=raylcdf(sort(test),alpha1);
% cdf of Exponential distribution according to the inputdata
y33=expcdf(sort(test),lamdal);
% cdf of gumbel distribution according to the inputdata
y34=exp(-exp(-all*(sort(test)-mul)));
for i=1:length(test)
    E_cdf(i)=i/length(test);
end
Dn_L_0(1)=max(abs(y31-E_cdf'));
Dn_R_0(1)=max(abs(y32-E_cdf'));
Dn_E_0(1)=max(abs(y33-E_cdf'));
Dn_G_0(1)=max(abs(y34-E_cdf'));
% maximum different value of CDF for 1/3, 1/10, one-hundredth highest
for i=1:length(H_nth)
    Dn_L_0(i+1)=max(abs(y31(ceil((1-1/H_nth(i))*length(test))):end-
                        E_cdf(ceil((1-1/H_nth(i))*length(test))):end)')
    Dn_R_0(i+1)=max(abs(y32(ceil((1-1/H_nth(i))*length(test))):end-
                        E_cdf(ceil((1-1/H_nth(i))*length(test))):end)')
    Dn_E_0(i+1)=max(abs(y33(ceil((1-1/H_nth(i))*length(test))):end-
                        E_cdf(ceil((1-1/H_nth(i))*length(test))):end)')
    Dn_G_0(i+1)=max(abs(y34(ceil((1-1/H_nth(i))*length(test))):end-
                        E_cdf(ceil((1-1/H_nth(i))*length(test))):end)')
end
Dn=[Dn_E_0;Dn_R_0;Dn_L_0;Dn_G_0]

%% Monte Carlo simulation
samples=1000000;
% Parameters for Lognormal distribution
X_log=exp(erfinv(2*rand(samples,1)-1)*sqrt(2)*stdll+meanll);
% X_log=exp(erfinv(2*rand(samples,1)-1)*2*stdll+meanll);
Xsort_log=sort(X_log);
```
\[
\begin{align*}
\text{Mean}_{\log} &= \text{mean}(X_{\log}) \\
\text{std}_{\log} &= \text{std}(X_{\log}) \\
% Parameters for Rayleigh distribution 
\alpha_{\text{al}} &= (\text{mean}_{\text{l}}/\sqrt{\pi/2} + \text{std}_{\text{l}}/\sqrt{2-\pi/2})/2; \\
X_{\text{ray}} &= \sqrt{-2\log(1-\text{rand}(\text{samples}, 1))} \times \alpha_{\text{al}}; \\
X_{\text{sort\_ray}} &= \text{sort}(X_{\text{ray}}); \\
\text{Mean}_{\text{ray}} &= \text{mean}(X_{\text{ray}}); \\
\text{std}_{\text{ray}} &= \text{std}(X_{\text{ray}}); \\
% Parameters for Exponential distribution 
X_{\text{exp}} &= -\log(1-\text{rand}(\text{samples}, 1))/\lambda_{\text{l}}; \\
X_{\text{sort\_exp}} &= \text{sort}(X_{\text{exp}}); \\
\text{Mean}_{\text{exp}} &= \text{mean}(X_{\text{exp}}); \\
\text{std}_{\text{exp}} &= \text{std}(X_{\text{exp}}); \\
% Parameters for Gumbel distribution 
X_{\text{gum}} &= -\log(-\log(\text{rand}(\text{samples}, 1)))/\alpha_{\text{l}} + \mu_{\text{l}}; \\
X_{\text{sort\_gum}} &= \text{sort}(X_{\text{gum}}); \\
\text{Mean}_{\text{gum}} &= \text{mean}(X_{\text{gum}}) \\
\text{std}_{\text{gum}} &= \text{std}(X_{\text{gum}}) \\
\text{for } i = 1: \text{length}(H_{\text{nth}}) \\
\quad H_{\text{log\_i}} &= \exp(\text{erfinv}((H_{\text{nth\_i}}-2)/H_{\text{nth\_i}}) \times \sqrt{2} \times \text{std}_{\text{l}} + \text{mean}_{\text{l}}); \\
\quad H_{\text{log\_direct\_i}} &= \text{Xsort\_log}(\text{ceil}((1-1/H_{\text{nth\_i}})\times\text{samples})); \\
\quad \text{aaa} &= 0.147; \\
\quad \alpha_{\text{l}} &= 2/\pi/\text{aaa}; \\
\quad \alpha_{\text{aa2}} &= \log(4 \times (H_{\text{nth\_i}}-1)/H_{\text{nth\_i}})^2); \\
\quad H_{\text{log\_theo\_i}} &= \exp(\text{mean}_{\text{l}} + \text{std}_{\text{l}} \times \sqrt{-\alpha_{\text{l}} - \alpha_{\text{aa2}}/2 + \sqrt{(-\alpha_{\text{l}} + \alpha_{\text{aa2}}/2)^2 - 1/\text{aaa} \times \alpha_{\text{aa2}}}}); \\
\quad X_{\text{sum\_log}} &= 0; \\
\quad \text{for } j = 1: \text{samples} \\
\quad \quad \text{if } X_{\text{log\_j}} \geq H_{\text{log\_i}} \\
\quad \quad \quad X_{\text{sum\_log}} &= X_{\text{sum\_log}} + X_{\text{log\_j}}; \\
\quad \quad \text{end} \\
\quad \text{end} \\
\quad H_{\text{avg\_log\_i}} &= H_{\text{nth\_i}} \times X_{\text{sum\_log}}/\text{samples}; \\
\quad H_{\text{avg\_log\_direct\_i}} &= \text{mean}(X_{\text{sort\_log}}(\text{ceil}((1-1/H_{\text{nth\_i}})\times\text{samples})); \\
% Rayleigh distribution 
\quad H_{\text{ray\_i}} &= \alpha_{\text{al}} \times \sqrt{2 \times \log(H_{\text{nth\_i}})}; \\
\quad H_{\text{ray\_direct\_i}} &= \text{Xsort\_ray}(\text{ceil}((1-1/H_{\text{nth\_i}})\times\text{samples})); \\
% \text{num\_ray} &= \text{sum}(X_{\text{ray}} \geq H_{\text{ray\_i}}); \\
\quad X_{\text{sum\_ray}} &= 0; \\
\quad \text{for } j = 1: \text{samples} \\
\quad \quad \text{if } X_{\text{ray\_j}} \geq H_{\text{ray\_i}} \\
\quad \quad \quad X_{\text{sum\_ray}} &= X_{\text{sum\_ray}} + X_{\text{ray\_j}}; \\
\quad \quad \text{end} \\
\quad \text{end} \\
\quad H_{\text{avg\_ray\_i}} &= H_{\text{nth\_i}} \times X_{\text{sum\_ray}}/\text{samples}; \\
\quad H_{\text{avg\_ray\_theo\_i}} &= \text{H_{\text{nth\_i}}} \times \alpha_{\text{al}} \times \sqrt{2} \times \log(H_{\text{nth\_i}}) - \frac{\sqrt{\text{pi}}}{2} \cdot \text{erf} \left( \log(H_{\text{nth\_i}}) \right) - 1 - \frac{\sqrt{\text{pi}}}{2} \cdot \text{erf} \left( \log(H_{\text{nth\_i}}) \right) \times \log(H_{\text{nth\_i}}))/H_{\text{nth\_i}}; \\
\quad H_{\text{avg\_ray\_direct\_i}} &= \text{mean}(X_{\text{sort\_ray}}(\text{ceil}((1-1/H_{\text{nth\_i}})\times\text{samples})); \\
% Exponential distribution 
\quad H_{\text{exp\_i}} &= 1/\lambda_{\text{l}} \times \log(H_{\text{nth\_i}}); \\
\quad H_{\text{exp\_direct\_i}} &= \text{Xsort\_log}(\text{ceil}((1-1/H_{\text{nth\_i}})\times\text{samples})); \\
% \text{num\_exp} &= \text{sum}(X_{\text{exp}} \geq H_{\text{exp\_i}}); \\
\quad X_{\text{sum\_exp}} &= 0; \\
\quad \text{for } j = 1: \text{samples} \\n\end{align*}
\]
if X_exp(j) >= Hn_exp(i)
    X_sum_exp = X_sum_exp + X_exp(j);
end

H_avg_exp(i) = H_nth(i) * X_sum_exp / samples;
H_avg_exp_direct(i) = mean(Xsort_exp(ceil(1 - 1/H_nth(i) * samples):end));
H_avg_exp_theo(i) = 1/lamda1 * (1 + log(H_nth(i))); % Gumbel distribution
Hn_gum(i) = -log(-log(l - 1/H_nth(i)))/all + mul;
Hn_gum_direct(i) = Xsort_log(ceil((1 - 1/H_nth(i)) * samples));
% num_gum = sum(X_gum >= Hn_gum);
X_sum_gum = 0;
for j = 1:samples
    if X_gum(j) >= Hn_gum(i)
        X_sum_gum = X_sum_gum + X_gum(j);
    end
end
H_avg_gum(i) = H_nth(i) * X_sum_gum / samples;
H_avg_gum_direct(i) = mean(Xsort_gum(ceil(1 - 1/H_nth(i)) * samples):end));
end
H_avg_gum
MATLAB Code
A.2. Peak Analysis
(Includes Selection of Peaks Using Buffer Method)

clear all
clc
clf
% format long
load Data403_0
load BoatA.txt
test=BoatA;
% load BoatB.txt
% test=BoatB
Data_Len=length(test)
A_mean=mean(test);
test=test-A_mean;
multipler=1.;
A_rms=0;
for i=1:Data_Len
    A_rms=A_rms+(test(i))^2;
end
A_rms=sqrt(A_rms/Data_Len);
%buffer
A_buffer=multipler*A_rms;
figure(1)
plot(test)
hold on
plot([1 Data_Len],[0 0])
% delete the consucessive equivalent data
% start=1;
% temp=999999;
% j=1;
% for i=start:Data_Len
%    if temp<--test(i)
%        Data_new(j)=test(i);
%        j=j+1;
%        temp=test(i);
%    end
% end
% figure(2)
% plot(Data_new)
% hold on
% plot([1 length(Data_new)],0 0)
start1=1;
start2=1;
pcount=1;% count the number of the peaks
tcount=1;% count the number of the troughs
for istart=start1:length(test)
    localmax=-999999;
    localmin=999999;
goback1=0;
    for i=start1:length(test)
if localmin>test(i)
    localmin=test(i);
    min_index=i;
else if localmin<test(i)
    break
end
for j=1:length(test)
    if (test(j)-localmin)>=A_buffer
        Trough_value(tcount)=localmin;
        %Trough_index(tcount)=min_index;
        Trough_index(tcount)=i-1;
        start2=j
        tcount=tcount+1;
        goback1=0;
        break
    end
end
if localmin> test(j)
% localmin=test(j)
% min_index=j;
% i=j;
    start1=j;
    goback1=1;
    break
end
if goback1==1
    continue;
end
% make sure find the trough values then go to the next step, otherwise
% go % back to the for loop to find the trough values again
if start1>length(test)-1
    if tcount==1
        disp('The troughs cannot be found, please decrease the buffer
value')
        break
    end
    break
end
goback2=1;
%goback2 should go back to find the new minimum
for ii = j: length(test)
    if goback2==0
        break;
    end
end
for mm=j:length(test)
    if localmax<test(mm)
        localmax=test(mm);
        max_index=mm;
    elseif localmax>test(mm)
        break
    end
end
for kk=mm:length(test)
    if (-test(kk)+localmax)>=A_buffer
        Peak_value(pcount)=localmax;
Peak_index(pcount)=mm-l+\min(\text{find(test(mm-1:end))==localmax)})-1;
Peak_index(pcount)=mm-l;
pcount=pcount+1;
goback2=0;
j=kk;
break
end
if localmax<test(kk)
j=kk;
goback2=1;
break
end
startl=kk;
if j>\text{length(test)}-1
if tcount==l
\text{disp('The peaks cannot be found, please decrease the buffer value')}
break
end
break
end
end
T_V=Trough_value(2:end);
T_I=Trough_index(2:end);
if length(T_I)<length(Peak_index)
Peak_value=Peak_value(1:end-1);
Peak_index=Peak_index(1:end-1);
end
Peak_Len=\text{length}(Peak_value)
% Cycle=Peak_value-T_V;
% [Cycle_max Cycle_index]=max(Cycle);
% Cycle_index=Trough_index+Cycle_index-1;
Peak_sort=\text{sort}(Peak_value);
[P_max P_max_index]=max(Peak_value)
P_max_index=Peak_index(P_max_index)
for i=1:length(Peak_index)-1
Cycle(i)=Peak_index(i+1)-Peak_index(i);
end
C_Pk_Cy=\text{corrcoef}(Peak_value(2:end),Cycle)
min(Cycle)
%% find the average of 1/3, 1/10, 1/100, 0.2/100 highest peaks and their cut values
H_nth=[3 10 100 500];
for jj=1:length(H_nth)
if length(Peak_value)<H_nth(jj)
PC_mean(jj)=0;
PC_cut(jj)=0;
else
PC_mean(jj)=\text{mean}(\text{Peak_sort(ceil((1-1./H_nth(jj))*length(Peak_value):length(Peak_value))))};
PC_cut(jj)=\text{Peak_sort(ceil((1-1./H_nth(jj))*length(Peak_value))))};
end
end
figure(2)
for i=1:length(T_I)
    plot([Peak_index(i) T_I(i)],[Peak_value(i) T_V(i)])
end
hold on

% Parameters for BoatADT
meanl=mean(Peak_value) % mean value
stdl=std(Peak_value) % stand deviation
M2_pk=mean((Peak_value-meanl).^2)
M3_pk=mean((Peak_value-meanl).^3)
M4_pk=mean((Peak_value-meanl).^4)
% Parameters for Lognormal distribution
stdll=sqrt(log(1+(stdl/meanl)^2));
meanll=log(meanl)-stdll^2/2;
% Parameters for Rayleigh distribution
alphal=(meanl/sqrt(pi/2)+stdl/sqrt(2-pi/2))/2;
% Parameters for Exponential distribution
lamlal=(1/meanl+1/stdl)/2;
% Parameters for Gumbel distribution
all=pi/stdl/sqrt(6);
mul=meanl-0.5772/all;
% Generate the value according to input
xl=linspace(min(Peak_value),max(Peak_value));
% Lognormal pdf
yll=lognpdf(xl,meanll,stdll);
% Rayleigh pdf
yrl=raylpdf(xl,alphal);
% Exponential pdf
y1l=lamlal*exp(-lamlal*xl);
% Gumbel distribution
y4l=all*exp(-all*(xl-mul)).*exp(-exp(-all*(xl-mul)));
% Normal pdf
y5l=normpdf(xl,meanl,stdl);
% CDF information for different distribution
y24l=exp(-exp(-all*(xl-mul)));

% Parameters for BoatADT
figure(4)
ksdensity(Peak_value)
hold on
plot(xl,yll,'-ro','markersize',4)
legend('Observed','Lognormal pdf')
title('BoatA Buffer')
figure(5)
ksdensity(Peak_value)
hold on
plot(xl,yll,'-ro','markersize',4)
legend('Observed','Rayleigh pdf')
title('BoatA Buffer')
figure(6)
ksdensity(Peak_value)
hold on
plot(xl,yll,'-ro','markersize',4)
legend('Observed','Exponential pdf')
title('BoatA Buffer')
figure(7)
ksdensity(Peak_value)
plot(x1,y24,'-ro','markersize' , 4)
legend('Observed', 'Lognormal', 'Rayleigh', 'Exponential', 'Gumbel cdf')
title('BoatA Buffer CDF comparison')

%% KS test
% cdf of lognormal distribution according to the inputdata
y31=logncdf(sort(Peak_value),mean11,stdll);
% cdf of Rayleigh distribution according to the inputdata
y32=raylcdf(sort(Peak_value),alphal);
% cdf of Exponential distribution according to the inputdata
y33=expcdf(sort(Peak_value),lamdal);
% cdf of gumbel distribution according to the inputdata
y34=exp(-exp(-all*(sort(Peak_value)-mul)))
for i=1:length(Peak_value)
    E_cdf(i)=i/length(Peak_value);
end
Dn_L(l)=max(abs(y31-E_cdf));
Dn_R(l)=max(abs(y32-E_cdf));
Dn_E(l)=max(abs(y33-E_cdf));
Dn_G(l)=max(abs(y34-E_cdf));
% maximum different value of CDF for 1/3, 1/10, one-hundredth highest
for i=1:length(H_nth)
    Dn_L(i+1)=max(abs(y31(ceil((1-1/H_nth(i))*length(Peak_value)))-
                     E_cdf(ceil((1-1/H_nth(i))*length(Peak_value))));
    Dn_R(i+1)=max(abs(y32(ceil((1-1/H_nth(i))*length(Peak_value)))-
                     E_cdf(ceil((1-1/H_nth(i))*length(Peak_value))));
    Dn_E(i+1)=max(abs(y33(ceil((1-1/H_nth(i))*length(Peak_value)))-
                     E_cdf(ceil((1-1/H_nth(i))*length(Peak_value))));
    Dn_G(i+1)=max(abs(y34(ceil((1-1/H_nth(i))*length(Peak_value)))-
                     E_cdf(ceil((1-1/H_nth(i))*length(Peak_value))));
end

%% monte calo simulation
samples=1000000;
% H_nth=[3 10 100];
% Parameters for Lognormal distribution
X__log=exp(erfinv(2*rand(samples,1)-1)*sqrt(2)*stdll+meanll);
% Parameters for Rayleigh distribution
alphal=(meanl/sqrt(pi/2)+stdl/sqrt(2-pi/2))
X_ray=sqrt(-2*log(1-rand(samples,1)))*alphal;
% Parameters for Exponential distribution
X_exp=-log(1-rand(samples,1))/lamdal;
% Parameters for Gumbel distribution
X_gum=-log(-log(rand(samples,1)))/all+mul;
for i=1:length(H_nth)
    Hn_log(i)=exp(erfinv((H_nth(i)-2)/H_nth(i))*sqrt(2)*stdll+meanll);
\[ H_{\text{log\_direct\_tho}}(i) = \exp(\text{mean} + \sqrt{2} \cdot \text{std} \cdot \sqrt{-aa_1 - aa_2/2 + \sqrt{(aa_1 + aa_2/2)^2 - 1/aaa \cdot aa_2}}) \]

\[ X_{\text{sum\_log}} = 0; \]
\[ \text{for } j = 1: \text{samples} \]
\[ \text{if } X_{\text{log}}(j) \geq H_{\text{log\_direct\_tho}}(i) \]
\[ X_{\text{sum\_log}} = X_{\text{sum\_log}} + X_{\text{log}}(j); \]
\[ \text{end} \]
\[ \text{end} \]
\[ H_{\text{avg\_log\_direct\_tho}}(i) = \frac{H_{\text{nth}}(i) \cdot X_{\text{sum\_log}}}{\text{samples}}; \]

\[ H_{\text{log\_direct\_tho}}(i) = \text{mean}(X_{\text{sort\_log}}(\text{ceil}((1 - 1/H_{\text{nth}}(i))^2 \cdot \text{samples})); \]

\[ H_{\text{avg\_log\_direct\_tho}}(i) = \text{mean}(X_{\text{sort\_log}}(\text{ceil}((1 - 1/H_{\text{nth}}(i))^2 \cdot \text{samples}))); \]

\% Rayleigh distribution
\[ H_{\text{ray\_direct}}(i) = \text{mean}(X_{\text{ray}}(\text{ceil}((1 - 1/H_{\text{nth}}(i))^2 \cdot \text{samples})); \]
\[ \text{for } j = 1: \text{samples} \]
\[ \text{if } X_{\text{ray}}(j) \geq H_{\text{ray\_direct}}(i) \]
\[ X_{\text{sum\_ray}} = X_{\text{sum\_ray}} + X_{\text{ray}}(j); \]
\[ \text{end} \]
\[ \text{end} \]
\[ H_{\text{avg\_ray\_direct\_tho}}(i) = \text{mean}(X_{\text{sort\_ray}}(\text{ceil}((1 - 1/H_{\text{nth}}(i))^2 \cdot \text{samples}))); \]

\% Exponential distribution
\[ H_{\text{exp\_direct}}(i) = \text{mean}(X_{\text{exp}}(\text{ceil}((1 - 1/H_{\text{nth}}(i))^2 \cdot \text{samples})); \]
\[ \text{for } j = 1: \text{samples} \]
\[ \text{if } X_{\text{exp}}(j) \geq H_{\text{exp\_direct}}(i) \]
\[ X_{\text{sum\_exp}} = X_{\text{sum\_exp}} + X_{\text{exp}}(j); \]
\[ \text{end} \]
\[ \text{end} \]
\[ H_{\text{avg\_exp\_direct\_tho}}(i) = \text{mean}(X_{\text{sort\_exp}}(\text{ceil}((1 - 1/H_{\text{nth}}(i))^2 \cdot \text{samples}))); \]

\% Gumbel distribution
\[ H_{\text{gum\_direct}}(i) = \text{mean}(X_{\text{sort\_log}}(\text{ceil}((1 - 1/H_{\text{nth}}(i))^2 \cdot \text{samples})); \]
\[ \text{for } j = 1: \text{samples} \]
\[ \text{if } X_{\text{gum}}(j) \geq H_{\text{gum\_direct}}(i) \]
\[ X_{\text{sum\_gum}} = X_{\text{sum\_gum}} + X_{\text{gum}}(j); \]
\[ \text{end} \]
\[ \text{end} \]
\[ H_{\text{avg\_gum\_direct\_tho}}(i) = \text{mean}(X_{\text{sort\_gum}}(\text{ceil}((1 - 1/H_{\text{nth}}(i))^2 \cdot \text{samples})); \]
H_avg_gum_direct(i) = mean(Xsort_gum(ceil((1-
1/H_nth(i)) * samples):end));
end
%% histogram and frequency diagram plot
NN = round(1+3.3*log10(length(Peak_value)));
binwidth = (max(Peak_value) - min(Peak_value)) / NN;
[nhist histbin] = hist(Peak_value, 4*NN);
figure(15)
bar(histbin, nhist/length(Peak_value)/binwidth)
Areacheck = sum(nhist/length(Peak_value)/binwidth)*binwidth
hold on
plot(xl, yl2, '-bo', 'markersize', 4)
hold on
plot(xl, yl3, '-ko', 'markersize', 4)
hold on
plot(xl, yl4, '-r', 'lineWidth', 2.5)
legend('Hist', 'Lognormal', 'Rayleigh', 'Exponential', 'Gumbel pdf'
title('Peaks value selected by Buffer Method(BufferSize=2rms)');
figure(16)
hist(Peak_value, NN)
% different method to generate bidwidth
binwidth1 = 2*iqr(Peak_value)/length(Peak_value)^(1/3)
bins1 = min(Peak_value):binwidth1:max(Peak_value);
[nhist1 histbins1] = hist(Peak_value, bins1);
figure(17)
bar(histbins1, nhist1/length(Peak_value)/binwidth1)
Areacheck1 = sum(nhist1/length(Peak_value)/binwidth1)*binwidth1
hold on
plot(xl, lognpdf(xl, meanll, stdll), '-go', 'markersize', 4)
hold on
plot(xl, yl2, '-bo', 'markersize', 4)
hold on
plot(xl, yl3, '-ko', 'markersize', 4)
hold on
plot(xl, yl4, '-r', 'lineWidth', 2.5)
legend('Hist', 'Lognormal', 'Rayleigh', 'Exponential', 'Gumbel pdf'
title('Peaks value selected by Buffer Method(BufferSize=2rms)');
%% Chi-square test
% interval determined using 75% percentage
P1_start = histbins1-binwidth1/2;
P1_end = histbins1+binwidth1/2;
% L1_th: theoretical results for lognormal, El_th, Rl_th, Gl_th
L1_th = (logncdf(P1_end, meanll, stdll)-
logncdf(P1_start, meanll, stdll))*length(Peak_value);
El_th = (expcdf(P1_end, lamdal)-
expcdf(P1_start, lamdal))*length(Peak_value);
Rl_th = (raylcdf(P1_end, alphal)-
raylcdf(P1_start, alphal))*length(Peak_value);
Gl_th = (exp(-exp(-alphal* (P1_end-mul)))-exp(-exp(-alphal* (P1_start-
mul))))*length(Peak_value);
% L1_cs: Chi-Square for lognormal
L1_cs = sum((nhist1-L1_th).^2./L1_th)
E1_cs = sum((nhist1-El_th).^2./El_th)
R1_cs = sum((nhist1-Rl_th).^2./Rl_th)
G1_cs = sum((nhist1-Gl_th).^2./G1_th)
% interval determined using the method in the textbook
P_start = histbin-binwidth/2;
\[ P_{\text{end}} = \text{histbin} + \text{binwidth}/2; \]

% L_th: theoretical results for lognormal, E_th, R_th, G_th
\[
L_{\text{th}} = (\logncdf(P_{\text{end}}, \text{meanll}, \text{stdll}) - \logncdf(P_{\text{start}}, \text{meanll}, \text{stdll})) \times \text{length(Peak\_value)};
\]
\[
E_{\text{th}} = (\expcdf(P_{\text{end}}, \text{lamdal}) - \expcdf(P_{\text{start}}, \text{lamdal}) \times \text{length(Peak\_value)};
\]
\[
R_{\text{th}} = (\raylcdf(P_{\text{end}}, \text{alphal}) - \raylcdf(P_{\text{start}}, \text{alphal}) \times \text{length(Peak\_value)};
\]
\[
G_{\text{th}} = (\exp(-\exp(-\text{all}\times(P_{\text{end}} - \text{mul})) - \exp(-\exp(-\text{all}\times(P_{\text{start}} - \text{mul})))) \times \text{length(Peak\_value)};
\]

% L_cs: Chi-Square for lognormal
\[
L_{\text{cs}} = \sum((\text{nhist} - L_{\text{th}})^2./L_{\text{th}})
\]
\[
E_{\text{cs}} = \sum((\text{nhist} - E_{\text{th}})^2./E_{\text{th}})
\]
\[
R_{\text{cs}} = \sum((\text{nhist} - R_{\text{th}})^2./R_{\text{th}})
\]
\[
G_{\text{cs}} = \sum((\text{nhist} - G_{\text{th}})^2./G_{\text{th}})
\]

D_S = [\text{meanl}; \text{stdl}; \text{PC\_mean}']
D_E = [\text{Mean\_exp}; \text{std\_exp}; H\_avg\_exp']
D_R = [\text{Mean\_ray}; \text{std\_ray}; H\_avg\_ray']
D_L = [\text{Mean\_log}; \text{std\_log}; H\_avg\_log']
D_G = [\text{Mean\_gum}; \text{std\_gum}; H\_avg\_gum']

% PC\_mean
% PC\_cut'
% Hn\_gum\_direct'
% Hn\_gum'
Dn_E = Dn_E
Dn_R = Dn_R
Dn_L = Dn_L
Dn_G = Dn_G
MATLAB Code
A.3. Peak Analysis
(Includes Selection of Peaks Using Vertical Threshold Method)

clear all
clc clf
format short
load BoatD.txt
test=BoatD;
% test=test(501:1000)
Data_Len=length(test)
A_mean=mean(test)
test=test-A_mean;
multipler=1;
A_rms=0;
for i=1:Data_Len
    A_rms=A_rms+(test(i))^2;
end
A_rms=sqrt(A_rms/Data_Len)
%buffer
A_threshold=multipler*A_rms
% test=test(90:200);
figure(1)
plot(test)
hold on
plot([1 length(test)],[0 0])
Data_Len=length(test)
% find the first local maximum, which larger than buffer
k=1;
for i=1:length(test)-1
    if test(i)<=A_threshold & test(i+1)>=A_threshold
        Aup_index(k)=i;
        k=k+1;
    end
end
kk=1;
for i=Aup_index(1):length(test)
    if test(i)>=A_threshold & test(i+1)<=A_threshold
        Adown_index(kk)=i+1;
        kk=kk+1;
    end
end
if size(Aup_index)>size(Adown_index)
    Aup_index=Aup_index(1:length(Adown_index));
end
for i=1:length(Adown_index)
    [Peak_value(i) P_I]=max(test(Aup_index(i):Adown_index(i)));
    Peak_index(i)=Aup_index(i)+P_I-1;
end
[P_max P_max_index]=max(Peak_value)
P_max_index=Peak_index(P_max_index)
Peak_sort=sort(Peak_value);
% find the average of 1/3, 1/10, 1/100 highest peaks and their cut values
H_nth=[3 10 100];
for jj=1:length(H_nth)
    if length(Peak_value)<H_nth(jj)
        PC_mean(jj)=0;
        PC_cut(jj)=0;
    else
        PC_mean(jj)=mean(Peak_sort(ceil((1-1./H_nth(jj))*length(Peak_value):length(Peak_value))));
        PC_cut(jj)=Peak_sort(ceil((1-1./H_nth(jj))*length(Peak_value)));
    end
end

% //Parameters for BoatADT/////
mean1=mean(Peak_value) % mean value
std1=std(Peak_value) % stand deviation
% Parameters for Lognormal distribution
std1l=sqrt(log(1+(std1/mean1)^2));
mean1l=log(mean1)-std1l^2/2;
% Parameters for Rayleigh distribution
alphal=(mean1/sqrt(pi/2)+std1/sqrt(2-pi/2))/2;
% Parameters for Exponential distribution
lamdall=(1/mean1+1/std1)/2;
% Parameters for Gumbel distribution
mul=mean1-0.5772/alphal;
% Generate the value accoring to input
x1=linspace(min(Peak_value),max(Peak_value));
% Lognormal pdf
y11=lognpdf(x1,mean1l,std1l);
% Rayleigh pdf
y12=raylpdf(x1,alphal);
% Exponential pdf
y13=lamdall*exp(-lamdall*x1);
% Gumbel distribution
y14=alphall*exp(-alphall*(x1-mul)).*exp(-exp(-alphall*(x1-mul)));
% Normal pdf
y15=normpdf(x1,mean1,std1);
% CDF information for different distribution
y24=exp(-exp(-alphall*(x1-mul)));
% //Parameters for BoatADT/////
figure(2)
plot(Peak_value)
hold on
plot([1,length(test)],[A_threshold,A_threshold],'r-','lineWidth',2.5)
hold on
plot(test)
hold on
plot(Peak_index,Peak_value,'ro')
figure(3)
plot(Peak_value)
hold on
plot([1,length(Peak_value)],[A_threshold,A_threshold],'r-','lineWidth',2.5)
% PDF plot
% /// Parameters for BoatADT ///
meanl=mean(Peak_value)  % mean value
stdl=std(Peak_value)  % stand deviation
% Parameters for Lognormal distribution
stdll=sqrt(log(1+(stdl/meanl)^2));
meanll=log(meanl)-stdll^2/2;
% Parameters for Rayleigh distribution
alphal=(meanl/sqrt(pi/2)+stdl/sqrt(2-pi/2))/2;
% Parameters for Exponential distribution
laml=(l/meanl+l/stdl)/2;
% Parameters for Gumbel distribution
all=pi/stdl/sqrt(6);
mul=meanl-0.5772/all;
% Generate the value according to input
xl=linspace(min(Peak_value),max(Peak_value));
% Lognormal pdf
y11=lognpdf(xl,meanll,stdll);
% Rayleigh pdf
y12=raylpdf(xl,alphal);
% Exponential pdf
y13=laml*exp(-laml*xl);
% Gumbel distribution
y14=all*exp(-all*(xl-mul)).*exp(-exp(-all*(xl-mul)));
% Normal pdf
y15=normpdf(xl,meanl,stdl);
% CDF information for different distribution
y24=exp(-exp(-all*(xl-mul)));
% /// Parameters for BoatADT ///
figure(4)
ksdensity(Peak_value)
hold on
plot(xl,y11,'-ro','markersize',4)
legend('Observed','Lognormal pdf')
title('BoatA VT (Threshold=lrms)')
figure(5)
ksdensity(Peak_value)
hold on
plot(xl,y12,'-ro','markersize',4)
legend('Observed','Rayleigh pdf')
title('BoatA VT (Threshold=lrms)')
figure(6)
ksdensity(Peak_value)
hold on
plot(xl,y13,'-ro','markersize',4)
legend('Observed','Exponential pdf')
title('BoatA VT (Threshold=lrms)')
figure(7)
ksdensity(Peak_value)
hold on
plot(xl,y14,'-ro','markersize',4)
legend('Observed','Gumbel pdf')
title('BoatA VT (Threshold=lrms)')
figure(8)
ksdensity(Peak_value)
hold on
plot(xl,y15,'-ro','markersize',4)
legend('Observed','Normal pdf')
title('BoatA VT (Threshold=lrms)')
figure(9)
ksdensity(Peak_value)
hold on
plot(xl,lognpdf(xl,meanll,stdll),'-go','markersize',4)
hold on
plot(xl,y12,'-bo','markersize',4)
plot(xl,y13,'-ko','markersize',4)
hold on
plot(xl,y14,'-ro','markersize',4)
legend('Observed', 'Lognormal', 'Rayleigh', 'Exponential', 'Gumbel')
title('BoatA VT (Threshold=lrms) PDF comparison')
figure(10)
ecdf(Peak_value)
hold on
plot(xl,logncdf(xl,meanll,stdll),'-ro','markersize',4)
hold on
% plot(xl,1/2*(1+erf((log(xl)-meanll)/(sqrt(2)*stdll))),'-bo','markersize',4)
legend('Observed','Lognormal Matlab','Lognormal equation')
title('BoatA VT (Threshold=lrms)')
figure(11)
ecdf(Peak_value)
hold on
plot(xl,expcdf(xl,lamdal),'-ro','markersize',4)
legend('Observed', 'Exponential cdf')
title('BoatA VT (Threshold=lrms)')
figure(12)
ecdf(Peak_value)
hold on
plot(xl,raylcdf(xl,alphal),'-ro','markersize',4)
legend('Observed','Rayleigh cdf')
title('BoatA VT (Threshold=lrms)')
figure(13)
ecdf(Peak_value)
hold on
plot(xl,y24,'-ro','markersize',4)
legend('Observed','Gumbel cdf')
title('BoatA VT (Threshold=lrms)')
figure(14)
ecdf(Peak_value)
hold on
plot(xl,logncdf(xl,meanll,stdll),'-go','markersize',4)
hold on
plot(xl,raylcdf(xl,alphal),'-bo','markersize',4)
hold on
plot(xl,expcdf(xl,lamdal),'-ko','markersize',4)
hold on
plot(xl,y24,'-ro','markersize',4)
legend('Observed', 'Lognormal', 'Rayleigh', 'Exponential', 'Gumbel cdf')
title('BoatA VT (Threshold=lrms) CDF comparison')

%% KS-test
% cdf of lognormal distribution according to the inputdata
y31=logncdf(sort(Peak_value),meanll,stdll);
% cdf of Rayleigh distribution according to the inputdata
y32=raylcdf(sort(Peak_value),alphal);
% cdf of Exponential distribution according to the inputdata
y33=expcdf(sort(Peak_value),lamdal);

% cdf of gumbel distribution according to the inputdata
y34=exp(-exp(-all*(sort(Peak_value)-mul)));
for i=1:length(Peak_value)
    E_cdf(i)=i/length(Peak_value);
end

Dn_L(l)=max(abs(y31-E_cdf));
Dn_R(l)=max(abs(y32-E_cdf));
Dn_E(l)=max(abs(y33-E_cdf));
Dn_G(l)=max(abs(y34-E_cdf));

% maximum different value of CDF for 1/3, 1/10, one-hundredth highest
for i=1:length(H_nth)
    Dn_L(i+l)=max(abs(y31(ceil((1-1/H_nth(i))/length(Peak_value))))-E_cdf(ceil((1-1/H_nth(i))/length(Peak_value))));
    Dn_R(i+l)=max(abs(y32(ceil((1-1/H_nth(i))/length(Peak_value))))-E_cdf(ceil((1-1/H_nth(i))/length(Peak_value))));
    Dn_E(i+l)=max(abs(y33(ceil((1-1/H_nth(i))/length(Peak_value))))-E_cdf(ceil((1-1/H_nth(i))/length(Peak_value))));
    Dn_G(i+l)=max(abs(y34(ceil((1-1/H_nth(i))/length(Peak_value))))-E_cdf(ceil((1-1/H_nth(i))/length(Peak_value))));
end

%% monte carlo simulation
samples=1000000;
% H_nth=[3 10 100];
% Parameters for Lognormal distribution
X_log=exp(erfinv(2*rand(samples,1)-1)*sqrt(2)*stdll+meanll);
Xsort_log=sort(X_log);
Mean_log=mean(X_log);
std_log=std(X_log);
% Parameters for Rayleigh distribution
alpha1=(mean1/sqrt(pi/2)+stdl/sqrt(2-pi/2))/2;
X_ray=sqrt(-2*log(1-rand(samples,1)))*alpha1;
Xsort_ray=sort(X_ray);
Mean_ray=mean(X_ray);
std_ray=std(X_ray);
% Parameters for Exponential distribution
X_exp=-log(1-rand(samples,1))/lamdal;
Xsort_exp=sort(X_exp);
Mean_exp=mean(X_exp);
std_exp=std(X_exp);
% Parameters for Gumbel distribution
X_gum=-log(-log(rand(samples,1)))/all+mul;
Xsort_gum=sort(X_gum);
Mean_gum=mean(X_gum);
std_gum=std(X_gum);
for i=1:length(H_nth)
    Hn_log(i)=exp(erfinv((H_nth(i)-2)/H_nth(i))*sqrt(2)*stdll+meanll))
    Hn_log_direct(i)=Xsort_log(ceil((1-1/H_nth(i))*samples))
    aaa=0.147;
    aal=2/pi/aaa;
    aa2=log(4*H_nth(i)-1)/H_nth(i)^2;
    Hn_log_theo(i)=exp(meanll+sqrt(2)*stdll*sqrt(-aal-aa2/2+sqrt((aal+aa2/2)^2-1/aaa*aa2)))
    X_sum_log=0;
    for j=1:samples
        if X_log(j)>=Hn_log(i)
\begin{verbatim}
X_sum_log=X_sum_log+X_log(j);
end
end
H_avg_log(i)=H_nth(i)*X_sum_log/samples
H_avg_log_direct(i)=mean(Xsort_log(ceil((1-1/H_nth(i))*samples):end))
%Rayleigh distribution
Hn_ray(i)=alpha*sqrt(2*log(H_nth(i)))
Hn_ray_direct(i)=Xsort_ray(ceil((1-1/H_nth(i))*samples))
% num_ray=sum(X_ray>=Hn_ray);
X_sum_ray=0;
for j=1:samples
    if X_ray(j)>=Hn_ray(i)
        X_sum_ray=X_sum_ray+X_ray(j);
    end
end
H_avg_ray(i)=H_nth(i)*X_sum_ray/samples
H_avg_ray_direct(i)=mean(Xsort_ray(ceil((1-1/H_nth(i))*samples):end))
% Exponential distribution
Hn_exp(i)=1/lambda*log(H_nth(i))
Hn_exp_direct(i)=Xsort_log(ceil((1-1/H_nth(i))*samples))
% num_exp=sum(X_exp>=Hn_exp);
X_sum_exp=0;
for j=1:samples
    if X_exp(j)>=Hn_exp(i)
        X_sum_exp=X_sum_exp+X_exp(j);
    end
end
H_avg_exp(i)=H_nth(i)*X_sum_exp/samples
H_avg_exp_direct(i)=mean(Xsort_exp(ceil((1-1/H_nth(i))*samples):end))
H_avg_exp_theo(i)=1/lambda*(1+log(H_nth(i)))
%Gumbel distribution
Hn_gum(i)=-log(-log(1-1/H_nth(i)))/all+mul
Hn_gum_direct(i)=Xsort_log(ceil((1-1/H_nth(i))*samples))
% num_gum=sum(X_gum>=Hn_gum);
X_sum_gum=0;
for j=1:samples
    if X_gum(j)>=Hn_gum(i)
        X_sum_gum=X_sum_gum+X_gum(j);
    end
end
H_avg_gum(i)=H_nth(i)*X_sum_gum/samples
H_avg_gum_direct(i)=mean(Xsort_gum(ceil((1-1/H_nth(i))*samples):end))
end
%% histogram and frequency diagram plot
NN=round(1+3.3*log10(length(Peak_value)));
binwidth=(max(Peak_value)-min(Peak_value))/NN;
[nhist histbin] =hist(Peak_value,4*NN);
figure(15)
bar(histbin,nhist/length(Peak_value)/binwidth)
Areacheck=sum(nhist/length(Peak_value)/binwidth)*binwidth
\end{verbatim}
hold on
plot(xl,y14,'-r','lineWidth',2.5)
legend('Hist','Gumbel pdf')
title('Peaks value selected by VT (Threshold=lrms)');
figure(16)
hist(Peak_value,NN)

Different method to generate bidwidth
binwidthl=2*iqr(Peak_value)/length(Peak_value)^{(1/3)}
bins1=min(Peak_value):binwidth1:max(Peak_value);
[nhisl histbinsl]=hist(Peak_value,bins1);
figure(17)
bar(histbinsl,nhisl/length(Peak_value)/binwidthl)
Areacheckl=sum(nhisl/length(Peak_value)/binwidthl)*binwidthl

hold on
plot(xl,lognpdf(xl,meanll,stdll),'-go','markersize',4)
hold on
plot(xl,yl2,'-bo','markersize',4)
hold on
plot(xl,yl3,'-ko','markersize',4)
hold on
plot(xl,yl4,'-r','lineWidth',2.5)
legend('Hist','Lognormal','Rayleigh','Exponential','Gumbel pdf')
title('Peaks value selected by VT (Threshold=lrms)');

Chi-square test
Interval determined using 75% percentage
P1_start=histbinsl-binwidthl/2;
P1_end=histbinsl+binwidthl/2;

L1_th=theoretical results for lognormal, E1_th, R1_th, G1_th
L1_th=(logncdf(P1_end,mean11,std11)-
logncdf(P1_start,mean11,std11))*length(Peak_value);
E1_th=(expcdf(P1_end,lamdal)-
expcdf(P1_start,lamdal))*length(Peak_value);
R1_th=(raylcdf(P1_end,alphal)-
raylcdf(P1_start,alphal))*length(Peak_value);
G1_th=(exp(-exp(-alphal*(P1_end-mul)))-exp(-exp(-alphal*(P1_start-
mul))))*length(Peak_value);

L1_cs: Chi-Square for lognormal
L1_cs=sum((nhisl-L1_th).^2./L1_th)
E1_cs=sum((nhisl-E1_th).^2./E1_th)
R1_cs=sum((nhisl-R1_th).^2./R1_th)
G1_cs=sum((nhisl-G1_th).^2./G1_th)

Interval determined using the method in the textbook
P_start=histbin-binwidthl/2;
P_end=histbin+binwidthl/2;

L_th=theoretical results for lognormal, E_th, R_th, G_th
L_th=(logncdf(P_end,mean11,std11)-
logncdf(P_start,mean11,std11))*length(Peak_value);
E_th=(expcdf(P_end,lamdal)-expcdf(P_start,lamdal))*length(Peak_value);
R_th=(raylcdf(P_end,alphal)-
raylcdf(P_start,alphal))*length(Peak_value);
G_th=(exp(-exp(-alphal*(P_end-mul)))-exp(-exp(-alphal*(P_start-
mul))))*length(Peak_value);

L_cs: Chi-Square for lognormal
L_cs=sum((nhisl-L_th).^2./L_th)
E_cs=sum((nhisl-E_th).^2./E_th)
R_cs=sum((nhisl-R_th).^2./R_th)
G_cs=sum((nhisl-G_th).^2./G_th)
Dn_E = Dn_E'
Dn_R = Dn_R'
Dn_L = Dn_L'
Dn_G = Dn_G'
D_S = [mean_l; std_l; PC_mean']
D_E = [Mean_exp; std_exp; H_avg_exp']
D_R = [Mean_ray; std_ray; H_avg_ray']
D_L = [Mean_log; std_log; H_avg_log']
D_G = [Mean_gum; std_gum; H_avg_gum']
MATLAB Code
A.4. Peak Analysis
(Includes Selection of Peaks Using Vertical Difference Method)

clear all
clc
cf
format short
load test.txt
load BoatD.txt
test=BoatD;
Data_Len=length(test)
A_mean=mean(test);
test=test-A_mean;
multipler=1.;
A_rms=0;
for i=1:length(test)
    A_rms=A_rms+(test(i))^2;
end
A_rms=sqrt(A_rms/length(test));
%buffer
A_threshold1=multipler*A_rms;
A_threshold2=multipler*A_rms;
% this distance should be larger than 2*A_rms
A_d=2*A_rms*multipler;
figure(1)
plot(test)
hold on
plot([1 length(test)],[0 0])
% find the peak which is between the threshold1
k=1;
for i=1:length(test)-1
    if test(i)<=A_threshold1 & test(i+1)>=A_threshold1
        Aup_index(k)=i;
        k=k+1;
    end
end
kk=1;
for i=Aup_index(1):length(test)-1
    if test(i)>=A_threshold1 & test(i+1)<=A_threshold1
        Adown_index(kk)=i+1;
        kk=kk+1;
    end
end
if size(Aup_index)>size(Adown_index)
    Aup_index=Aup_index(1:length(Adown_index));
end
for i=1:length(Adown_index)
    [ Peak_value(i) P_I] =max(test(Aup_index(i):Adown_index(i)));
    Peak_index(i)=Aup_index(i)+P_I-1;
end
% find the trough which is between -threshold2
k=1;
for i=1:length(test)-1
    if test(i)\geq-A_{\text{threshold}2} & test(i+1)\leq-A_{\text{threshold}2}
        T_{\text{down}}(k)=i;
        k=k+1;
    end
end

kk=1;
for i=T_{\text{down}}(1):length(test)-1
    if test(i)\leq-A_{\text{threshold}2} & test(i+1)\geq-A_{\text{threshold}2}
        T_{\text{up}}(kk)=i;
        kk=kk+1;
    end
end
if size(T_{\text{down}})>size(T_{\text{up}})
    T_{\text{down}}=T_{\text{down}}(1:length(T_{\text{up}}));
end

for i=1:length(T_{\text{up}})
    [T_{\text{trough}}(i) T_{\text{I}}]=\min(test(T_{\text{down}}(i):T_{\text{up}}(i)));
    T_{\text{down}}(i)=T_{\text{down}}(i)+T_{\text{I}}-1;
end

Peak_index(1)
Trough_index(1)
figure(2)
plot(test)
hold
plot(Peak_index,Peak_value,'ro')
hold on
plot([1,length(test)],[A_{\text{threshold}1},A_{\text{threshold}1}],'r-','\text{lineWidth}',2.5)
hold on
plot(Trough_index,Trough_value,'gd')
hold on
plot([1,length(test)],[A_{\text{threshold}2},A_{\text{threshold}2}],'r-','\text{lineWidth}',2.5)
start1=1;
kk=1;
for i=start1:length(Peak_index)
    if Peak_index(start1)\leq T_{\text{trough}}(end)
        aa=\min(find(T_{\text{trough}}>Peak_index(start1)));
        if T_{\text{trough}}(end)\leq Peak_index(\text{end})
            start2=\min(find(Peak_index>T_{\text{trough}}(\text{aa})));  
            if T_{\text{trough}}(\text{end})=Peak_index(\text{end})
                bb=\max(find(T_{\text{trough}}<Peak_index(start2)));
                [P_{\text{max}}(kk) P_{\text{index}}]=\max(Peak_value(start1:start2-1));
                P_{\text{max}}(kk)=Peak_index(start1+P_{\text{index}}-1);
                [T_{\text{min}}(kk) T_{\text{index}}]=\min(Trough_value(aa:bb));
                T_{\text{min}}(kk)=Trough_index(aa+T_{\text{index}}-1);
                kk=kk+1;
                start1=start2;
            else
                [P_{\text{max}}(kk) P_{\text{index}}]=\max(Peak_value(start1:start2-1);
                P_{\text{max}}(kk)=Peak_index(start1+P_{\text{index}}-1);
                [T_{\text{min}}(kk) T_{\text{index}}]=\min(Trough_value(aa:end));
                T_{\text{min}}(kk)=Trough_index(aa+T_{\text{index}}-1);
                kk=kk+1
                start1=length(Peak_index)
                break
            end
        end
    end
end
else
  \% start2=length(Peak_index); \\
  \% bb=length(Trough_index); \\
  [Pmax(kk) P_index] =max(Peak_value(start1:end)); \\
  Pmax_index(kk)=Peak_index(start1+P_index-1); \\
  [Tmin(kk) T_index] =min(Trough_value(aa:end)); \\
  Tmin_index (kk) =Trough__index (aa+T_index-1) ; \\
  kk=kk+1;  \\
  start1=length(Peak_index);  \\
  break \\
else \\
  break \\
end \\
end \\
figure (3) \\
plot(Pmax_index,Pmax,'ro') \\
hold on \\
plot(Tmin_index,Tmin,'gd') \\
hold on \\
plot(test) \\
hold on \\
plot([1,length(test)],[A_threshold1,A_threshold1],'r-','linewidth',2.5) \\
hold on \\
plot([1,length(test)],[-A_threshold2,-A_threshold2],'r- \\
','linewidth',2.5) \\
title('Peak and Trough change alternately') \\
\%check the min value and corresponding location \\
\% Tmin(4) \\
\% test(Tmin_index(4)) \\
\%find the peak and trough that satisfies the distance \\
jj=1; \\
for i=1:length(Pmax) \\
  if Pmax(i)-Tmin(i)>=A_d 
    Peak_max_V(jj)=Pmax(i); 
    Peak_max_I(jj)=Pmax_index(i); 
    Trough_min_V(jj)=Tmin(i); 
    Trough_min_I(jj)=Tmin_index(i); 
    jj=jj+1; 
  end 
end \\
jj \\
if jj=1 
  disp('no peaks is selected, please reduce the distance') 
end \\
[P_max P_max_index]=max(Peak_max_V) \\
P_max_index=Peak_index(P_max_index) \\
Peak_sort=sort(Peak_max_V); \\
\%find the average of 1/3,1/10,1/100 highest peaks and their cut values \\
if length(Peak_max_V)<3 
  disp('the selected peak values are less than 3, there is no average 
  highest exist') 
  Pmean_13=0; 
  P_13=0; 
  Pmean_110=0; 
  P_110=0; 
  Pmean_1100=0;
P_1100=0;
elseif length(Peak_max_V)<10
  disp('Only 1/3 highest exist')
  Pmean_13=mean(Peak_sort(ceil(2/3*length(Peak_max_V))):end))
  P_13=Peak_sort(ceil(2/3*length(Peak_max_V)));
  Pmean_110=0;
  P_110=0;
  Pmean_1100=0;
  P_1100=0;
elseif length(Peak_max_V)<100
  disp('The average of 1/100 highest does not exist')
  Pmean_13=mean(Peak_sort(ceil(2/3*length(Peak_max_V))):end));
  P_13=Peak_sort(ceil(2/3*length(Peak_max_V)));
  Pmean_110=mean(Peak_sort(ceil(9/10*length(Peak_max_V))):end));
  P_110=Peak_sort(ceil(9/10*length(Peak_max_V)));
  Pmean_1100=0;
  P_1100=0;
else
  disp('The average of 1/3,1/10,1/100 highest and their cut values
  are as follows:')
  Pmean_13=mean(Peak_sort(ceil(2/3*length(Peak_max_V))):end));
  P_13=Peak_sort(ceil(2/3*length(Peak_max_V)));
  Pmean_110=mean(Peak_sort(ceil(9/10*length(Peak_max_V))):end));
  P_110=Peak_sort(ceil(9/10*length(Peak_max_V)));
  Pmean_1100=mean(Peak_sort(ceil(99/100*length(Peak_max_V))):end));
  P_1100=Peak_sort(ceil(99/100*length(Peak_max_V)));
end
% PDF plot
%\\\\\\\\\///Parameters for BoatADT\\\\\\\\\///
meanl=mean(Peak_max_V) % mean value
stdl=std(Peak_max_V) %stand deviation
%Parameters for Lognormal distribution
stdll=sqrt(log(1+(stdl/meanl)^2));
meanll=log(meanl)-stdll^2/2;
%Parameters for Rayleigh distribution
alphal=(meanl/sqrt(pi/2)+stdl/sqrt(2-pi/2))^2/2;
%Parameters for Exponential distribution
laml=(1/(meanl+1/stdl))/2;
%Parameters for Gumbel distribution
all=pi/stdl/sqrt(6);
mul=meanl-0.5772/all;
% Generate the value accoring to input
x1=linspace(min(Peak_max_V),max(Peak_max_V));
%Lognormal pdf
y11=lognpdf(x1,mean1, stdll);
%Rayleigh pdf
y12=raylpdf(x1,alphal);
% Exponential pdf
y13=lamdal*exp(-lamdal*x1);
% Gumbel distribution
y14=all*exp(-all*(x1-mul)).*exp(-exp(-all*(x1-mul)));
%Normal pdf
y15=normpdf(x1,mean1, stdll);
% CDF information for different distribution
y24=exp(-exp(-all*(x1-mul)));
%\\\\\\\\\///Parameters for BoatADT\\\\\\\\\///
figure(4)
ksdensity(Peak_max_V)
hold on
plot(xl,yl1,'-ro','markersize',4)
legend('Observed','Lognormal pdf')
title('BoatA DT Method (threshold=lrms Peak-Trough=3rms)')
figure(5)
ksdensity(Peak_max_V)
hold on
plot(xl,yl2,'-ro','markersize',4)
legend('Observed','Rayleigh pdf')
title('BoatA DT Method (threshold=lrms Peak-Trough=3rms)')
figure(6)
ksdensity(Peak_max_V)
hold on
plot(xl,yl3,'-ro','markersize',4)
legend('Observed','Exponential pdf')
title('BoatA DT Method (threshold=lrms Peak-Trough=3rms)')
figure(7)
ksdensity(Peak_max_V)
hold on
plot(xl,yl4,'-ro','markersize',4)
legend('Observed','Gumbel pdf')
title('BoatA DT Method (threshold=lrms Peak-Trough=3rms)')
figure(8)
ksdensity(Peak_max_V)
hold on
plot(xl,yl5,'-ro','markersize',4)
legend('Observed','Normal pdf')
title('BoatA DT Method (threshold=lrms Peak-Trough=3rms)')
figure(9)
ecdf(Peak_max_V)
hold on
plot(xl,logncdf(xl,meanll,stdll),'-ro','markersize',4)
legend('Observed','Lognormal Matlab','Lognormal equation')
title('BoatA DT Method (threshold=lrms Peak-Trough=3rms)')
figure(10)
ecdf(Peak_max_V)
hold on
plot(xl,expcdf(xl,lamdal),'-ro','markersize',4)
legend('Observed','Exponential cdf')
title('BoatA DT Method (threshold=lrms Peak-Trough=3rms)')
figure(11)
ecdf(Peak_max_V)
hold on
plot(xl,raylcdf(xl,alphal),'-ro','markersize',4)
legend('Observed','Rayleigh cdf')
title('BoatA DT Method (threshold=lrms Peak-Trough=3rms)')
figure(13)
ecdf(Peak_max_V)
hold on
plot(xl,y24,'-ro','markersize',4)
legend('Observed','Gumbel cdf')
title('BoatA DT Method (threshold=lrms Peak-Trough=3rms)')
figure(14)
ecdf(Peak_max_V)
hold on
plot(xl,logncdf(xl,meanll,stdll),'-go','markersize',4)
hold on
plot(xl,raylcdf(xl,alphal),'-bo','markersize',4)
hold on
plot(xl,expcdf(xl,ladlal),'-ko','markersize',4)
hold on
plot(xl,y24,'-ro','markersize',4)
legend('Observed','Lognormal','Rayleigh','Exponential','Gumbel cdf')
title('BoatA DT Method (threshold=lrms Peak-Trough=3rms) CDF comparison')

% cdf of lognormal distribution according to the input data
y31=logncdf(sort(Peak_max_V),meanll,stdll);
% cdf of Rayleigh distribution according to the input data
y32=raylcdf(sort(Peak_max_V),alphal);
% cdf of Exponential distribution according to the input data
y33=expcdf(sort(Peak_max_V),ladlal);
% cdf of Gumbel distribution according to the input data
y34=exp(-exp(-all*(sort(Peak_max_V)-mul)));
for i=1:length(Peak_max_V)
    E_cdf(i)=i/length(Peak_max_V);
end
Dn_L_overall=max(abs(y31-E_cdf))
Dn_R_overall=max(abs(y32-E_cdf))
Dn_E_overall=max(abs(y33-E_cdf))
Dn_G_overall=max(abs(y34-E_cdf))

% 1/3rd maximum difference
Dn_L_13=max(abs(y31(ceil(2/3*length(Peak_max_V)):end)-
    E_cdf(ceil(2/3*length(Peak_max_V)):end)))
Dn_R_13=max(abs(y32(ceil(2/3*length(Peak_max_V)):end)-
    E_cdf(ceil(2/3*length(Peak_max_V)):end)))
Dn_E_13=max(abs(y33(ceil(2/3*length(Peak_max_V)):end)-
    E_cdf(ceil(2/3*length(Peak_max_V)):end)))
Dn_G_13=max(abs(y34(ceil(2/3*length(Peak_max_V)):end)-
    E_cdf(ceil(2/3*length(Peak_max_V)):end)))
% one-tenth maximum difference
Dn_L_110=max(abs(y31(ceil(9/10*length(Peak_max_V)):end)-
    E_cdf(ceil(9/10*length(Peak_max_V)):end)))
Dn_R_110=max(abs(y32(ceil(9/10*length(Peak_max_V)):end)-
    E_cdf(ceil(9/10*length(Peak_max_V)):end)))
Dn_E_110=max(abs(y33(ceil(9/10*length(Peak_max_V)):end)-
    E_cdf(ceil(9/10*length(Peak_max_V)):end)))
Dn_G_110=max(abs(y34(ceil(9/10*length(Peak_max_V)):end)-
    E_cdf(ceil(9/10*length(Peak_max_V)):end)))
% one-hundredth maximum difference
Dn_L_1100=max(abs(y31(ceil(99/100*length(Peak_max_V)):end)-
E_cdf(ceil(99/100*length(Peak_max_V)):end))
Dn_R_1100=max(abs(y32(ceil(99/100*length(Peak_max_V)):end)-
E_cdf(ceil(99/100*length(Peak_max_V)):end))
Dn_E_1100=max(abs(y33(ceil(99/100*length(Peak_max_V)):end)-
E_cdf(ceil(99/100*length(Peak_max_V)):end))
Dn_G_1100=max(abs(y34(ceil(99/100*length(Peak_max_V)):end)-
E_cdf(ceil(99/100*length(Peak_max_V)):end))

% monte carlo simulation
samples=1000000;
H_nth=[3 10 100];

% Parameters for Lognormal distribution
X_log=exp(erfinv(2*rand(samples,1)-1)*sqrt(2)*stdll+meanll);
X_sort_log=sort(X_log);
Mean_log=mean(X_log);
std_log=std(X_log);

% Parameters for Rayleigh distribution
alpha1=(meanl/sqrt(pi/2)+stdl/sqrt(2-pi/2))/2;
X_ray=sqrt(-2*log(1-rand(samples,1)))*alpha1;
X_sort_ray=sort(X_ray);
Mean_ray=mean(X_ray);
std_ray=std(X_ray);

% Parameters for Exponential distribution
X_exp=-log(1-rand(samples,1))/lamdal;
X_sort_exp=sort(X_exp);
Mean_exp=mean(X_exp);
std_exp=std(X_exp);

% Parameters for Gumbel distribution
X_gum=-log(-log(rand(samples,1)))/all+mul;
X_sort_gum=sort(X_gum);
Mean_gum=mean(X_gum);
std_gum=std(X_gum);
for i=1:length(H_nth)
    Hn_log(i)=exp(erfinv((H_nth(i)-2)/H_nth(i))*sqrt(2)*stdll+meanll)
    Hn_log_direct(i)=X_sort_log(ceil((1-1/H_nth(i))*samples))
    aaa=0.147;
    aal=2/pi/aaa;
    a2=4*H_nth(i)-1)/H_nth(i)^2;
    Hn_log_theo(i)=exp(meanll+sqrt(2)*stdll*sqrt(-aal-
    a2/2+sqrt((aal+a2/2)^2-1/aaa*a2))
    X_sum_log=0;
    for j=1:samples
        if X_log(j)>=Hn_log(i)
            X_sum_log=X_sum_log+X_log(j);
        end
    end
    H_avg_log(i)=H_nth(i)*X_sum_log/samples
    H_avg_log_direct(i)=mean(X_sort_log(ceil((1-
    1/H_nth(i))*samples):end))

    % Rayleigh distribution
    Hn_ray(i)=alpha1*sqrt(2*log(H_nth(i)))
    Hn_ray_direct(i)=X_sort_ray(ceil((1-1/H nth(i))*samples))
    % num_ray=sum(X_ray>=Hn_ray);
    X_sum_ray=0;
    for j=1:samples
        if X_ray(j)>=Hn_ray(i)
            X_sum_ray=X_sum Ray+X Ray(j);
        end
    end
    H_avg Ray(i)=H nth(i)*X_sum Ray/samples
    H_avg Ray direct(i)=mean(X_sort Ray(ceil((1-
    1/H nth(i))*samples):end))
%
if X_ray(j)>=Hn_ray(i)
    X_sum_ray=X_sum_ray+X_ray(j);
end
end
H_avg_ray(i)=H_nth(i)*X_sum_ray/samples
H_avg_ray_theo(i)=
    H_nth(i)*alphal*sqrt(2)*(sqrt(pi)/Z*(erf(log(H_nth(i)))-1)-
    sqrt(log(H_nth(i)))/H_nth(i))
    H_avg_ray_direct(i)=mean(Xsort_ray(ceil((1-
    1/H_nth(i))*samples):end))
% Exponential distribution
Hn_exp(i)=(1/lamdal)*log(H_nth(i))
Hn_exp_direct(i)=Xsort_log(ceil((1-
    1/H_nth(i))*samples))
% num_exp=sum(X_exp<=Hn_exp);
X_sum_exp=0;
for j=1:samples
    if X.exp(j)>=Hn_exp(i)
        X_sum_exp=X_sum_exp+X_exp(j);
    end
end
H_avg_exp(i)=H_nth(i)*X_sum_exp/samples
H_avg_exp_direct(i)=mean(Xsort_exp(ceil((1-
    1/H_nth(i))*samples):end))
H_avg_exp_theo(i)=(1/lamdal+(1+log(H_nth(i))))
% Gumbel distribution
Hn_gum(i)=-log(-log(1-l/H_nth(i)))/all+mul
Hn_gum_direct(i)=Xsort_log(ceil((1-
    1/H_nth(i))*samples))
% num_gum=sum(X_gum>Hn_gum);
X_sum_gum=0;
for j=1:samples
    if X_gum(j)>Hn_gum(i)
        X_sum_gum=X_sum_gum+X_gum(j);
    end
end
H_avg_gum(i)=H_nth(i)*X_sum_gum/samples
H_avg_gum_direct(i)=mean(Xsort_gum(ceil((1-
    1/H_nth(i))*samples):end))
end
NN=round(1+3.3*log10(length(Peak_max_V)));
binwidth=(max(Peak_max_V)-min(Peak_max_V))/NN;
[ nhist histbin] =hist(Peak_max_V,4*NN);
figure(15)
bar(histbin, nhist/length(Peak_max_V)/binwidth)
Areacheck=sum(nhist/length(Peak_max_V)/binwidth)*binwidth
hold on
plot(xl,yl4,'-r','lineWidth',2.5)
legend('Hist','Gumbel pdf')
title('Peaks value selected by DT Method (Threshold=lrms Peak-
Trough=3rms)');
figure(16)
hist(Peak_max_V,NN)
% different method to generate bidwidth
binwidthl=2*iqr(Peak_max_V)/length(Peak_max_V)^(1/3)
binsl=min(Peak_max_V):binwidthl:max(Peak_max_V);
[ nhist1 histbinsl] =hist(Peak_max_V,binsl);
figure(17)
bar(histbinsl,nhist1/length(Peak_max_V)/binwidth1)
Area check 1 = sum(nhistl/length(Peak_max_V)/binwidthl)*binwidthl
hold on
plot(xl,lognpdf(xl,meanll,stderrll),'-go','markersize',4)
hold on
plot(xl,yl2,'-bo','markersize',4)
hold on
plot(xl,yl3,'-ko','markersize',4)
hold on
plot(xl,yl4,'-r','lineWidth',2.5)
legend('Hist','Lognormal','Rayleigh','Exponential','Gumbel pdf')
title('Peaks value selected by DT Method(Threshold=1rms, Peak-Trough=3rms)');
Dn_E=[ Dn_E_overall;Dn_E_13;Dn_E_110;Dn_E_1100]
Dn_R=[ Dn_R_overall;Dn_R_13;Dn_R_110;Dn_R_1100]
Dn_L=[ Dn_L_overall;Dn_L_13;Dn_L_110;Dn_L_1100]
Dn_G=[ Dn_G_overall;Dn_G_13;Dn_G_110;Dn_G_1100]
D_S=[ meanl;stderrl;Pmean_13;Pmean_110;Pmean_1100]
D_E=[ Mean_exp;std_exp;H_avg_exp]
D_R=[ Mean_ray;std_ray;H_avg_ray]
D_L=[ Mean_log;std_log;H_avg_log]
D_G=[ Mean_gum;std_gum;H_avg_gum]
MATLAB Code

A.5. Peak Analysis
(Includes Selection of Peaks Using Horizontal Threshold Method)

clear all
clc
clf
format short
% Find the peaks using HT method
format short
load BoatB.txt
test=BoatB;
load Data403_0.txt;
test=Data403_08(:,5);
% test=test(1:1000)
Data_Len=length(test)
A_mean=mean(test);
test=test-A_mean;
% peaks found based on Hirisontal threshold method
HT_len=256;
A_Index=floor(length(test)/HT_len);
for i=1:A_index
    [ Peak_value(i) P_I] =max(test((1+(i-1)*HT_len):i*HT_len));
    Peak_index(i)=(i-1)*HT_len+P_I;
    [Trough_value(i) T_I] =min(test((1+(i-1)*HT_len):i*HT_len));
    Trough_index(i)=(i-1)*HT_len+T_I;
end
if length(test)>A_index*HT_len
    [ Peak_value(A_index+1) aa] =max(test((1+A_index*HT_len):end));
    Peak_index(A_index+1)=A_index*HT_len+aa;
    [Trough_value(A_index+1) bb] =min(test((1+A_index*HT_len):end));
    Trough_index(A_index+1)=A_index*HT_len+bb;
end
HT_len=256;
[ P_max P_max_index] =max(Peak_value);
P_max_index=Peak_index(P_max_index);
Peak_sort=sort(Peak_value);
% find the average of 1/3,1/10,1/100 and 0.2/100 highest peaks and their cut values
H_nth=[3 10 100 500];
for jj=1:length(H_nth)
    if length(Peak_value)<H_nth(jj)
        PC_mean(jj)=0;
        PC_cut(jj)=0;
        else
        PC_mean(jj)=mean(Peak_sort(ceil((1-
                        1./H_nth(jj))*length(Peak_value):length(Peak_value))));
        PC_cut(jj)=Peak_sort(ceil((1-
                        1./H_nth(jj))*length(Peak_value)));
        end
end
%////////////Parameters for BoatADT///////
mean1=mean(Peak_value);  % mean value
\[
\text{stdl} = \text{std}(\text{Peak}_\text{value}); \quad \%\text{stand deviation}
\]
\[
\%\text{Parameters for Lognormal distribution}
\]
\[
\text{stdll} = \sqrt{\log(1 + (\text{stdl}/\text{mean1})^2)};
\]
\[
\text{meanll} = \log(\text{mean1}) - \text{stdll}^2/2;
\]
\[
\%\text{Parameters for Rayleigh distribution}
\]
\[
\alpha_\text{l} = (\text{mean1}/\sqrt{\pi/2} + \text{stdl}/\sqrt{2-\pi/2})/2;
\]
\[
\%\text{Parameters for Exponential distribution}
\]
\[
\lambda_\text{l} = (1/\text{mean1} + 1/\text{stdl})/2;
\]
\[
\%\text{Parameters for Gumbel distribution}
\]
\[
\alpha_\text{l} = \pi/\text{stdl}/\sqrt{6};
\]
\[
\mu_\text{l} = \text{mean1} - 0.5772/\alpha_\text{l};
\]
\[
\%\text{Generate the value according to input}
\]
\[
x_\text{l} = \text{linspace}(\text{min}(\text{Peak}_\text{value}), \text{max}(\text{Peak}_\text{value}));
\]
\[
\%\text{Lognormal pdf}
\]
\[
y_{\text{l}} = \text{lognpdf}(x_\text{l}, \text{meanll}, \text{stdll});
\]
\[
\%\text{Rayleigh pdf}
\]
\[
y_{\text{l2}} = \text{raylpdf}(x_\text{l}, \alpha_\text{l});
\]
\[
\%\text{Exponential pdf}
\]
\[
y_{\text{l3}} = \lambda_\text{l} \cdot \exp(-\lambda_\text{l} \cdot x_\text{l});
\]
\[
\%\text{Gumbel distribution}
\]
\[
y_{\text{l4}} = \alpha_\text{l} \cdot \exp(-\alpha_\text{l} \cdot (x_\text{l} - \mu_\text{l})) \cdot \exp(-\exp(-\alpha_\text{l} \cdot (x_\text{l} - \mu_\text{l})))
\]
\[
\%\text{Normal pdf}
\]
\[
y_{\text{l5}} = \text{normpdf}(x_\text{l}, \text{mean1}, \text{stdl});
\]
\[
\%\text{CDF information for different distribution}
\]
\[
y_{\text{l4}} = \exp(-\exp(-\alpha_\text{l} \cdot (x_\text{l} - \mu_\text{l})))
\]
\[
\%\text{////Parameters for BoatADT/////}
\]
\[
\text{figure}(1)
\]
\[
\text{plot}(\text{test})
\]
\[
\text{hold on}
\]
\[
\text{plot}([1 \text{ length(\text{test})}],[0 0],'r','linwidth',2.5)
\]
\[
\text{figure}(2)
\]
\[
\text{plot(\text{Peak}_\text{index}, \text{Peak}_\text{value},'r')}
\]
\[
\text{hold on}
\]
\[
\text{plot(\text{test})}
\]
\[
\%\text{////Parameters for BoatADT/////}
\]
\[
\text{mean1} = \text{mean}(\text{Peak}_\text{value}) \%\text{mean value}
\]
\[
\text{stdl} = \text{std}(\text{Peak}_\text{value}) \%\text{stand deviation}
\]
\[
\%\text{Parameters for Lognormal distribution}
\]
\[
\text{stdll} = \sqrt{\log(1 + (\text{stdl}/\text{mean1})^2)};
\]
\[
\text{meanll} = \log(\text{mean1}) - \text{stdll}^2/2;
\]
\[
\%\text{Parameters for Rayleigh distribution}
\]
\[
\alpha_\text{l} = (\text{mean1}/\sqrt{\pi/2} + \text{stdl}/\sqrt{2-\pi/2})/2;
\]
\[
\%\text{Parameters for Exponential distribution}
\]
\[
\lambda_\text{l} = (1/\text{mean1} + 1/\text{stdl})/2;
\]
\[
\%\text{Parameters for Gumbel distribution}
\]
\[
\alpha_\text{l} = \pi/\text{stdl}/\sqrt{6};
\]
\[
\mu_\text{l} = \text{mean1} - 0.5772/\alpha_\text{l};
\]
\[
\%\text{Generate the value according to input}
\]
\[
x_\text{l} = \text{linspace}(\text{min}(\text{Peak}_\text{value}), \text{max}(\text{Peak}_\text{value}));
\]
\[
\%\text{Lognormal pdf}
yll = lognpdf(xl, mean11, stdll);
% Rayleigh pdf
yl2 = raylpdf(xl, alphal);
% Exponential pdf
yl3 = lamdal*exp(-lamdal*xl);
% Gumbel distribution
yl4 = all*exp(-all*(xl-mul)).*exp(-exp(-all*(xl-mul)));
% Normal pdf
yl5 = normpdf(xl, meanl, stdl);
% CDF information for different distribution
y24 = exp(-exp(-all*(xl-mul)));

%/////Parameters for BoatADT/////
figure(4)
ksdensity(Peak_value)
hold on
plot(xl, yll, '-ro', 'markersize', 4)
legend('Observed', 'Lognormal pdf')
title('BoatA HT (HT=256)')
figure(5)
ksdensity(Peak_value)
hold on
plot(xl, yl2, '-ro', 'markersize', 4)
legend('Observed', 'Rayleigh pdf')
title('BoatA HT (HT=256)')
figure(6)
ksdensity(Peak_value)
hold on
plot(xl, yl3, '-ro', 'markersize', 4)
legend('Observed', 'Exponential pdf')
title('BoatA HT (HT=256)')
figure(7)
ksdensity(Peak_value)
hold on
plot(xl, yl4, '-ro', 'markersize', 4)
legend('Observed', 'Gumbel pdf')
title('BoatA HT (HT=256)')
figure(8)
ksdensity(Peak_value)
hold on
plot(xl, yl5, '-ro', 'markersize', 4)
legend('Observed', 'Normal pdf')
title('BoatA HT (HT=256)')
figure(9)
ksdensity(Peak_value)
hold on
plot(xl, lognpdf(xl, mean11, stdll), '-go', 'markersize', 4)
hold on
plot(xl, yl2, '-bo', 'markersize', 4)
plot(xl, yl3, '-ko', 'markersize', 4)
hold on
plot(xl, yl4, '-ro', 'markersize', 4)
legend('Observed', 'Lognormal', 'Rayleigh', 'Exponential', 'Gumbel')
title('BoatA HT (HT=256) PDF comparison')
%% CDF plot
figure(10)
ecdf(Peak_value)
hold on
plot(xl,logncdf(xl,meanll,stdll),'-ro','markersize',4)
hold on
% plot(xl,1/2*(1+erf((log(xl)-meanll)/(sqrt(2)*stdll))),'-bd','markersize',4)
legend('Observed','Lognormal Matlab','Lognormal equation')
title('BoatA HT (HT=256)')
figure(11)
ecdf(Peak_value)
hold on
plot(xl,expcdf(xl,lamdal),'-ro','markersize',4)
legend('Observed','Exponential cdf')
title('BoatA HT (HT=256)')
figure(12)
ecdf(Peak_value)
hold on
plot(xl,raylcdf(xl,alphal),'-ro','markersize',4)
legend('Observed','Rayleigh cdf')
title('BoatA HT (HT=256)')
figure(13)
ecdf(Peak_value)
hold on
plot(xl,y24,'-ro','markersize',4)
legend('Observed','Gumbel cdf')
title('BoatA HT (HT=256) CDF comparison')

%% KS test
% cdf of lognormal distribution according to the inputdata
y31=logncdf(sort(Peak_value),meanll,stdll);
% cdf of Rayleigh distribution according to the inputdata
y32=raylcdf(sort(Peak_value),alphal);
% cdf of Exponential distribution according to the inputdata
y33=expcdf(sort(Peak_value),lamdal);
% cdf of gumbel distribution according to the inputdata
y34=exp(-exp(-all*(sort(Peak_value)-mul)));
for i=1:length(Peak_value)
    E_cdf(i)=i/length(Peak_value);
end
Dn_L(1)=max(abs(y31-E_cdf));
Dn_R(1)=max(abs(y32-E_cdf));
Dn_E(1)=max(abs(y33-E_cdf));
Dn_G(1)=max(abs(y34-E_cdf));
% maximum different value of CDF for 1/3, 1/10, one-hundredth highest
for i=1:length(H_nth)
    Dn_L(i+1)=max(abs(y31(ceil((1-1/H_nth(i))*length(Peak_value))):end)-E_cdf(ceil((1-1/H nth(i))*length(Peak_value))):end))
Dn_R(i+1)=max(abs(y32(ceil((1-1/H_nth(i))*length(Peak_value))):end)-E_cdf(ceil((1-1/H_nth(i))*length(Peak_value))):end)
Dn_E(i+1)=max(abs(y33(ceil((1-1/H_nth(i))*length(Peak_value))):end)-E_cdf(ceil((1-1/H_nth(i))*length(Peak_value))):end)
Dn_G(i+1)=max(abs(y34(ceil((1-1/H_nth(i))*length(Peak_value))):end)-E_cdf(ceil((1-1/H_nth(i))*length(Peak_value))):end))
end

%% monte calo simulation
samples=1000000;
% H_nth=[3 10 100];
% Parameters for Lognormal distribution
X_log=exp(erfinv(2*rand(samples,1)-1)*sqrt(2)*stdll+meanll);
% X_log=exp(erfinv(2*(2*rand(samples,1)-1))*2*stdll+meanll);
Xsort_log=sort(X_log);
Mean_log=mean(X_log);
std_log=std(X_log);
%Parameters for Rayleigh distribution
alphal=(meanl/sqrt(pi/2)+stdl/sqrt(2-pi/2))/2;
X_ray=sqrt(-2*log(1-rand(samples,1)))*alphal;
Xsort_ray=sort(X_ray);
Mean_ray=mean(X_ray);
std_ray=std(X_ray);
%Parameters for Exponential distribution
X_exp=-log(1-rand(samples,1))/lamdal;
Xsort_exp=sort(X_exp);
Mean_exp=mean(X_exp);
std_exp=std(X_exp);
%Parameters for Gumbel distribution
X_gum=-log(-log(rand(samples,1)))/al+mul;
Xsort_gum=sort(X_gum);
Mean_gum=mean(X_gum);
std_gum=std(X_gum)
for i=1:length(H_nth)
    Hn_log(i)=exp(erfinv((H_nth(i)-2)/H_nth(i))*sqrt(2)*stdll+meanll)
    Hn_log_direct(i)=Xsort_log(ceil((1-1/H_nth(i))*samples))
    aaa=0.147;
aal=2/pi/aaa;
aa2=log(4*(H_nth(i)-1)/H_nth(i)^2);
    Hn_log_theo(i)=exp(meanll+sqrt(2)*stdll*sqrt(-aal-aa2/2+sqrt((aal+aa2/2)^2-aal*aa2)))
    X_sum_log=0;
    for j=1:samples
        if X_log(j)>=Hn_log(i)
            X_sum_log=X_sum_log+X_log(j);
        end
    end
    H_avg_log(i)=H_nth(i)*X_sum_log/samples
    H_avg_log_direct(i)=mean(Xsort_log(ceil((1-1/H_nth(i))*samples)));
end
%Rayleigh distribution
Hn_ray(i)=alphal*sqrt(2*log(H nth(i)))
Hn_ray_direct(i)=Xsort_ray(ceil((1-1/H_nth(i))*samples))
% num_ray=sum(X ray>=Hn ray);
X_sum_ray=0;
for j=1:samples
    if X ray(j)>=Hn_ray(i)
\[ X_{\text{sum\_ray}} = X_{\text{sum\_ray}} + X_{\text{ray}(j)}; \]
\end
\]
\[ H_{\text{avg\_ray}}(i) = H_{\text{nth}(i)} \times X_{\text{sum\_ray}} / \text{samples} \]
\[ H_{\text{avg\_ray\_theo}}(i) = \]
\[ H_{\text{nth}(i)} \times \alpha_1 \times \sqrt{2} \times \left( \frac{\sqrt{\pi}}{2} \left( \text{erf}(\log(H_{\text{nth}(i)})) - 1 \right) - \frac{\sqrt{\log(H_{\text{nth}(i)})}}{H_{\text{nth}(i)}} \right) \]
\[ H_{\text{avg\_ray\_direct}}(i) = \text{mean}(X_{\text{sort\_ray}}(\text{ceil}((1 - 1/H_{\text{nth}(i)}) \times \text{samples})); \text{end}) \]
\% Exponential distribution
\[ H_{n_{\text{exp}}}(i) = 1/\lambda_1 \times \log(H_{\text{nth}(i)}) \]
\[ H_{\text{exp\_direct}}(i) = \text{Xsort\_log}(\text{ceil}((1 - 1/H_{\text{nth}(i)}) \times \text{samples})) \]
\% num\_exp = \text{sum}(X_{\text{exp}} \geq H_{n_{\text{exp}}});
\[ X_{\text{sum\_exp}} = 0; \]
\% for \ j = 1: \text{samples}
\[ \text{if } X_{\text{exp}}(j) \geq H_{n_{\text{exp}}}(i) \]
\[ X_{\text{sum\_exp}} = X_{\text{sum\_exp}} + X_{\text{exp}}(j); \]
\end
\% for \ j = 1: \text{samples}
\[ H_{\text{avg\_exp\_theo}}(i) = 1/\lambda_1 \times (1 + \log(H_{\text{nth}(i)})) \]
\% Gumbel distribution
\[ H_{n_{\text{gum}}}(i) = -\log(-\log(1 - l/H_{\text{nth}(i)}))/\alpha_2 + \mu_1 \]
\[ H_{\text{gum\_direct}}(i) = \text{Xsort\_log}(\text{ceil}((1 - 1/H_{\text{nth}(i)}) \times \text{samples})) \]
\% num\_gum = \text{sum}(X_{\text{gum}} \geq H_{n_{\text{gum}}});
\[ X_{\text{sum\_gum}} = 0; \]
\% for \ j = 1: \text{samples}
\[ \text{if } X_{\text{gum}}(j) \geq H_{n_{\text{gum}}}(i) \]
\[ X_{\text{sum\_gum}} = X_{\text{sum\_gum}} + X_{\text{gum}}(j); \]
\end
\% Histogram and frequency diagram plot
\[ \text{NN} = \text{round}(1 + 3.3 \times \log(10(\text{length(Peak\_value))))); \]
\[ \text{binwidth} = (\text{max(Peak\_value)} - \text{min(Peak\_value)}) / \text{NN}; \]
\[ [\text{nhist\_histbin}] = \text{hist(Peak\_value,NN)}; \]
\[ \text{figure}(15) \]
\[ \text{bar(histbin, nhist/length(Peak\_value)/binwidth)} \]
\[ \text{Areacheck} = \text{sum(nhist/length(Peak\_value)/binwidth)*binwidth} \]
\hold on
\[ \text{plot(xl, yl, 'r', 'lineWidth', 2.5)} \]
\[ \text{legend('Hist', 'Gumbel pdf')} \]
\[ \text{title('Peaks value selected by HT Method (HT=256)' )} \]
\[ \text{figure}(16) \]
\text{hist(Peak\_value,NN)}
\% different method to generate binwidth
\[ \text{binwidthl} = 2 * \text{iqr(Peak\_value)/length(Peak\_value)}^{(1/3)} \]
\[ \text{binsl} = \text{min(Peak\_value)} : \text{binwidthl} : \text{max(Peak\_value)}; \]
\[ [\text{nhistl histbinsl}] = \text{hist(Peak\_value,binsl)}; \]
\[ \text{figure}(17) \]
\[ \text{bar(histbinsl, nhistl/length(Peak\_value)/binwidthl)} \]
\[ \text{Areacheckl} = \text{sum(nhistl/length(Peak\_value)/binwidthl)*binwidthl} \]
hold on
plot(xl,lognpdf(xl,meanll,stdll),'-go','markersize',4)
hold on
plot(xl,yl2,'-bo','markersize',4)
hold on
plot(xl,yl3,'-ko','markersize',4)
hold on
plot(xl,yl4,'-r','lineWidth',2.5)
legend('Hist','Lognormal','Rayleigh','Exponential','Gumbel pdf')
title('Peaks value selected by HT Method(HT=256)');

%% Chi-square test
% interval determined using 75% percentage
P1_start=histbins1-binwidth1/2;
P1_end=histbins1+binwidth1/2;
% L1_th: theoretical results for lognormal, E1_th, R1_th, G1_th
L1_th=(logncdf(P1_end,meanll,stdll) -
    logncdf(P1_start,meanll,stdll))*length(Peak_value);
E1_th=(expcdf(P1_end, lamdal) -
    expcdf(P1_start, lamdal))*length(Peak_value);
R1_th=(raylcdf(P1_end, alphal) -
    raylcdf(P1_start, alphal))*length(Peak_value);
G1_th=(exp(-exp(-all*{P1_end-mul})) -
    exp(-exp(-all*{P1_start-mul})))'*length(Peak_value);

% L1_cs: Chi-Square for lognormal
L1_cs=sum((nhist1-L1_th).*2./L1_th)
E1_cs=sum((nhist1-E1_th).*2./E1_th)
R1_cs=sum((nhist1-R1_th).*2./R1_th)
G1_cs=sum((nhist1-G1_th).*2./G1_th)
% interval determined using the method in the textbook
P2_start=histbins1-binwidth1/2;
P2_end=histbins1+binwidth1/2;
% L2_th: theoretical results for lognormal, E2_th, R2_th, G2_th
L2_th=(logncdf(P2_end,meanll,stdll) -
    logncdf(P2_start,meanll,stdll))*length(Peak_value);
E2_th=(expcdf(P2_end, lamdal) -
    expcdf(P2_start, lamdal))*length(Peak_value);
R2_th=(raylcdf(P2_end, alphal) -
    raylcdf(P2_start, alphal))*length(Peak_value);
G2_th=(exp(-exp(-all*{P2_end-mul})) -
    exp(-exp(-all*{P2_start-mul})))'*length(Peak_value);

% L2_cs: Chi-Square for lognormal
L2_cs=sum((nhist2-L2_th).*2./L2_th)
E2_cs=sum((nhist2-E2_th).*2./E2_th)
R2_cs=sum((nhist2-R2_th).*2./R2_th)
G2_cs=sum((nhist2-G2_th).*2./G2_th)
Dn_E=Dn_E'
Dn_R=Dn_R'
Dn_L=Dn_L'
Dn_G=Dn_G'
D_S=[ meanl;stdl;PC_mean']
D_E=[ Mean_exp;std_exp;H_avg_exp']
D_R=[ Mean_ray;std_ray;H_avg_ray']
D_L=[ Mean_log;std_log;H_avg_log']
D_G=[ Mean_gum;std_gum;H_avg_gum']
Data_Mon=[ D_S D_E D_R D_L D_G]
Data_CDF=[ Dn_E Dn_R Dn_L Dn_G]
APPENDIX B: Four Moments of Test Cases
First Four Moments of Parent Data Sets for Full Scale Test Cases

<table>
<thead>
<tr>
<th>Test Case</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Case 1</td>
<td>0.00</td>
<td>0.22</td>
<td>0.06</td>
<td>0.29</td>
</tr>
<tr>
<td>Test Case 2</td>
<td>0.00</td>
<td>0.14</td>
<td>0.06</td>
<td>0.22</td>
</tr>
<tr>
<td>Test Case 3</td>
<td>0.00</td>
<td>0.25</td>
<td>0.09</td>
<td>0.44</td>
</tr>
<tr>
<td>Test Case 4</td>
<td>0.00</td>
<td>0.36</td>
<td>0.54</td>
<td>1.64</td>
</tr>
<tr>
<td>Test Case 5</td>
<td>0.00</td>
<td>0.48</td>
<td>0.59</td>
<td>2.56</td>
</tr>
<tr>
<td>Test Case 6</td>
<td>0.00</td>
<td>0.07</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>Test Case 7</td>
<td>0.00</td>
<td>41.86</td>
<td>284.15</td>
<td>11775.27</td>
</tr>
<tr>
<td>Test Case 8</td>
<td>0.00</td>
<td>0.47</td>
<td>0.51</td>
<td>2.44</td>
</tr>
<tr>
<td>Test Case 9</td>
<td>0.00</td>
<td>0.46</td>
<td>0.37</td>
<td>1.73</td>
</tr>
<tr>
<td>Test Case 10</td>
<td>0.00</td>
<td>0.34</td>
<td>0.23</td>
<td>1.25</td>
</tr>
<tr>
<td>Test Case 11</td>
<td>0.00</td>
<td>0.30</td>
<td>0.12</td>
<td>0.51</td>
</tr>
<tr>
<td>Test Case 12</td>
<td>0.00</td>
<td>0.17</td>
<td>0.05</td>
<td>0.13</td>
</tr>
<tr>
<td>Test Case 13</td>
<td>0.00</td>
<td>0.20</td>
<td>0.03</td>
<td>0.14</td>
</tr>
<tr>
<td>Test Case 14</td>
<td>0.00</td>
<td>0.12</td>
<td>0.03</td>
<td>0.18</td>
</tr>
<tr>
<td>Test Case 15</td>
<td>0.00</td>
<td>0.08</td>
<td>0.01</td>
<td>0.06</td>
</tr>
<tr>
<td>Test Case 16</td>
<td>0.00</td>
<td>0.27</td>
<td>0.16</td>
<td>0.50</td>
</tr>
<tr>
<td>Test Case 17</td>
<td>0.00</td>
<td>0.26</td>
<td>0.14</td>
<td>0.39</td>
</tr>
<tr>
<td>Test Case 18</td>
<td>0.00</td>
<td>2.59</td>
<td>3.01</td>
<td>64.12</td>
</tr>
<tr>
<td>Test Case 19</td>
<td>0.00</td>
<td>0.04</td>
<td>0.00</td>
<td>0.01</td>
</tr>
</tbody>
</table>

First Four Moments of Parent Data Sets for Model Scale Test Cases

<table>
<thead>
<tr>
<th>Test Case</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test Case 1</td>
<td>0.00</td>
<td>0.37</td>
<td>0.64</td>
<td>4.13</td>
</tr>
<tr>
<td>Test Case 2</td>
<td>0.00</td>
<td>0.61</td>
<td>1.14</td>
<td>5.68</td>
</tr>
<tr>
<td>Test Case 3</td>
<td>0.00</td>
<td>2.09</td>
<td>8.18</td>
<td>95.01</td>
</tr>
<tr>
<td>Test Case 4</td>
<td>0.00</td>
<td>0.19</td>
<td>0.15</td>
<td>0.46</td>
</tr>
<tr>
<td>Test Case 5</td>
<td>0.00</td>
<td>0.39</td>
<td>0.60</td>
<td>2.73</td>
</tr>
<tr>
<td>Test Case 6</td>
<td>0.00</td>
<td>1.64</td>
<td>6.43</td>
<td>90.13</td>
</tr>
<tr>
<td>Test Case 7</td>
<td>0.00</td>
<td>0.22</td>
<td>0.08</td>
<td>0.22</td>
</tr>
<tr>
<td>Test Case 8</td>
<td>0.00</td>
<td>0.24</td>
<td>0.11</td>
<td>0.31</td>
</tr>
<tr>
<td>Test Case 9</td>
<td>0.00</td>
<td>0.68</td>
<td>0.98</td>
<td>5.88</td>
</tr>
</tbody>
</table>
VITA

The author was born in 1976 to G. Michael and Roberta Grimsley. She graduated from Denbigh High School in Newport News, Virginia in 1994. That year she enrolled at Webb Institute in Glen Cove, New York. She graduated from Webb Institute in 1998 with a Bachelor of Science in Naval Architecture and Marine Engineering. In 1999 she completed her Master of Science in Ocean Systems Management from Massachusetts Institute of Technology. For ten years she worked for the United States (U.S.) Navy’s Combatant Craft Division where she was responsible for research and development of combatant craft technologies and special programs for the U.S. Department of Defense. She currently works in industry providing technical direction and oversight in the development of advanced maritime systems for U.S. and Coalition Special Forces. She is the 1998 recipient of the Society of Naval Architects and Marine Engineers (SNAME) Paper of the Year Award, the 2006 recipient of the American Society of Naval Engineers (ASNE) Lester Rosenblatt Young Naval Engineer Award, and the 2007 recipient of the U.S. Navy’s Rear Admiral Taylor Award for Outstanding Scientific Contributions. Ms. Grimsley is a licensed, Professional Engineer in the state of Virginia and a Member of ASNE, SNAME and the Royal Institute of Naval Architects (RINA).