

CHARACTERIZATION OF VEHICLE TIME HEADWAY IN CLEAR AND RAINY  
WEATHER

by

Atousa Tangestanipour  
BA, Shiraz University, 2005  
MSc, Alghadir University, 2008

A Thesis Submitted in Partial Fulfillment  
of the Requirements for the Degree of  
MASTER OF APPLIED SCIENCE

in the Department of Electrical and Computer Engineering

© Atousa Tangestanipour, 2018  
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy  
or other means, without the permission of the author.

CHARACTERIZATION OF VEHICLE TIME HEADWAY IN CLEAR AND RAINY  
WEATHER

by

Atousa Tangestanipour

BA, Shiraz University, 2005

MSc, Alghadir University, 2008

**Supervisory Committee**

Dr. T. Aaron Gulliver, Supervisor  
Department of Electrical and Computer Engineering

Dr. A. Baniasadi, Committee member  
Department of Electrical and Computer Engineering

## Abstract

Adverse weather has a direct effect on traffic congestion, capacity, and road safety. Rain can influence traffic flow and headway. Thus, it is important to study the impact of weather conditions on traffic. In this thesis, headway data from a north-south highway in Tehran is categorized according to weather conditions and traffic flow. A statistical analysis of this data is presented which show that the mean time headway increases in rainy weather and the traffic flow rate is lower. The probability density functions fit to the headway data and these distributions are then evaluated using the Chi-Squared (C-S) and Kolmogorov-Smirnov (K-S) tests to determine which is the most suitable. To generalize results for different traffic flow rates, another highway was selected for data collection. The results obtained show that the Burr distribution is the best to model the headway data in clear and rainy weather with a higher traffic flow. Moreover, the generalized extreme value distribution is the best to model the headway data in clear and rainy weather with a lower traffic flow. This justifies the use of different time headway distributions in rainy and clear weather conditions and different traffic flow rates for traffic modeling.

# Contents

|   |     |
|---|-----|
| Supervisory Committee.....                    | ii  |
| Abstract.....                                 | iii |
| Contents.....                                 | iv  |
| List of Tables.....                           | vi  |
| List of Figures.....                          | vii |
| Acknowledgments.....                          | ix  |
| 1 Introduction                                | 1   |
| 1.1 Motivation and Background .....           | 1   |
| 1.2 Thesis Organization.....                  | 4   |
| 2 Data Collection and Analysis                | 6   |
| 2.1 Data Classification.....                  | 8   |
| 3 Headway Statistical Distributions           | 13  |
| 3.1 Distribution Parameter Estimation .....   | 21  |
| 3.2 Goodness of Fit .....                     | 28  |
| 3.2.1 The Kolmogorov-Smirnov (K-S) Test ..... | 46  |
| 3.2.2 The Chi-Squared (C-S) Test .....        | 47  |
| 4 Analysis of New Clear Weather Data          | 53  |
| 4.1 Distribution Parameter Estimation .....   | 56  |
| 4.2 Goodness of Fit .....                     | 59  |
| 5 Conclusion                                  | 68  |
| 5.1 Future Work.....                          | 69  |
| Bibliography                                  | 70  |



## List of Tables

|   |    |
|---|----|
| Table 2.1: Dataset parameters.....  | 9  |
| Table 2.2: Statistical characteristics of the headway data .....              | 10 |
| Table 2.3: The percentage of vehicles with a headway less than 2 s .....      | 12 |
| Table 3.1: Statistical distribution parameters for the headway data .....     | 26 |
| Table 3.2: Goodness of fit test results for the headway data .....            | 48 |
| Table 4.1: The new dataset parameters.....                                    | 55 |
| Table 4.2: Statistical characteristics of the new headway data .....          | 56 |
| Table 4.3: Statistical distribution parameters for the new headway data ..... | 58 |
| Table 4.4: Goodness of fit test results for the new headway data.....         | 66 |
| Table A.1: Critical values for the C-S test statistic.....                    | 74 |

## List of Figures

|  |    |
|--|----|
| Figure 2.1: The location of the data collection point on Google maps.....  | 7  |
| Figure 2.2: An image taken from the video data showing the two reference points.....   | 7  |
| Figure 2.3: The cumulative probability distributions for the eight headway datasets.....   | 12 |
| Figure 3.1: The PDF of the Burr distribution with shape parameters 2 and 5 and different scale parameters .....  | 14 |
| Figure 3.2: The PDF of the generalized extreme value distribution with scale and shape parameters 1 and 0, respectively, and different location parameters ..... | 14 |
| Figure 3.3: The PDF of the Burr distribution with scale parameter 1 and different shape parameters.....  | 15 |
| Figure 3.4: The PDFs of all distributions fit to the N1 dataset.....   | 22 |
| Figure 3.5: The PDFs of all distributions fit to the N2 dataset.....   | 22 |
| Figure 3.6: The PDFs of all distributions fit to the N3 dataset.....   | 23 |
| Figure 3.7: The PDFs of all distributions fit to the N4 dataset.....   | 23 |
| Figure 3.8: The PDFs of all distributions fit to the R1 dataset.....   | 24 |
| Figure 3.9: The PDFs of all distributions fit to the R2 dataset.....   | 24 |
| Figure 3.10: The PDFs of all distributions fit to the R3 dataset.....  | 25 |
| Figure 3.11: The PDFs of all distributions fit to the R4 dataset.....  | 25 |
| Figure 3.12: The Q-Q plots for the N1 dataset.....   | 31 |
| Figure 3.13: The Q-Q plots for the N2 dataset.....   | 33 |
| Figure 3.14: The Q-Q plots for the N3 dataset.....   | 35 |
| Figure 3.15: The Q-Q plots for the N4 dataset.....   | 37 |
| Figure 3.16: The Q-Q plots for the R1 dataset.....   | 39 |

|   |    |
|---|----|
| Figure 3.17: The Q-Q plots for the R2 dataset.....  | 41 |
| Figure 3.18: The Q-Q plots for the R3 dataset.....  | 43 |
| Figure 3.19: The Q-Q plots for the R4 dataset.....  | 45 |
| Figure 4.1: The location of the new data collection point on Google maps.....                       | 53 |
| Figure 4.2: The locations on the two highways on Google maps where headway data was collected ..... | 54 |
| Figure 4.3: An image taken from the video data showing the reference point.....                     | 54 |
| Figure 4.4: The PDFs of all distributions for the T1 dataset.....                                   | 57 |
| Figure 4.5: The PDFs of all distributions for the T2 dataset.....                                   | 57 |
| Figure 4.6: The PDFs of all distributions for the T3 dataset.....                                   | 58 |
| Figure 4.7: The Q-Q plots for the T1 dataset.....   | 61 |
| Figure 4.8: The Q-Q plots for the T2 dataset.....   | 63 |
| Figure 4.9: The Q-Q plots for the T3 dataset.....   | 65 |

## **Acknowledgments**

I would like to thank my supervisor, Dr. Gulliver, for giving me the opportunity to work under his supervision in the field of traffic systems, and for his useful feedback and patience during my studies. I would also like to extend my gratitude to Dr. Baniasadi who served on my supervisory committee and provided useful feedback on my thesis. I am also indebted to Dr. Phalguni Mukhopadhyaya for serving as external examiner.

I would like to express my special thanks to Dr. Ghanbarnejad for his valuable feedback and Mr. Valibakht from the United Bus Company of Tehran for providing the headway data without any hesitation. It would not have been possible to complete this research without their support.

Special thanks and love to my family members, especially my husband who emotionally supported me during my education and my parents.

# Chapter 1

## Introduction

### 1.1 Motivation and background

Congestion on a road occurs when demand exceeds the capacity of road facilities. Traffic congestion is increasing in every major metropolitan area. Consequently, traffic modeling has become an important field of research in solving traffic engineering problems. There are two main types of traffic models based on the level of detail: macroscopic and microscopic. Macroscopic models consider the average traffic flow and related parameters such as speed and density. Microscopic models consider individual vehicles and their interactions. These models characterize vehicle behaviour on a road and focus on small changes in the traffic flow over time and space. This requires collecting data for individual vehicles [1].

Headway is one of the vehicle parameters used in traffic modeling. There are two types of headway: time headway and distance headway. Distance headway is usually called spacing, and time headway is referred to here as just headway. Headway is defined as the time between two consecutive vehicles and is the difference in time when the front bumpers pass a reference point on a road in a lane [2]. If the  $i$ th vehicle ( $x_i$ ) passes the reference point at time  $t_{x_i}$  and the next vehicle ( $x_{i+1}$ ) passes this point at time  $t_{x_{i+1}}$ , the headway is

$$h_{x_i} = t_{x_{i+1}} - t_{x_i} \quad (1.1)$$

and the distance headway is

$$d_{x_i} = h_{x_i} \times s_{x_i} \quad (1.2)$$

where  $s_{x_i}$  is the  $i$ th vehicle speed. Traffic engineers typically use headway data for analysis because it is easier to collect than spacing data.

On highways with free flow traffic, the capacity is defined as the maximum rate at which vehicles traverse a specified segment of the roadway. This depends on the road and traffic conditions, and is related to the minimum headway. In fact, there is an inverse relationship between headway and highway capacity. A qualitative measure of traffic flow is the level of service [2]. There are six levels, A to F ranging from the best quality of traffic when it is free flow to the worst quality when vehicles have very low speeds. At higher service levels, vehicles have more freedom to choose their speed and headway and to change lanes. Headway is an important indicator of traffic congestion, level of service, and road capacity [3].

The complexity of real-world traffic behaviour makes simulation an efficient way to analyze traffic [4]. It is used to plan and improve traffic conditions based on models of real traffic data [5]. Parameters such as speed and headway are required to accurately simulate traffic flow. For instance, the effect of road width and weather on the interaction between drivers has been determined based on the car type, vehicle speed, and headway [6]. The distributions of these values can be determined from headway statistics.

Driving behaviour depends on laws and their implementation, culture, social behaviour, and road conditions. Therefore, traffic statistics differ depending on where and when the data has been obtained. The influence of weather and type of vehicles on traffic flow rates

can be quantified by collecting data and developing probabilistic distributions to characterize parameters such as the headway, traffic flow, and speed [7]. These models can be used to characterize the effect of vehicle size, time of day, visibility, and weather conditions on the headway and traffic [8].

Rainfall causes reduced visibility which increases headway [9]. A study in Putrajaya, Malaysia showed that the speed is reduced by 4.2% from dry weather to wet weather [10]. A study of vehicle headway during rainfall at night in Pontian, Malaysia showed that the mean headway increases with rainfall intensity. The headway increased by 3.39% in light rain, and 16.52% in moderate rain, while there was a 46.83% increase in heavy rain [11]. Another study of a principal road in Johor Bahru, Malaysia indicated that the mean headway was reduced by 15.66% in light rain. In addition, there was a 19.97% headway reduction in medium rain, while the reduction in heavy rain was 25.65% [7].

Several statistical distributions have been proposed to describe headway under different traffic flow conditions. The negative exponential, shifted exponential, and Erlang distributions have been used to characterize traffic flow [12]. Furthermore, a lognormal distribution was used for low traffic flow and a log-logistic distribution for high traffic flow [13]. A semi-Poisson distribution was used to model headway in normal traffic flow conditions on a two-lane road in Dunedin, New Zealand [14]. Shifted gamma and shifted lognormal distributions have been used to model headway data from different freeway sections such as a basic section without any interruptions in traffic flow, a ramp merge, a lane drop, and a ramp weave [15]. A double displaced negative exponential distribution was proposed in [16] to model headway in urban areas.

The effect of different times on the headway has been examined for High Occupancy Vehicle (HOV) and general purpose lanes which are used by all vehicles without any restrictions [17]. A double displaced negative exponential distribution was proposed to model headway data in HOV lanes with high traffic flow rates, and a shifted lognormal distribution was used to model headway in general purpose lanes [18].

The effect of adverse weather on the headway has also been analyzed. The Burr probability distribution was used in [7] to model headway for different rainfall intensities. Further, freeway traffic flow under different rain conditions was studied in [19] to explore the effect of rainfall intensity on traffic speed. It was found that traffic speed is inversely proportional to the rain intensity in free flow traffic.

In this thesis, headway under clear and rainy weather conditions is studied using data collected from two highways in Tehran, Iran. Several statistical distributions are considered for headway to determine which are the best. The distribution parameters are obtained and three goodness of fit tests are used to evaluate and compare the distributions.

## **1.2 Thesis Organization**

Chapter 1 provided an introduction to headway and its importance in traffic flow modeling. Based on previous studies, the factors which influence the headway were presented. The statistical distributions that have been used to model headway in different conditions were presented.

Chapter 2 explains the collection of headway data and the corresponding statistical parameters. This data is categorized into eight datasets according to weather conditions,

traffic flow rates, and average speeds. Statistical characteristics for each dataset are determined.

In Chapter 3, the probability distributions considered in this thesis are presented. The methodology and selection of appropriate distributions are explained. The distribution parameters are determined by fitting the probability density functions to the data. These distributions are then evaluated using the Chi-Squared (C-S) and Kolmogorov-Smirnov (K-S) tests. The results of the goodness of fit tests are discussed in detail to determine the best distributions.

In Chapter 4, headway data from another highway is modeled to compare with the headway distributions obtained in Chapter 3. The goodness of fit tests are used to determine the best distributions.

Chapter 5 concludes the thesis and compares the headway distributions for the two different highways. A summary of the thesis contributions is given along with some suggestions for future work.

## Chapter 2

### Data Collection and Analysis

Tehran does not have many rainy days and has high air pollution levels. These conditions result in slippery roads during the first few hours of rain. Slippery roads and reduced visibility lead to reduced control which may increase accident risk. Understanding how weather conditions can affect headway is important in making policies and installing variable message signs on highways [10].

The headway data was obtained from the Emam Ali Highway, Tehran, Iran. It is an eight-lane highway with the two center lanes dedicated to the Bus Rapid Transit (BRT) system and emergency vehicles. The speed limit on this highway is 90 kilometers per hour (km/h). The location of the data collection point on Google maps is shown in Figure 2.1. This section of highway is free of emergency refuge areas, ramps, and bus stops as well as traffic lights and intersections, so there is free flow traffic.

Video was used to collect the headway data. The video camera belongs to the United Bus Company of Tehran. It is used to monitor bus lanes and detect cars using these lanes illegally. It is installed on the Kosar Bus Rapid Transit stop of BRT line 9. This camera also captures traffic on the highway moving in a north-south direction. The guardrail base plates serve as reference points to extract data as they can be easily distinguished in the video. These reference points are indicated in Figure 2.2.

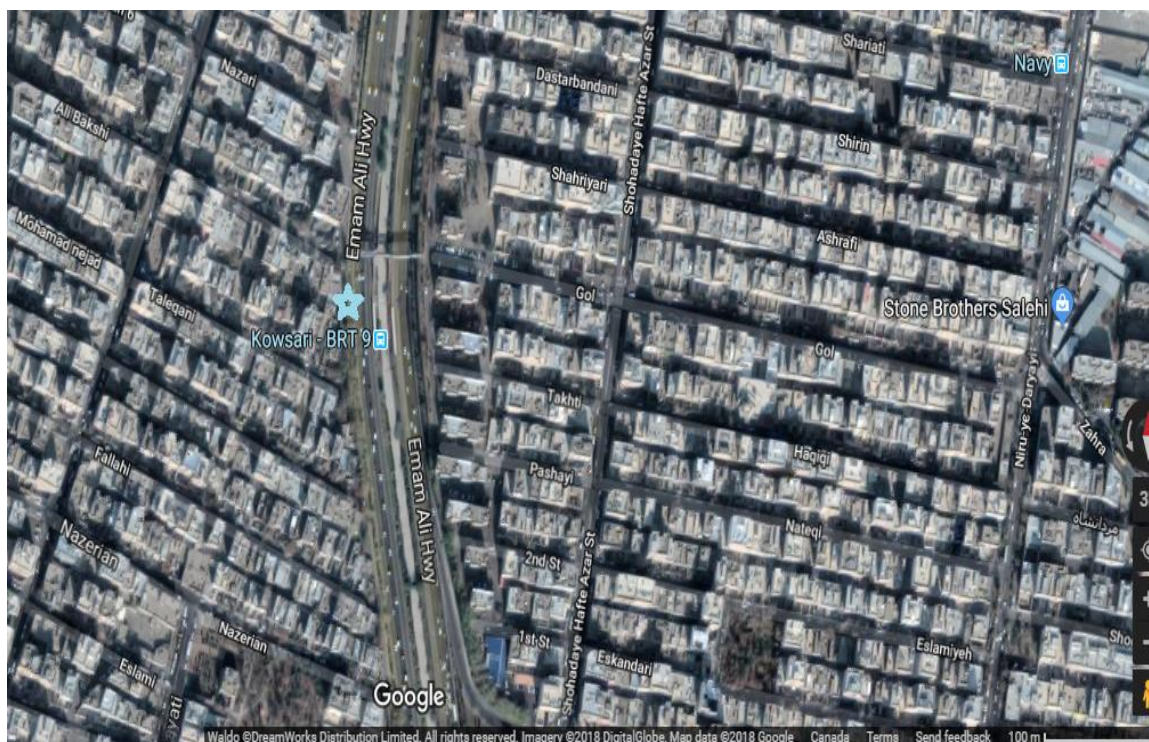


Figure 2.1: The location of the data collection point on Google maps [20].



Figure 2.2: An image taken from the video data showing the two reference points.

The video was recorded at a rate of 30 frames per second and the traffic data was extracted manually. Information such as the vehicle type, vehicle headway, and vehicle speed was obtained. The time headway was determined using the time when the front bumper of a vehicle crosses the first reference point. Vehicle speed was obtained using the two reference points on the highway which are six meters apart.

The video was recorded on four weekdays between 12:30 and 13:00 in the fall of 2017, two rainy days (November 5th and 11th) and two clear days (November 6th and 12th). It was recorded in 15 min segments. The data was filtered to remove unusual vehicle behaviour. For example, motorcycles and taxis often move parallel to each other in the right lane and sometimes taxis and automobiles stop to drop off or pick up passengers in the right lane as shown on the left in Figure 2.1. Thus, to obtain accurate data, headway data was extracted only from the two left lanes in one direction (north to south). Trucks over 3.5 tonnes are not allowed on this highway during the daytime. As a result, heavy vehicles had a minimal effect on the data during the observation times.

## 2.1 Data Classification

Both headway and speed data were collected. Each 15 min segment of video was divided into three 5 min sections for analysis. The average speed and traffic flow rate were obtained for each section. The traffic flow rate is defined as the number of vehicles per hour per lane (vphpl)

$$f_h = n_l \times 12 \quad (2.1)$$

where  $n_l$  is the number of vehicles per lane in the 5 min interval. Based on the average speeds and traffic flow rates obtained, the data was divided into eight datasets as indicated

in Table 2.1. The data was divided into two groups according to the weather, rainy or clear. For the clear weather data, there are two average speeds: 80 to 90 (km/h) and 70 to 80 (km/h). For the rainy weather data, there are also two average speeds: 80 to 90 (km/h) and 60 to 80 (km/h). The data was further classified based on traffic flow rates: less than 1,600 vphpl, between 1,600 and 1,900 vphpl, and greater than 1,900 vphpl for clear weather, and in rainy weather: less than 1,000 vphpl, between 1,000 and 1,300 vphpl, and greater than 1,300 vphpl.

Table 2.1: Dataset parameters

| <b>Dataset</b> | <b>Weather condition</b> | <b>Traffic flow rate<br/>(vphpl)</b> | <b>Average speed<br/>(km/h)</b> |
|----------------|--------------------------|--------------------------------------|---------------------------------|
| N1             | Clear                    | >1900                                | 80 to 90                        |
| N2             | Clear                    | Between 1600-1900                    | 80 to 90                        |
| N3             | Clear                    | <1600                                | 70 to 80                        |
| N4             | Clear                    | Between 1600-1900                    | 70 to 80                        |
| R1             | Rainy                    | >1300                                | 80 to 90                        |
| R2             | Rainy                    | Between 1000-1300                    | 80 to 90                        |
| R3             | Rainy                    | Between 1000-1300                    | 60 to 80                        |
| R4             | Rainy                    | <1000                                | 60 to 80                        |

The statistical parameters considered in this thesis are as follows.

### **Mean**

The mean is [21]

$$\mu = \frac{1}{n} \sum_{i=1}^n h_i \quad (2.2)$$

where  $n$  is the number of data values and  $h_i$  is the  $i$ th value.

### **Median**

Put the data  $h_i$  in ascending order  $h'_1 = \min h_i, h'_2, \dots, h'_{n-1}, h'_n = \max h_i$ . The median corresponds to the middle of the ordered values

$$m = h'_k, k = \frac{n+1}{2} \quad n \text{ odd} \quad (2.3)$$

$$m = \frac{1}{2} \left( h'_{\left(\frac{n}{2}\right)} + h'_{\left(\frac{n+1}{2}\right)} \right) \quad n \text{ even} \quad (2.4)$$

### Variance

The variance is defined as

$$var = \frac{1}{n-1} \sum_{i=1}^n (h_i - \mu)^2 \quad (2.5)$$

where  $\mu$  is the mean.

### Standard Deviation

The standard deviation is

$$s = \sqrt{var} \quad (2.6)$$

Table 2.2: Statistical characteristics of the headway data

| Dataset            | N1    | N2    | N3    | N4    | R1    | R2    | R3    | R4    |
|--------------------|-------|-------|-------|-------|-------|-------|-------|-------|
| <b>Sample size</b> | 693   | 948   | 869   | 563   | 650   | 603   | 583   | 254   |
| <b>Mean</b>        | 1.72  | 2.10  | 2.42  | 1.91  | 2.35  | 2.77  | 2.99  | 4.09  |
| <b>Median</b>      | 1.32  | 1.65  | 1.86  | 1.49  | 1.83  | 2.26  | 2.49  | 2.80  |
| <b>Variance</b>    | 1.83  | 2.04  | 3.23  | 2.18  | 3.16  | 3.18  | 3.80  | 12.29 |
| <b>Std Dev</b>     | 1.35  | 1.55  | 1.80  | 1.48  | 1.78  | 1.78  | 1.95  | 3.51  |
| <b>Max</b>         | 11.00 | 11.45 | 14.00 | 12.12 | 15.34 | 15.34 | 13.34 | 23.00 |
| <b>Min</b>         | 0.47  | 0.46  | 0.50  | 0.52  | 0.66  | 0.67  | 0.80  | 1.07  |

The number of headway values obtained was 5163. The number of values for each dataset is shown in Table 2.2 along with the corresponding statistical parameters. This shows that the median headway is smaller than the mean in all datasets and this difference decreases with increasing traffic flow rate. Thus, the data is more concentrated in the low values and is not symmetric.

In clear weather, the mean headway in heavy traffic flow (1600-2200 vphpl) is 1.72 s, whereas this value in light traffic flow (less than 1600 vphpl) is 2.42 s. In rainy weather, the mean headway is between 2.32 s and 4.09 s, while this value in clear weather is between 1.69 s and 2.42 s. This shows that drivers are more careful in rain as the headway is larger to decrease the probability of an accident. These results confirm that rainy weather has a considerable impact on headway.

Table 2.2 shows that the variance and standard deviation of headway increase with decreasing traffic flow rate. The safe headway is defined as 2 s in good visibility and dry road conditions [22]. Figure 2.3 presents the headway cumulative probability distributions. In clear weather, a large number of drivers adopt a headway lower than the safe headway because of the higher traffic flow and speed. During clear weather with a high traffic flow (N1 dataset), 80% of the vehicles have a headway less than 2 s. This percentage is the highest among the datasets, followed by N4 and N2 (lower traffic flow rates with 73% and 65% of the vehicles with a headway less than 2s, respectively). Table 2.3 shows that these percentages decrease with a reduction in traffic flow rate and a change in weather from clear to rainy. For example, 31% of the vehicles have a headway less than 2 s in rainy weather and a low traffic flow. Because the risk of an accident is higher with a small

headway, especially during peak hours, the need arises to use safety measures such as variable message signs on highways and following distance warning systems.

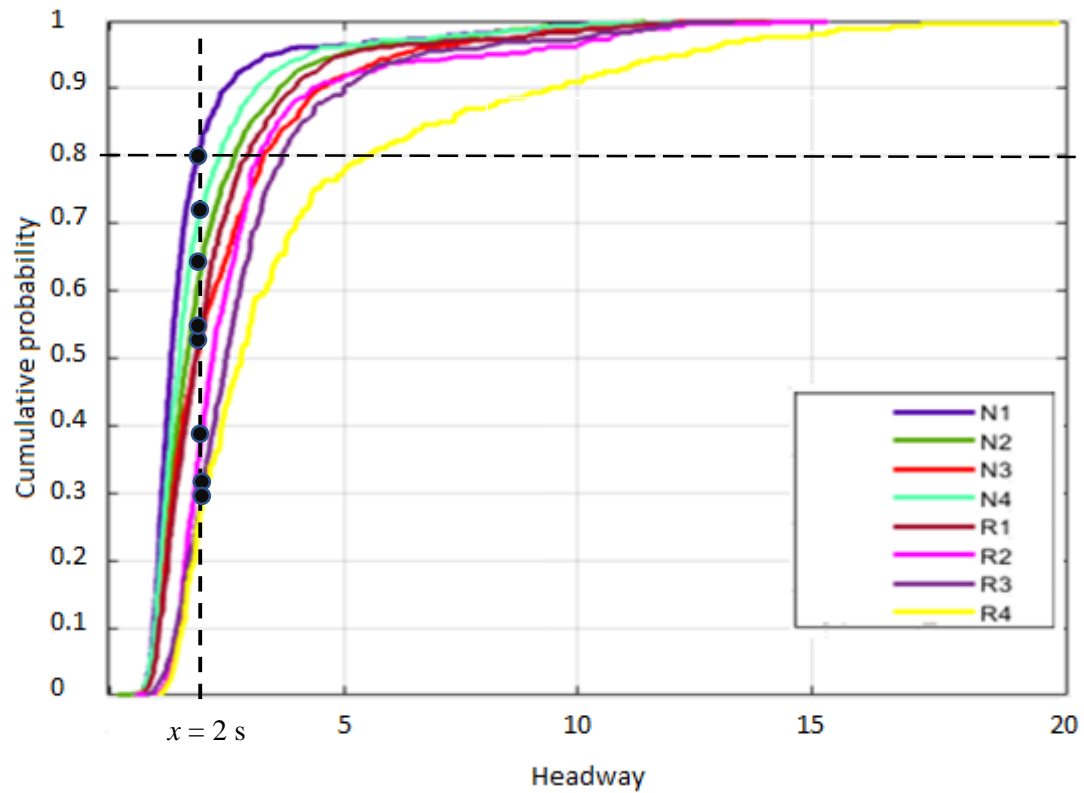


Figure 2.3: The cumulative probability distributions for the eight headway datasets.

Table 2.3: The percentage of vehicles with a headway less than 2 s.

| Dataset                | N1 | N2 | N3 | N4 | R1 | R2 | R3 | R4 |
|------------------------|----|----|----|----|----|----|----|----|
| Percentage of vehicles | 80 | 65 | 56 | 73 | 57 | 40 | 32 | 31 |

## Chapter 3

### Headway Statistical Distributions

Both single distributions and a mix of two or more distributions have been used to model headway [23]. The Probability Density Functions (PDFs) of mixed models are complicated and so parameter estimation is difficult [24]. Thus, single probability distributions are used in this thesis to model the headway data for both clear and rainy weather. Three goodness of fit tests are used to evaluate how well a distribution fits a dataset.

To be consistent with previous research, the following distributions are used here to model the headway data: Burr, generalized extreme value, logistic, lognormal, t location-scale, log-logistic, Weibull, exponential, extreme value, gamma, and normal. Some of these distributions have scale, location, and shape parameters. The scale parameter determines the dispersion or spread of the probability distribution. Figure 3.1 shows the PDF of the Burr distribution with the same shape parameter but different scale parameters. The distribution with a scale parameter of 1 is shown in green and is more dispersed than with a scale parameter of 10 which is shown in red. The location parameter determines the location or shift of a distribution. It indicates where the distribution is centered on the horizontal axis. Figure 3.2 shows the PDF of the generalized extreme value distribution with the same shape and scale parameters but different location parameters. The shape parameter affects the shape of a distribution. Figure 3.3 shows the PDF of the Burr distribution with different shape parameters. The distributions used to model the headway data are described below.

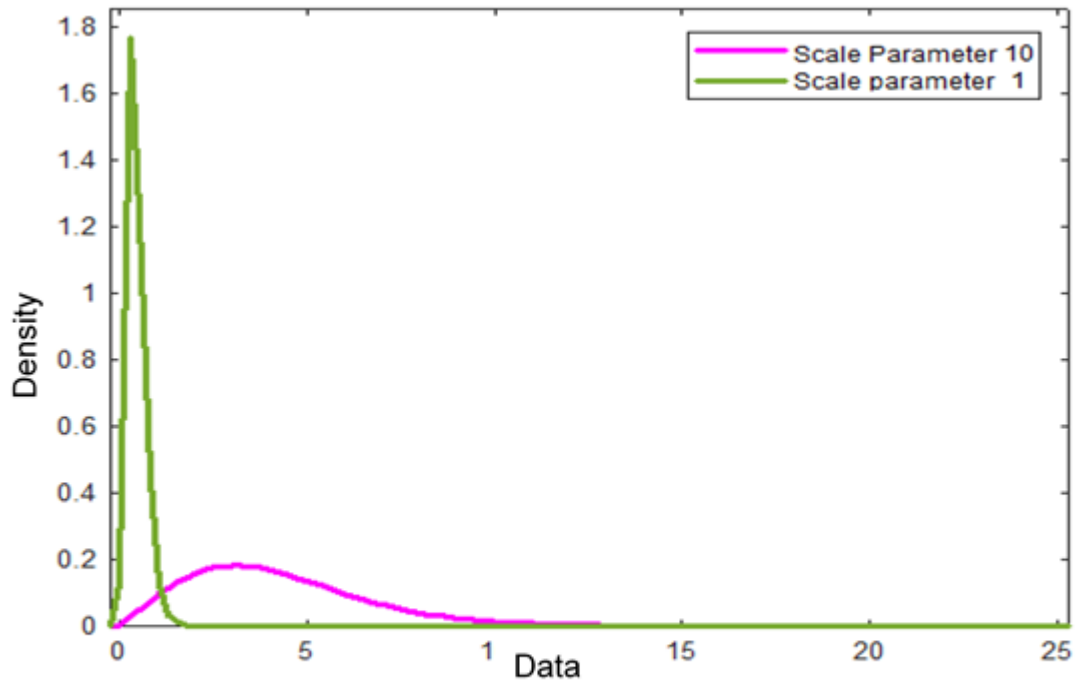


Figure 3.1: The PDF of the Burr distribution with shape parameters 2 and 5 and different scale parameters.

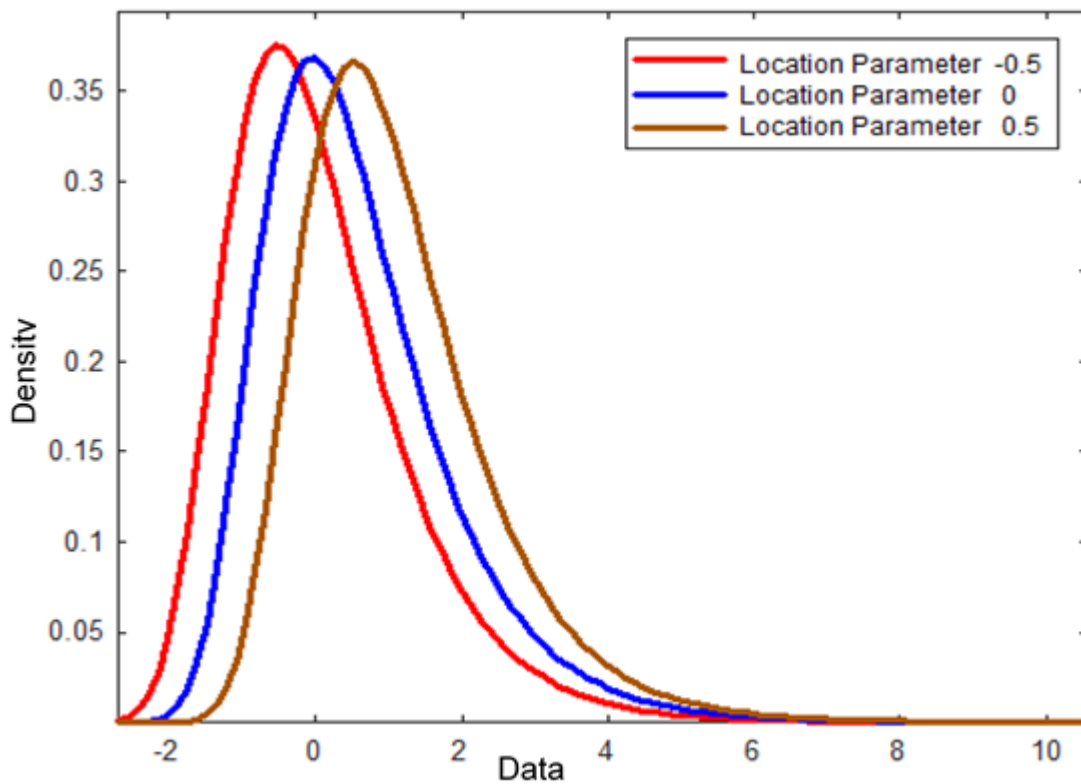


Figure 3.2: The PDF of the generalized extreme value distribution with scale and shape parameters 1 and 0, respectively, and different location parameters.

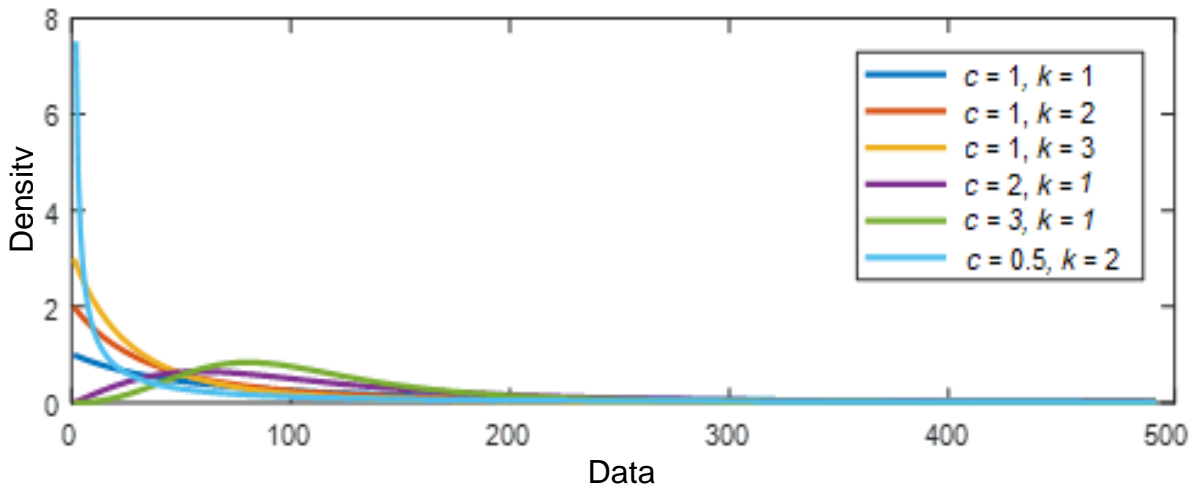


Figure 3.3: The PDF of the Burr distribution with scale parameter 1 and different shape parameters.

### Burr distribution

The Burr distribution has been used to analyze headway data and the effect of rainy weather on vehicle behaviour [7]. The PDF of the Burr distribution Type XII for non-negative  $x$  is [21]

$$f(x) = \frac{kc \left(\frac{x}{\alpha}\right)^{c-1}}{\left(1 + \left(\frac{x}{\alpha}\right)^c\right)^{k+1}} \quad (3.1)$$

where  $c$  and  $k$  are shape parameters and  $\alpha$  is the scale parameter. The Cumulative Distribution Function (CDF) of the Burr distribution is

$$F(x) = 1 - \frac{1}{\left(1 + \left(\frac{x}{\alpha}\right)^c\right)^k} \quad (3.2)$$

### Generalized extreme value distribution

The generalized extreme value distribution is a combination of the Gumbel, Fréchet, and Weibull distributions. In Type I,  $k = 0$ , in Type II  $k > 0$ , and in Type III  $k < 0$ . The PDF of the generalized extreme value distribution is [21]

$$f(x) = \left(\frac{1}{\alpha}\right) \exp\left(-\left(1 + k\frac{(x-\beta)}{\alpha}\right)^{\frac{-1}{k}}\right) \left(\left(1 + k\frac{(x-\beta)}{\alpha}\right)\right)^{-1-\frac{1}{k}} \quad k \neq 0 \quad (3.3)$$

$$f(x) = \left(\frac{1}{\alpha}\right) \exp\left(-\exp\left(\frac{-(x-\beta)}{\alpha}\right) - \frac{(x-\beta)}{\alpha}\right) \quad k = 0 \quad (3.4)$$

where  $k$  is the shape parameter,  $\beta$  is the location parameter, and  $\alpha$  is a non-negative scale parameter. The CDF of the generalized extreme value distribution is

$$F(X) = \exp\left(-\left(1 + k\frac{(x-\beta)}{\alpha}\right)^{\frac{-1}{k}}\right) \quad k \neq 0 \quad (3.5)$$

$$F(X) = \exp\left(-\exp\left(\frac{-(x-\beta)}{\alpha}\right)\right) \quad k = 0 \quad (3.6)$$

### Logistic distribution

The PDF of the logistic distribution is [21]

$$f(x) = \frac{\exp\left(\frac{x-\beta}{\alpha}\right)}{\alpha(1 + \exp\left(\frac{x-\beta}{\alpha}\right))^2} \quad (3.7)$$

where  $\beta$  is the location parameter and  $\alpha$  is a non-negative scale parameter. The CDF of the logistic distribution is

$$F(x) = \frac{\exp\left(\frac{x-\beta}{\alpha}\right)}{1 + \exp\left(\frac{x-\beta}{\alpha}\right)} \quad (3.8)$$

### Lognormal distribution

The lognormal distribution has been used to model headway data in non-peak hours [25].

The PDF of the lognormal distribution for non-negative  $x$  is [26]

$$f(x) = \frac{1}{xk\sqrt{2\pi}} \exp\left(\frac{-(\log(x/\alpha))^2}{2k^2}\right) \quad (3.9)$$

where  $k$  and  $\alpha$  are shape and scale parameters, respectively. The CDF of the lognormal distribution is

$$F(x) = \int_{-\infty}^x \frac{1}{xk\sqrt{2\pi}} \exp\left(\frac{-(\log(x/\alpha))^2}{2k^2}\right) du \quad (3.10)$$

### **t location-scale distribution**

This distribution is suitable for data with a large standard deviation [21]. The PDF of the t location-scale distribution is [26]

$$f(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\alpha\sqrt{k\pi}\Gamma\left(\frac{k}{2}\right)} \left[ \frac{k + \left(\frac{x-\beta}{\alpha}\right)^2}{k} \right]^{-\left(\frac{k+1}{2}\right)} \quad (3.11)$$

where  $\Gamma(\cdot)$  is the gamma function,  $k$  is the shape parameter,  $\alpha$  is the scale parameter, and  $\beta$  is the location parameter. The CDF of the t location-scale distribution is

$$F(x) = \int_{-\infty}^x \frac{\Gamma\left(\frac{k+1}{2}\right)}{\alpha\sqrt{k\pi}\Gamma\left(\frac{k}{2}\right)} \left[ \frac{k + \left(\frac{u-\beta}{\alpha}\right)^2}{k} \right]^{-\left(\frac{k+1}{2}\right)} du \quad (3.12)$$

### **Log-logistic distribution**

The log-logistic distribution has been used to model headway data in moderate traffic flows and during peak hours [25]. The PDF of the log-logistic distribution for non-negative  $x$  is [27]

$$f(x) = \left(\frac{k}{\alpha}\right) \left(\frac{x}{\alpha}\right)^{k-1} \left[1 + \left(\frac{x}{\alpha}\right)^k\right]^{-2} \quad x > 0 \quad (3.13)$$

where  $k$  and  $\alpha$  are the shape and scale parameters, respectively. Moreover, the CDF of the log-logistic distribution is

$$F(x) = \frac{x^k}{\alpha^k + x^k} \quad (3.14)$$

### **Weibull distribution**

The PDF of the Weibull distribution is [26]

$$f(x) = \frac{kx^{k-1}}{\alpha^k} \times \exp(-(x/\alpha)^k) \quad (3.15)$$

where  $k$  and  $\alpha$  are the shape and scale parameters, respectively. The CDF of the Weibull distribution for non-negative  $x$  is

$$F(x) = 1 - \exp(-(x/\alpha)^k) \quad (3.16)$$

### **Exponential distribution**

The PDF of the exponential distribution is [26]

$$f(x) = (1/\alpha) - \exp(-x/\alpha) \quad (3.17)$$

where  $\alpha$  is the scale parameter. The CDF of the exponential distribution is

$$F(x) = 1 - \exp(-x/\alpha) \quad (3.18)$$

### Extreme value distribution

The PDF of the extreme value distribution is [26]

$$f(x) = (1/\alpha) \times \exp[-(x - \beta)/\alpha] \times \exp(-\exp[-(x - \beta)/\alpha]) \quad (3.19)$$

where  $\beta$  and  $\alpha$  are the location and scale parameters, respectively. The CDF of the extreme value distribution is

$$F(x) = \exp(-\exp[-(x - \beta)/\alpha]) \quad (3.20)$$

### Gamma distribution

The PDF of the Gamma distribution is [26]

$$f(x) = (x/\alpha)^{k-1} [\exp(-x/\alpha)]/\alpha\Gamma(k) \quad (3.21)$$

where  $\Gamma(\cdot)$  is the Gamma function,  $k$  is the shape parameter and  $\alpha$  is the scale parameter.

The CDF of the Gamma distribution is

$$F(x) = \frac{1}{\alpha^k \Gamma(k)} \int_{-\infty}^x t^{k-1} \exp((-t/\alpha)) dt \quad (3.22)$$

### Normal distribution

The normal distribution is applicable to a wide range of data. The PDF of the normal distribution is [26]

$$f(x) = \frac{\exp(-(x - \beta)^2/2\alpha^2)}{\alpha\sqrt{2\pi}} \quad (3.23)$$

where  $\beta$  and  $\alpha$  are the location and scale parameters, respectively. The CDF of the normal distribution is

$$F(x) = \frac{1}{\alpha\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-\beta)^2}{2\alpha^2}} dt \quad (3.24)$$

### Maximum likelihood estimation

In statistics, Maximum Likelihood Estimation (MLE) is used to estimate the parameters of a statistical distribution. MLE finds the parameter values that maximize the likelihood function given the data (observations). The likelihood function is a function of the parameters of a statistical distribution. This method uses the observed data and obtains values of the distribution parameters,  $\theta$ , that maximize the likelihood function

$$L(\theta|X) = \prod_{i=1}^n f(x_i, \theta) \quad (3.25)$$

where  $x_i$  is the  $i$ th observed data value,  $n$  is number of values, and  $f(x_i, \theta)$  is the PDF of the statistical distribution with parameters  $\theta = [\theta_1, \theta_2, \dots, \theta_k]$  [28].

To find  $\theta_i$ , take the first derivative of the likelihood function with respect to  $\theta_i$  [29] and set it to zero which gives

$$\frac{d}{d\theta_i} \prod_{i=1}^n f(x_i, \theta_i) = 0 \quad (3.26)$$

### 3.1. Distribution Parameter Estimation

The procedure of fitting a distribution to the data begins with a histogram which provides a representation of the shape of the underlying probability density function. The shape of the histogram helps to identify distribution candidates. In this thesis, the PDF of each distribution was fit to each headway dataset separately using the MLE technique. For example, the PDF of Burr distribution for the N1 dataset has parameters  $k = 0.25$ ,  $c = 9.07$  and  $\alpha = 0.99$ .

To find the best headway distribution, the eleven PDFs described previously distributions were used to model each headway dataset. The parameters of these distributions were calculated using the MLE technique. The parameters for each distribution for the eight datasets are given in Table 3.1. This shows that the scale parameter increases with decreasing flow for most datasets. The PDFs of these distributions and the histogram of the headway data for each dataset are presented in Figures 3.4 to 3.11. These figures show that the histograms of the headway datasets are right skewed. Figure 3.4 shows that the N1 headway dataset has the largest percentage of very small headway values (less than 1.1 s) and this is because of congestion. The histogram of the R4 dataset for rainy weather with a traffic flow rate less than 1000 vphpl in Figure 3.11 shows that there is a large variation in headway. This indicates that vehicle behaviour varies more in rainy weather. In the next section, the goodness of fit tests are conducted.

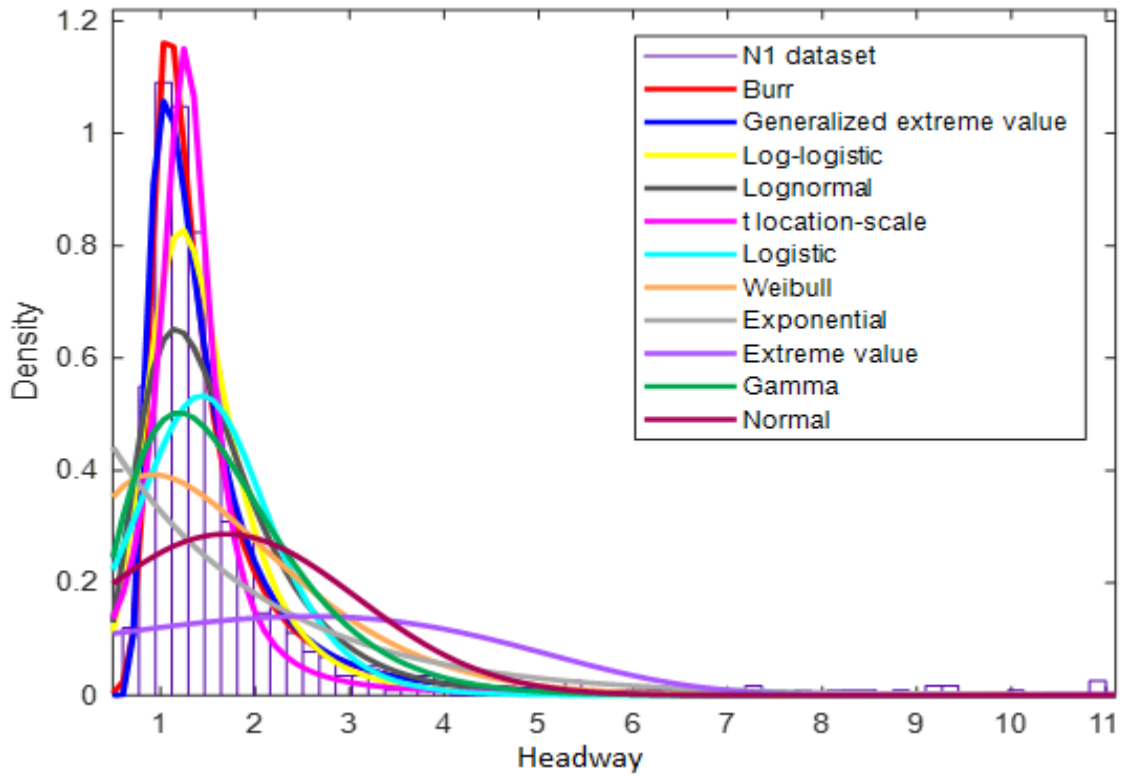


Figure 3.4: The PDFs of all distributions fit to the N1 dataset.

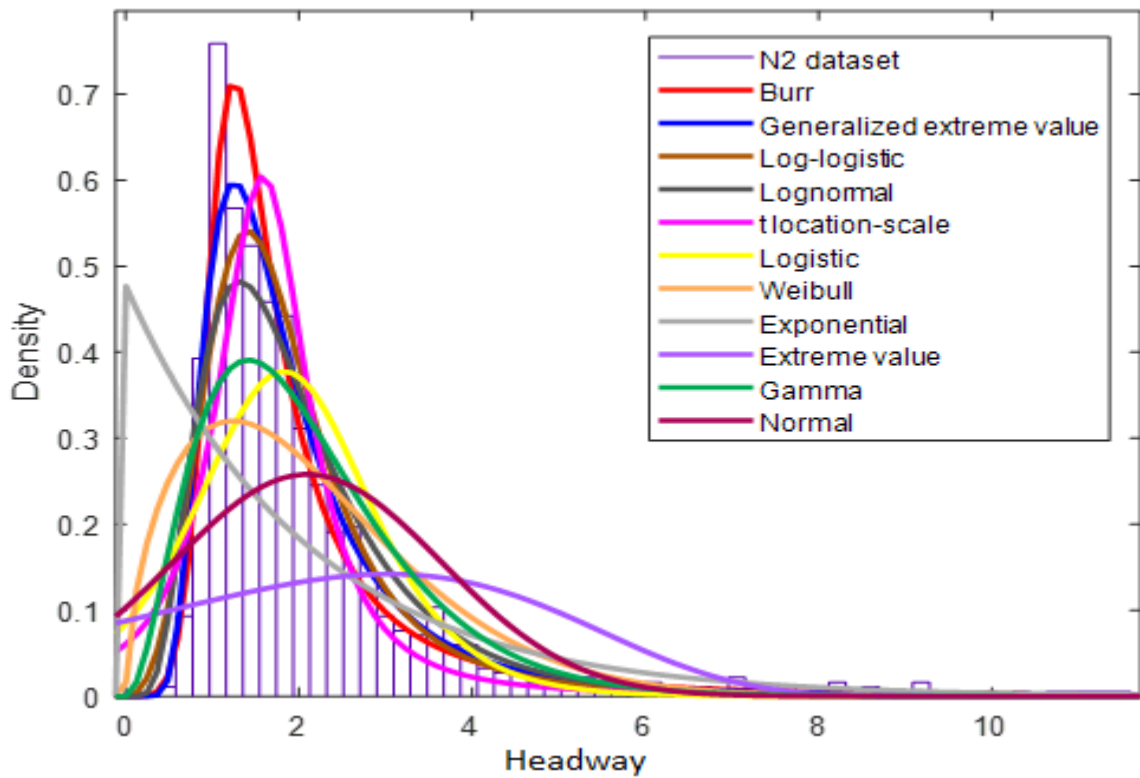


Figure 3.5: The PDFs of all distributions fit to the N2 dataset.

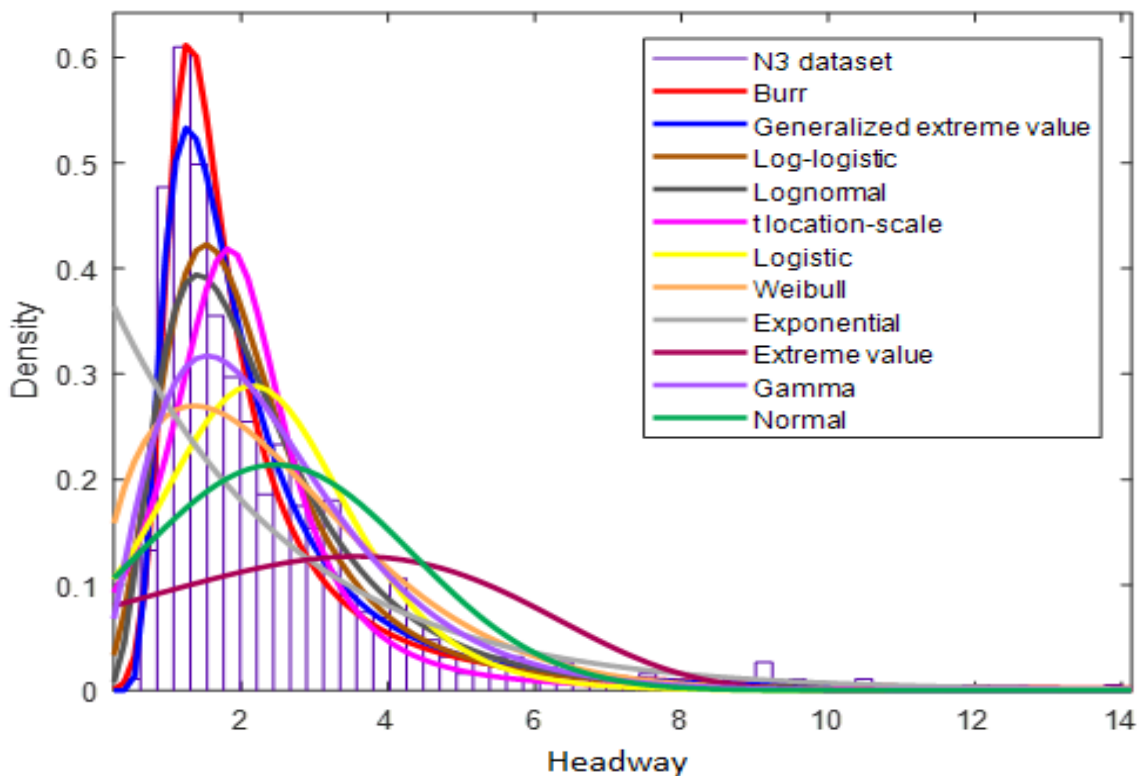


Figure 3.6: The PDFs of all distributions fit to the N3 dataset.

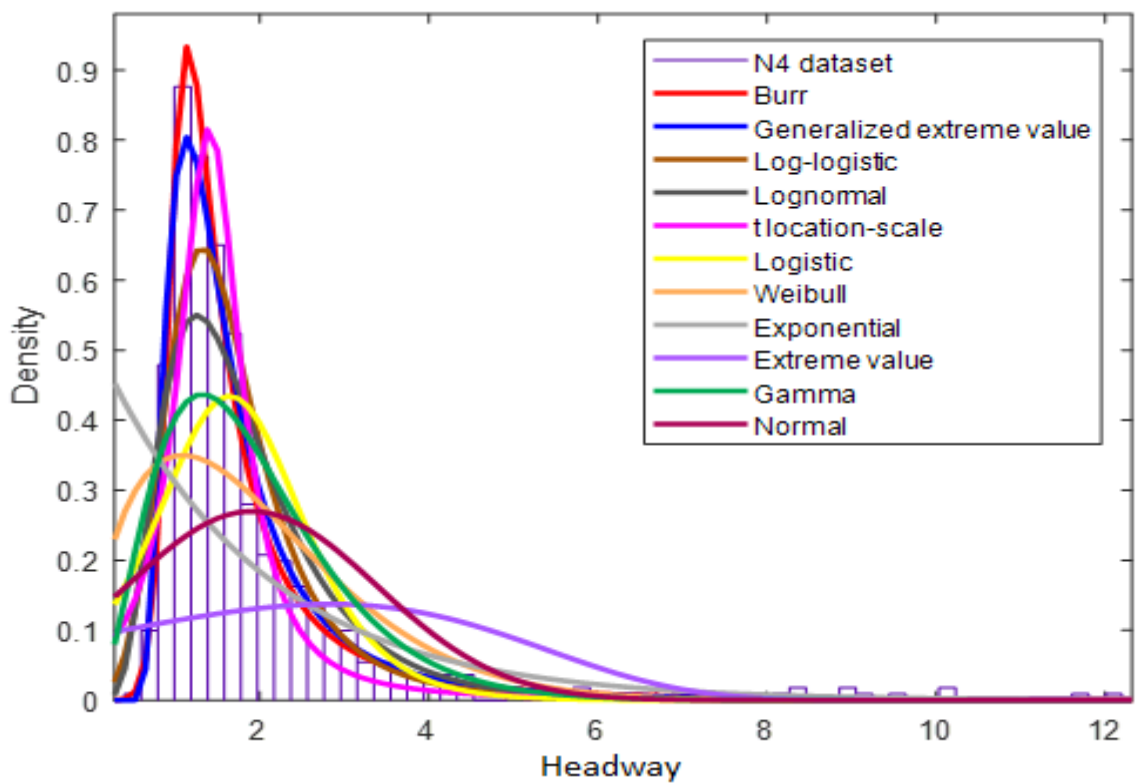


Figure 3.7: The PDFs of all distributions fit to the N4 dataset.

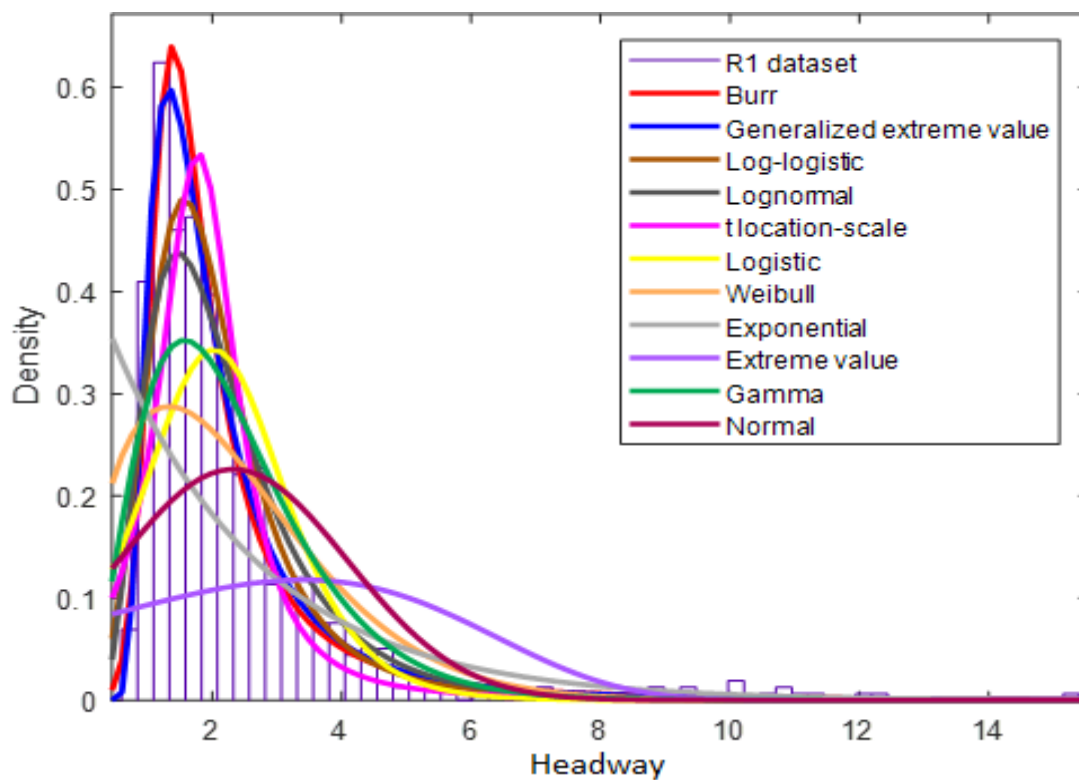


Figure 3.8: The PDFs of all distributions fit to the R1 dataset.

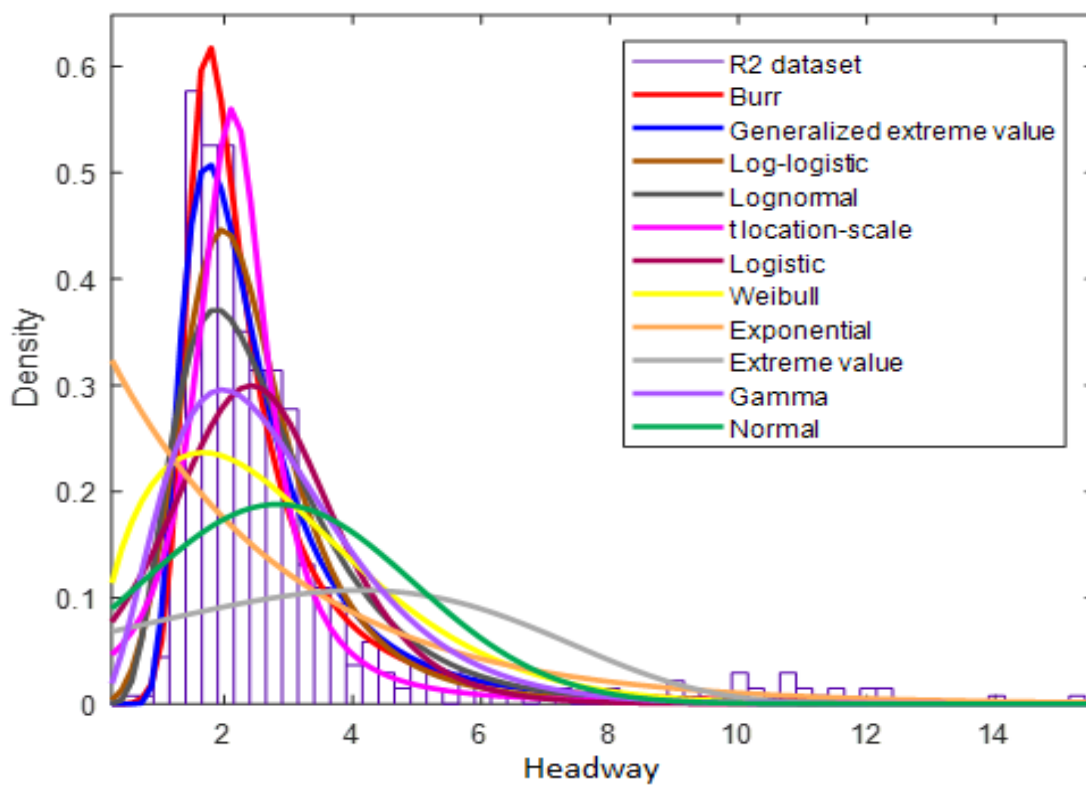


Figure 3.9: The PDFs of all distributions fit to the R2 dataset.

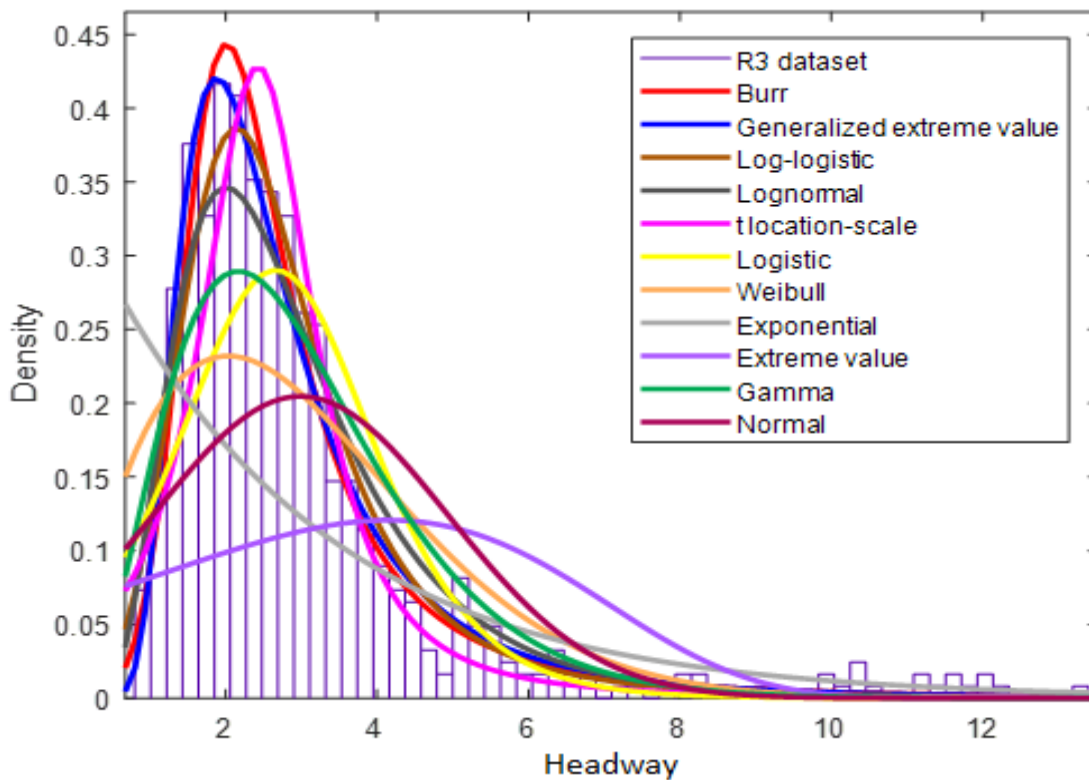


Figure 3.10: The PDFs of all distributions fit to the R3 dataset.

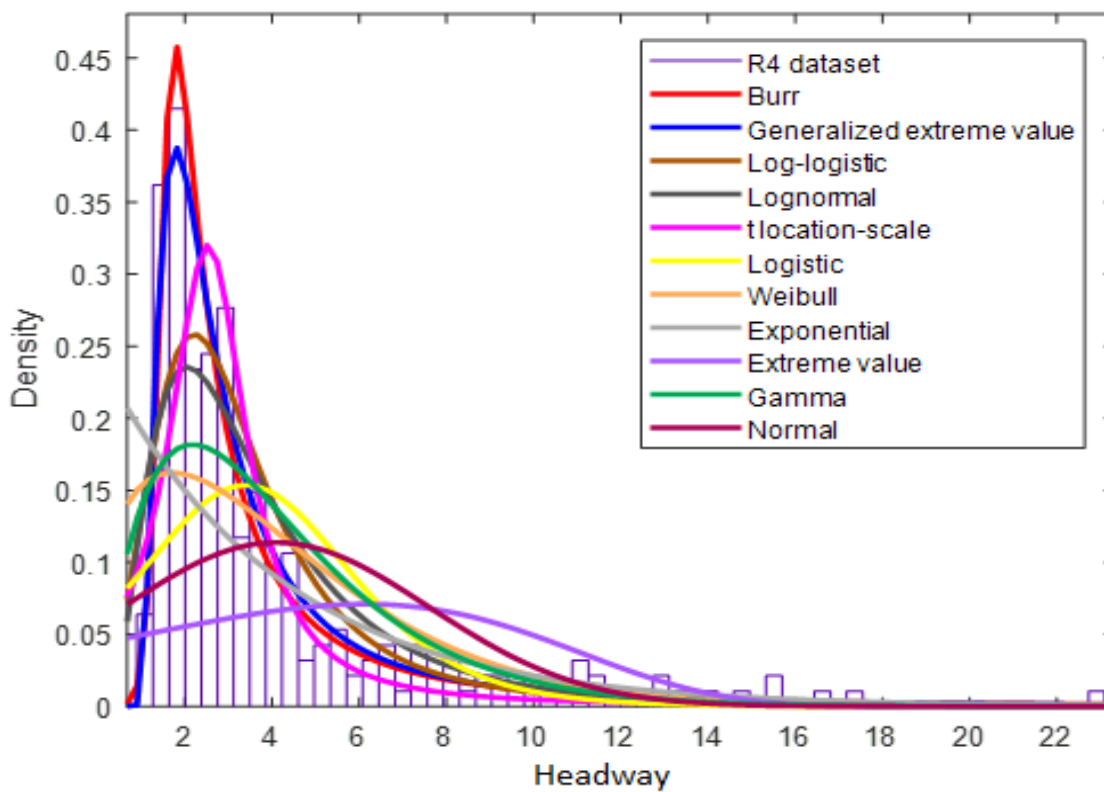


Figure 3.11: The PDFs of all distributions fit to the R4 dataset.

Table 3.1: Statistical distribution parameters for the headway data

| Probability distribution         | Dataset | Shape parameters      | Scale parameter | Location parameter |
|----------------------------------|---------|-----------------------|-----------------|--------------------|
| <b>Burr</b>                      | N1      | $k = 0.25$ $c = 9.07$ | $\alpha = 0.99$ | -                  |
|                                  | N2      | $k = 0.33$ $c = 5.98$ | $\alpha = 1.15$ | -                  |
|                                  | N3      | $k = 0.37$ $c = 5.08$ | $\alpha = 1.28$ | -                  |
|                                  | N4      | $k = 0.24$ $c = 8.45$ | $\alpha = 1.03$ | -                  |
|                                  | R1      | $k = 0.33$ $c = 6.01$ | $\alpha = 1.28$ | -                  |
|                                  | R2      | $k = 0.42$ $c = 5.88$ | $\alpha = 1.75$ | -                  |
|                                  | R3      | $k = 0.52$ $c = 4.83$ | $\alpha = 1.97$ | -                  |
|                                  | R4      | $k = 0.16$ $c = 8.50$ | $\alpha = 1.56$ | -                  |
| <b>Generalized extreme value</b> | N1      | $k = 0.36$            | $\alpha = 0.43$ | $\beta = 1.19$     |
|                                  | N2      | $k = 0.31$            | $\alpha = 0.64$ | $\beta = 1.42$     |
|                                  | N3      | $k = 0.42$            | $\alpha = 0.75$ | $\beta = 1.55$     |
|                                  | N4      | $k = 0.41$            | $\alpha = 0.49$ | $\beta = 1.29$     |
|                                  | R1      | $k = 0.42$            | $\alpha = 0.67$ | $\beta = 1.55$     |
|                                  | R2      | $k = 0.30$            | $\alpha = 0.77$ | $\beta = 1.95$     |
|                                  | R3      | $k = 0.29$            | $\alpha = 0.91$ | $\beta = 2.10$     |
|                                  | R4      | $k = 0.65$            | $\alpha = 1.13$ | $\beta = 2.27$     |
| <b>Log-logistic</b>              | N1      | $k = 0.33$            | $\alpha = 0.25$ | -                  |
|                                  | N2      | $k = 0.53$            | $\alpha = 0.30$ | -                  |
|                                  | N3      | $k = 0.65$            | $\alpha = 0.33$ | -                  |
|                                  | N4      | $k = 0.43$            | $\alpha = 0.27$ | -                  |
|                                  | R1      | $k = 0.64$            | $\alpha = 0.30$ | -                  |
|                                  | R2      | $k = 0.83$            | $\alpha = 0.26$ | -                  |
|                                  | R3      | $k = 0.92$            | $\alpha = 0.28$ | -                  |
|                                  | R4      | $k = 1.09$            | $\alpha = 0.38$ | -                  |
| <b>Lognormal</b>                 | N1      | $k = 0.39$            | $\alpha = 0.49$ | -                  |
|                                  | N2      | $k = 0.57$            | $\alpha = 0.54$ | -                  |
|                                  | N3      | $k = 0.69$            | $\alpha = 0.58$ | -                  |
|                                  | N4      | $k = 0.48$            | $\alpha = 0.51$ | -                  |
|                                  | R1      | $k = 0.68$            | $\alpha = 0.54$ | -                  |
|                                  | R2      | $k = 0.87$            | $\alpha = 0.48$ | -                  |
|                                  | R3      | $k = 0.95$            | $\alpha = 0.51$ | -                  |
|                                  | R4      | $k = 1.16$            | $\alpha = 0.66$ | -                  |
| <b>t location-scale</b>          | N1      | $k = 1.33$            | $\alpha = 0.32$ | $\beta = 1.27$     |
|                                  | N2      | $k = 1.58$            | $\alpha = 0.57$ | $\beta = 1.70$     |
|                                  | N3      | $k = 1.79$            | $\alpha = 0.74$ | $\beta = 1.74$     |
|                                  | N4      | $k = 1.41$            | $\alpha = 0.41$ | $\beta = 1.44$     |
|                                  | R1      | $k = 1.78$            | $\alpha = 0.64$ | $\beta = 1.72$     |
|                                  | R2      | $k = 2.21$            | $\alpha = 0.69$ | $\beta = 1.86$     |
|                                  | R3      | $k = 2.43$            | $\alpha = 0.82$ | $\beta = 1.90$     |
|                                  | R4      | $k = 2.53$            | $\alpha = 1.03$ | $\beta = 1.21$     |

|                      |    |            |                 |                |
|----------------------|----|------------|-----------------|----------------|
| <b>Logistic</b>      | N1 | -          | $\alpha = 0.47$ | $\beta = 1.44$ |
|                      | N2 | -          | $\alpha = 0.66$ | $\beta = 1.83$ |
|                      | N3 | -          | $\alpha = 0.87$ | $\beta = 2.15$ |
|                      | N4 | -          | $\alpha = 0.58$ | $\beta = 1.64$ |
|                      | R1 | -          | $\alpha = 0.73$ | $\beta = 2.02$ |
|                      | R2 | -          | $\alpha = 0.83$ | $\beta = 2.42$ |
|                      | R3 | -          | $\alpha = 0.86$ | $\beta = 2.66$ |
|                      | R4 | -          | $\alpha = 1.63$ | $\beta = 3.40$ |
| <b>Weibull</b>       | N1 | $k = 1.51$ | $\alpha = 1.90$ | -              |
|                      | N2 | $k = 1.57$ | $\alpha = 2.36$ | -              |
|                      | N3 | $k = 1.52$ | $\alpha = 2.77$ | -              |
|                      | N4 | $k = 1.54$ | $\alpha = 2.15$ | -              |
|                      | R1 | $k = 1.55$ | $\alpha = 2.61$ | -              |
|                      | R2 | $k = 1.57$ | $\alpha = 3.19$ | -              |
|                      | R3 | $k = 1.72$ | $\alpha = 3.38$ | -              |
|                      | R4 | $k = 1.36$ | $\alpha = 4.53$ | -              |
| <b>Exponential</b>   | N1 | -          | $\alpha = 1.69$ | -              |
|                      | N2 | -          | $\alpha = 2.10$ | -              |
|                      | N3 | -          | $\alpha = 2.47$ | -              |
|                      | N4 | -          | $\alpha = 1.91$ | -              |
|                      | R1 | -          | $\alpha = 2.32$ | -              |
|                      | R2 | -          | $\alpha = 2.83$ | -              |
|                      | R3 | -          | $\alpha = 2.98$ | -              |
|                      | R4 | -          | $\alpha = 4.09$ | -              |
| <b>Extreme value</b> | N1 | -          | $\alpha = 2.61$ | $\beta = 2.59$ |
|                      | N2 | -          | $\alpha = 2.59$ | $\beta = 3.04$ |
|                      | N3 | -          | $\alpha = 2.90$ | $\beta = 3.57$ |
|                      | N4 | -          | $\alpha = 2.68$ | $\beta = 2.83$ |
|                      | R1 | -          | $\alpha = 3.12$ | $\beta = 3.41$ |
|                      | R2 | -          | $\alpha = 3.44$ | $\beta = 4.13$ |
|                      | R3 | -          | $\alpha = 3.05$ | $\beta = 4.15$ |
|                      | R4 | -          | $\alpha = 5.18$ | $\beta = 6.13$ |
| <b>Gamma</b>         | N1 | $k = 3.44$ | $\alpha = 0.49$ | -              |
|                      | N2 | $k = 3.09$ | $\alpha = 0.68$ | -              |
|                      | N3 | $k = 2.65$ | $\alpha = 0.93$ | -              |
|                      | N4 | $k = 3.25$ | $\alpha = 0.59$ | -              |
|                      | R1 | $k = 3.08$ | $\alpha = 0.75$ | -              |
|                      | R2 | $k = 3.30$ | $\alpha = 0.86$ | -              |
|                      | R3 | $k = 3.63$ | $\alpha = 0.82$ | -              |
|                      | R4 | $k = 2.13$ | $\alpha = 1.92$ | -              |
| <b>Normal</b>        | N1 | -          | $\alpha = 1.39$ | $\beta = 1.69$ |
|                      | N2 | -          | $\alpha = 1.55$ | $\beta = 2.10$ |
|                      | N3 | -          | $\alpha = 1.87$ | $\beta = 2.47$ |
|                      | N4 | -          | $\alpha = 1.48$ | $\beta = 1.91$ |

|    |   |                 |                |
|----|---|-----------------|----------------|
| R1 | - | $\alpha = 1.76$ | $\beta = 2.32$ |
| R2 | - | $\alpha = 2.12$ | $\beta = 2.83$ |
| R3 | - | $\alpha = 1.95$ | $\beta = 2.99$ |
| R4 | - | $\alpha = 3.51$ | $\beta = 4.09$ |

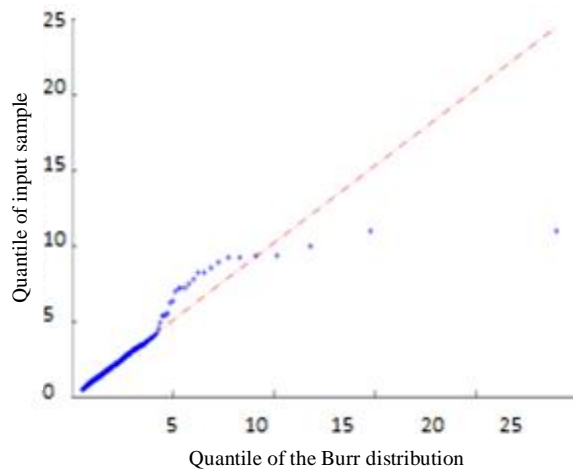
### 3.2. Goodness of Fit

The goodness of fit of a statistical distribution indicates how well it fits a set of observations. Because there are many statistical distributions, three goodness of fit tests are used here to evaluate how well the distributions fit the headway datasets [30].

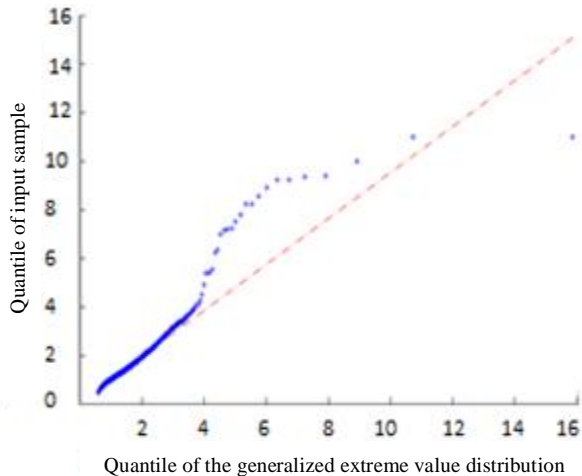
First, a quantile-quantile (Q-Q) plot is obtained to indicate graphically the closeness of the data to the specific distribution. First, the dataset of size  $n$  ( $x_1, x_2, \dots, x_n$ ) is ordered from smallest to largest [31]. Then the  $i$ th ordered value of the dataset is plotted against  $F^{-1}([i - 0.5]/n)$  where  $F$  is the CDF of the distribution. If the points in the Q-Q plot fall along a 45 degree line, then the dataset likely comes from this distribution. The Q-Q plots of all eleven distributions for the eight datasets are presented in Figures 3.12 to 3.19.

Figure 3.12 shows the Q-Q plots of the eleven distributions for the N1 dataset. Comparing these plots shows that the Burr distribution Q-Q plot in Figure 3.12(a) is closest to the 45 degree line than the other ten distributions, so it provides the best fit for this dataset. Figures 3.13 and 3.19 show that the Burr distribution is also closer to the 45 degree line compared to the Q-Q plots of the other ten distributions. Figure 3.14 shows that the Q-Q plots for the Burr, generalized extreme value, log-logistic, and exponential distributions are closer to the 45 degree line than the other seven distributions. The Q-Q plots for the R4 dataset show that the Burr and generalized extreme value distributions are closest to the 45 degree line than the other nine distributions. Comparing the Q-Q plots for the datasets indicates that

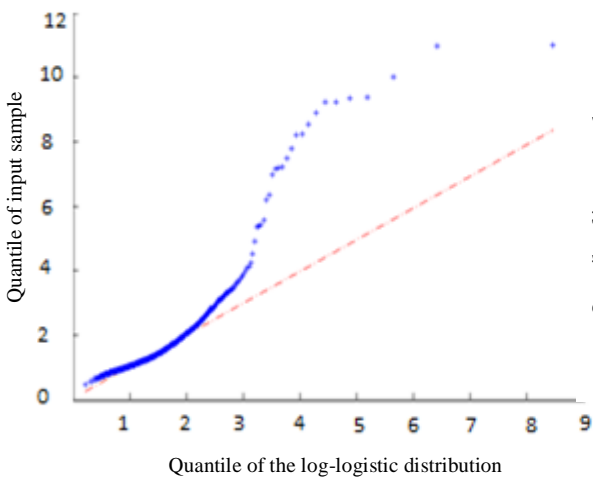
the Burr and generalized extreme value distributions better fit the datasets compared to the other distributions. The nonlinear shape of the Q-Q plots for the  $t$  location-scale, logistic, lognormal, Weibull, extreme value, Gamma, and normal distributions for all datasets suggests that are not suitable for modeling headway data.



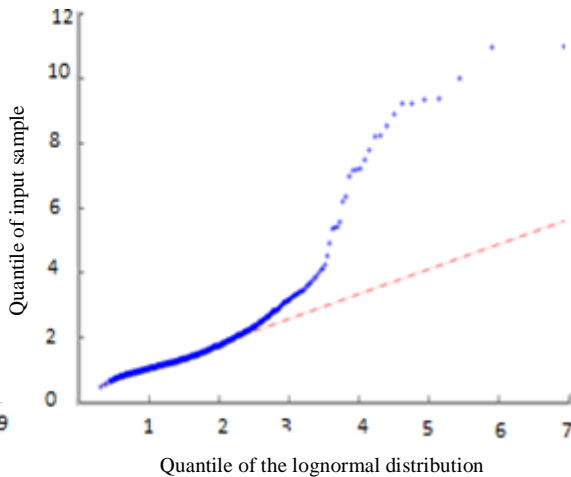
a) The Q-Q plot for the Burr distribution.



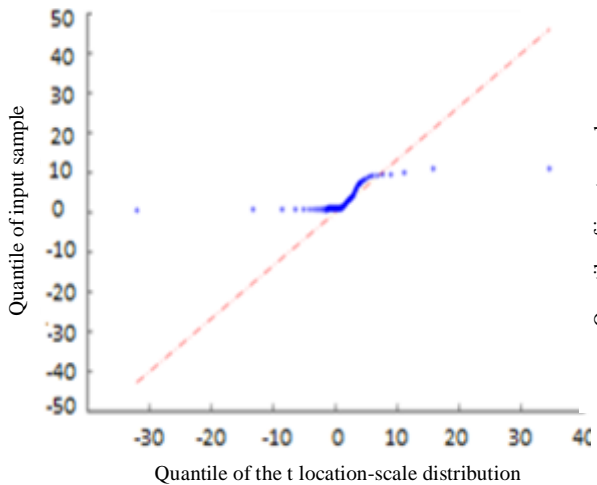
b) The Q-Q plot for the generalized extreme value distribution.



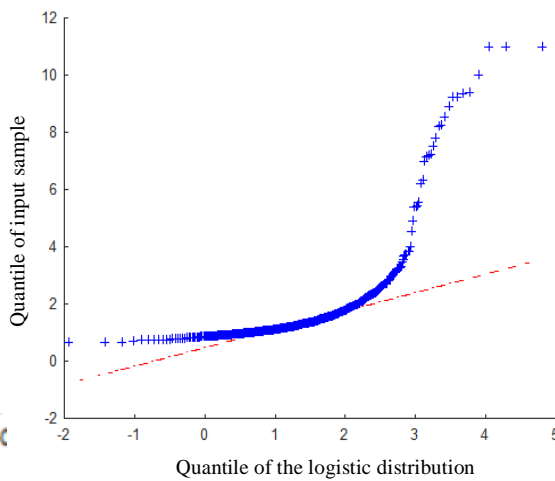
c) The Q-Q plot for the log-logistic distribution.



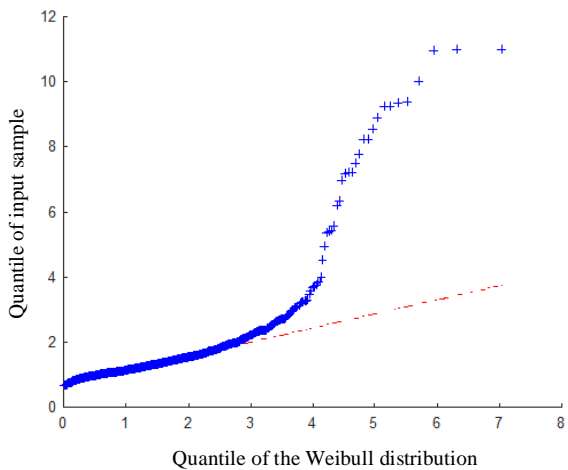
d) The Q-Q plot for the lognormal distribution.



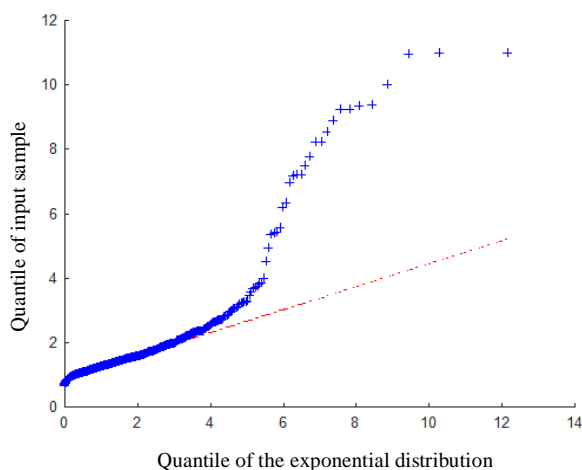
e) The Q-Q plot for the t location-scale distribution.



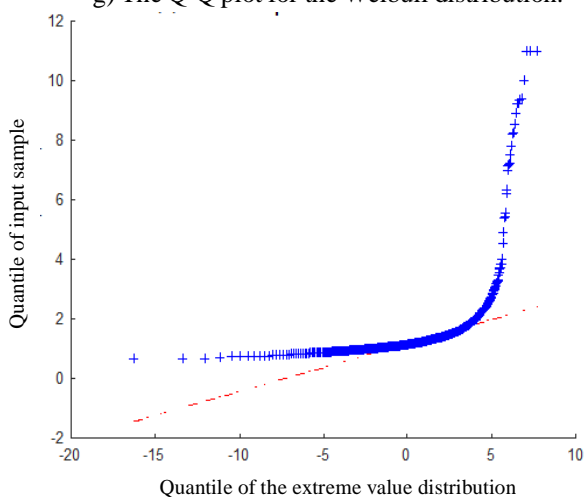
f) The Q-Q plot for the logistic distribution.



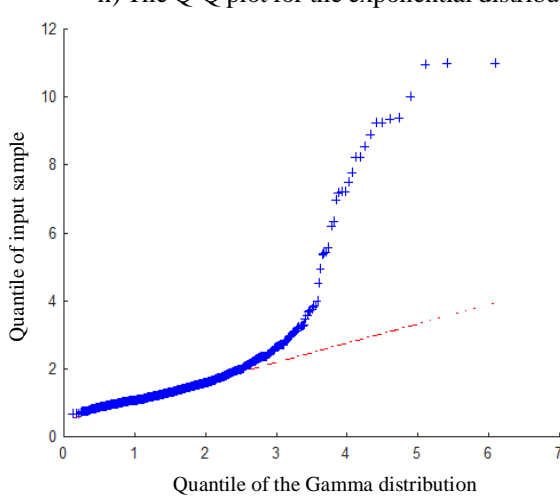
g) The Q-Q plot for the Weibull distribution.



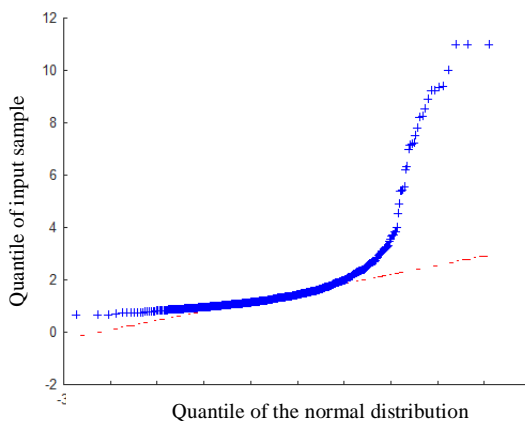
h) The Q-Q plot for the exponential distribution.



i) The Q-Q plot for the extreme value distribution.

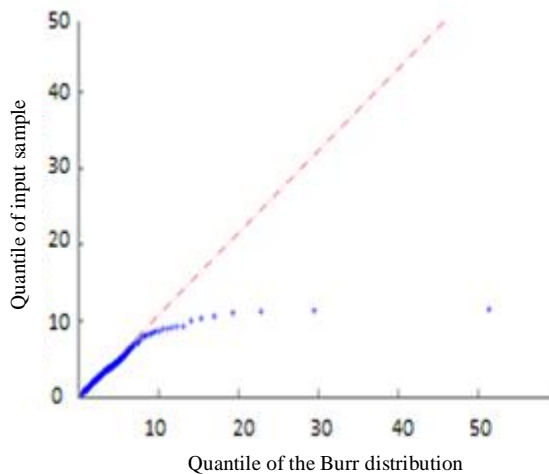


j) The Q-Q plot for the Gamma distribution.

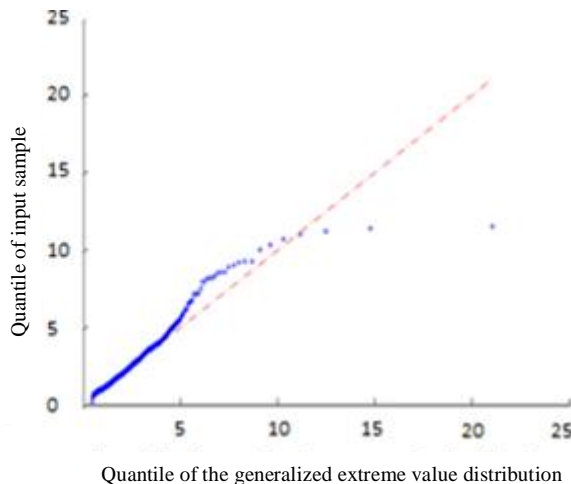


k) The Q-Q plot for the normal distribution.

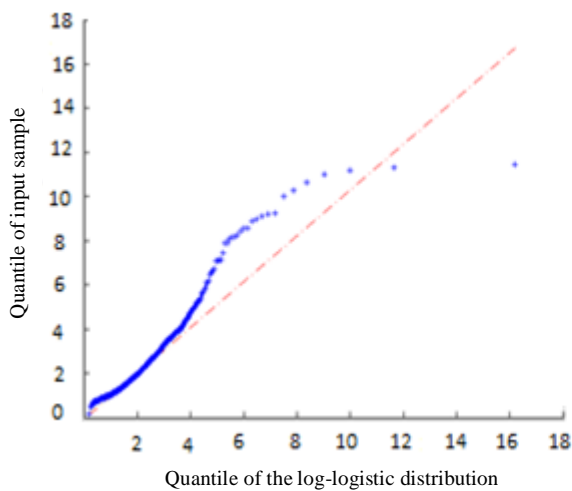
Figure 3.12: The Q-Q plots for the N1 dataset.



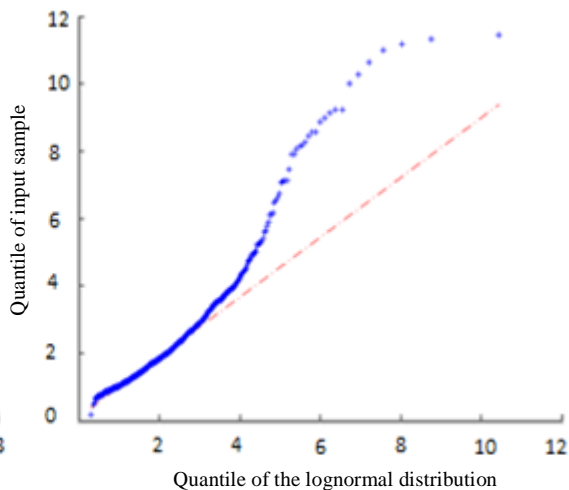
a) The Q-Q plot for the Burr distribution.



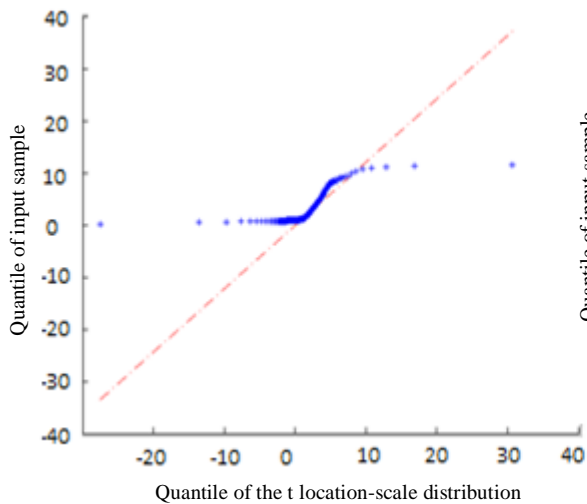
b) The Q-Q plot for the generalized extreme value distribution.



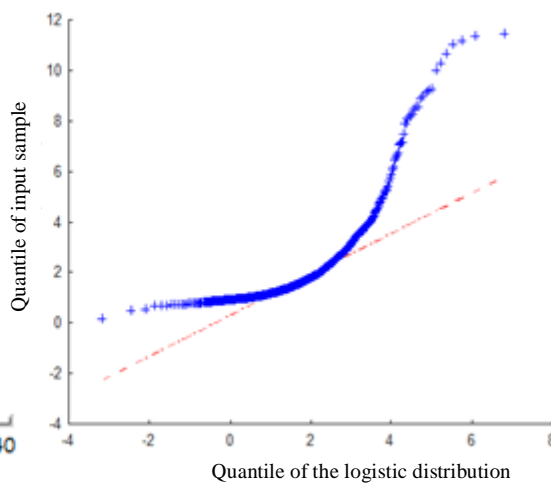
c) The Q-Q plot for the log-logistic distribution.



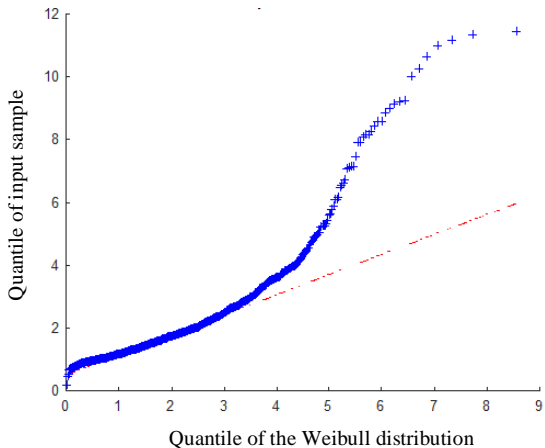
d) The Q-Q plot for the lognormal distribution.



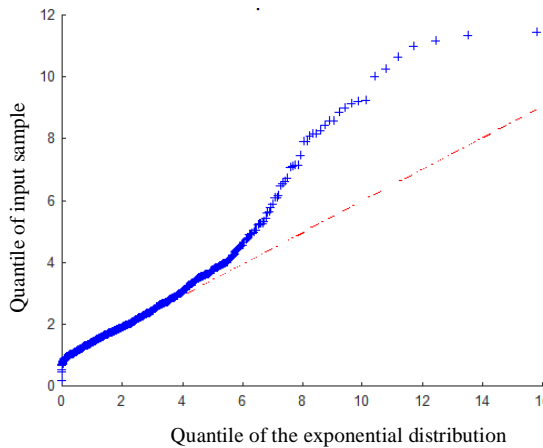
e) The Q-Q plot for the t location-scale distribution.



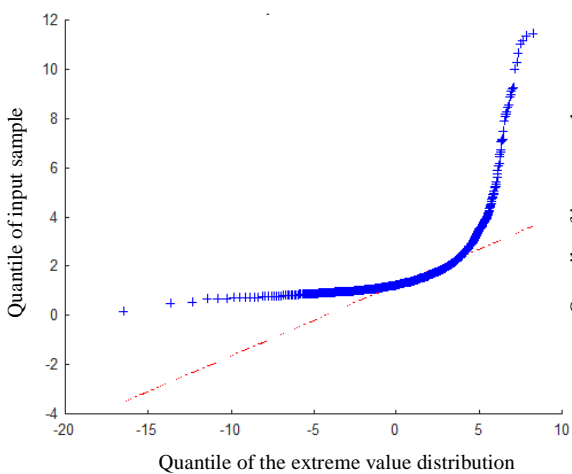
f) The Q-Q plot for the logistic distribution.



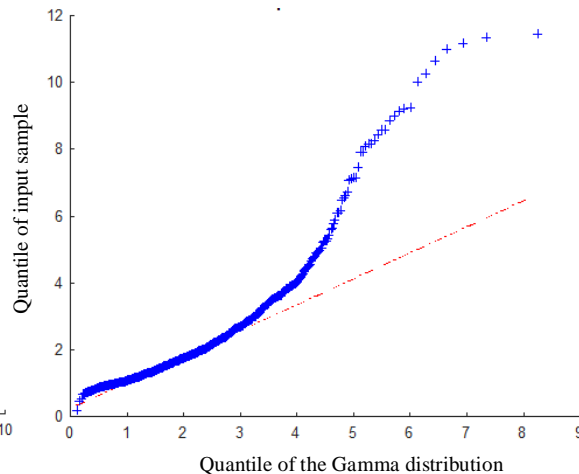
g) The Q-Q plot for the Weibull distribution.



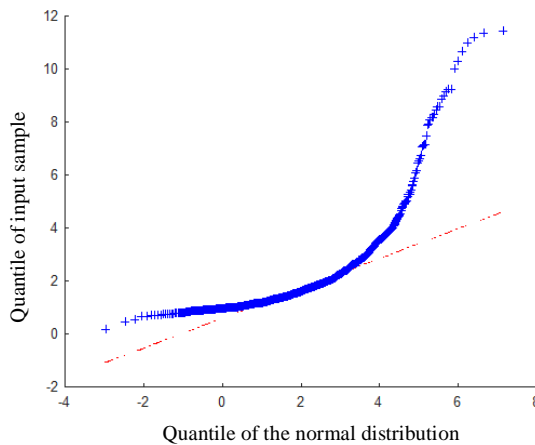
h) The Q-Q plot for the exponential distribution.



i) The Q-Q plot for the extreme value distribution.

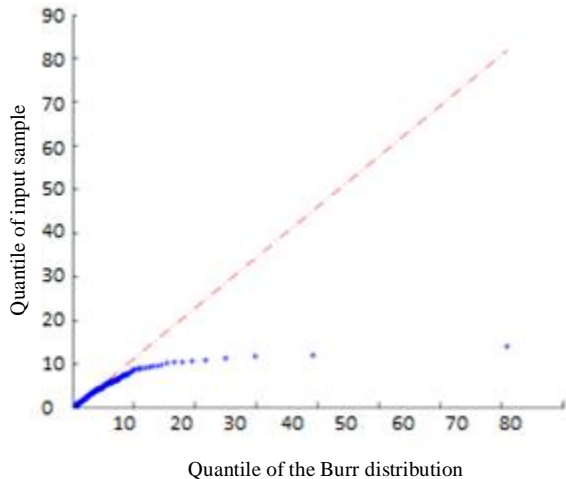


j) The Q-Q plot for the Gamma distribution.

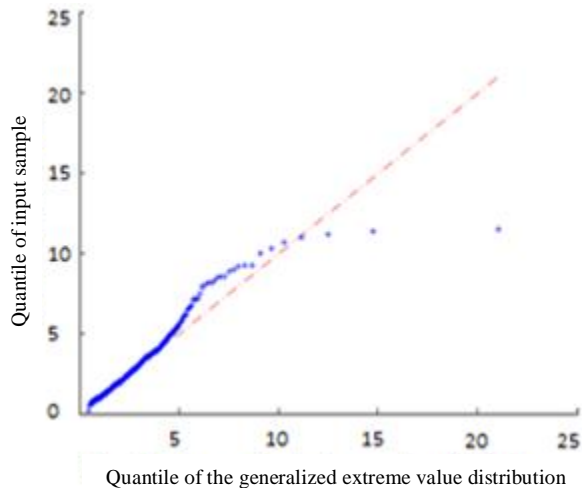


k) The Q-Q plot for the normal distribution.

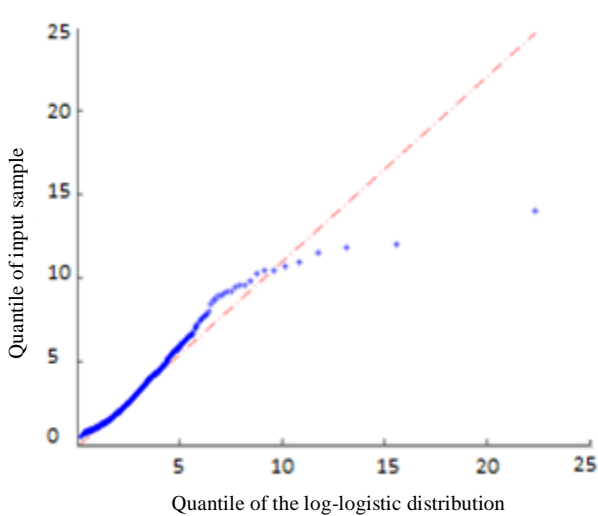
Figure 3.13: The Q-Q plots for the N2 dataset.



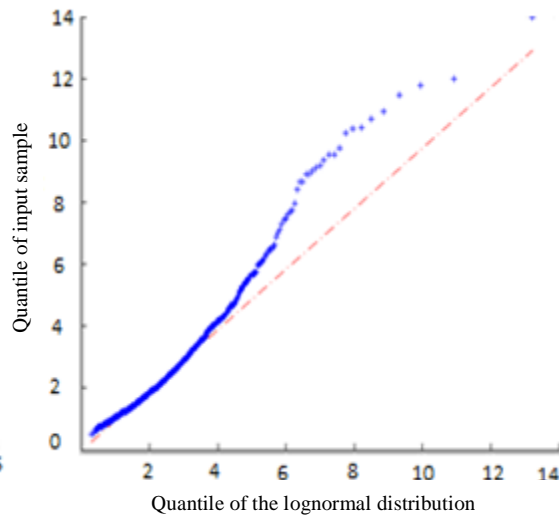
a) The Q-Q plot for the Burr distribution.



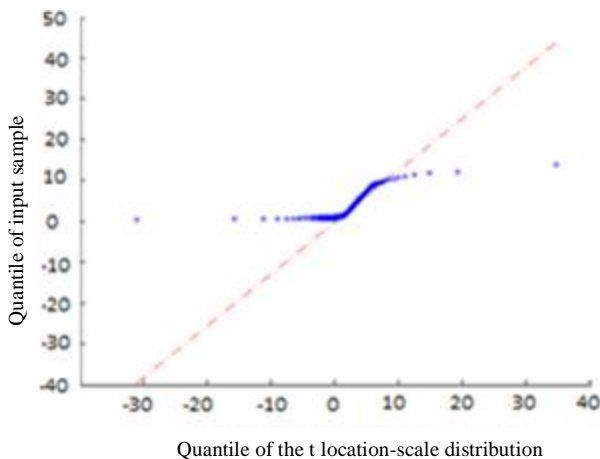
b) The Q-Q plot for the generalized extreme value distribution.



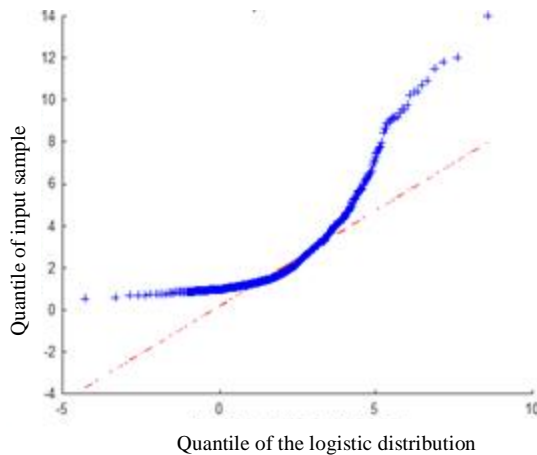
c) The Q-Q plot for the log-logistic distribution.



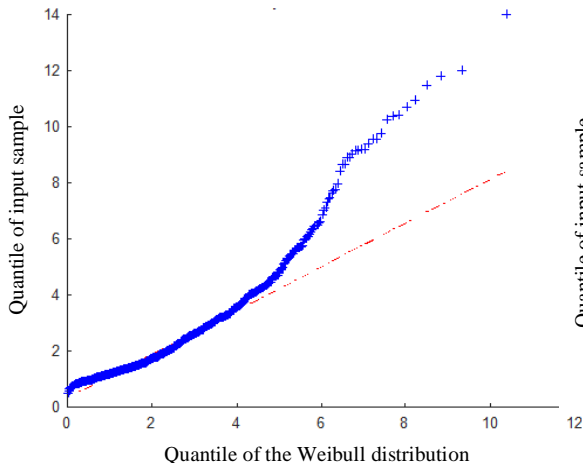
d) The Q-Q plot for the lognormal distribution.



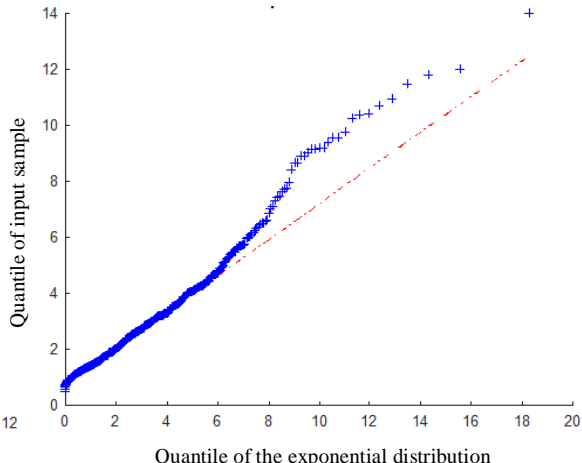
e) The Q-Q plot for the t location-scale distribution.



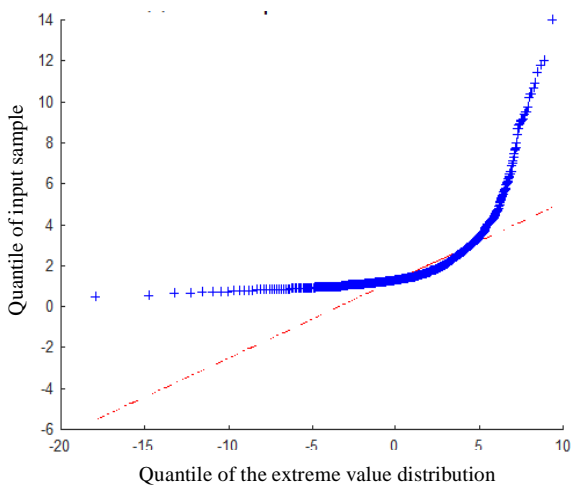
f) The Q-Q plot for the logistic distribution.



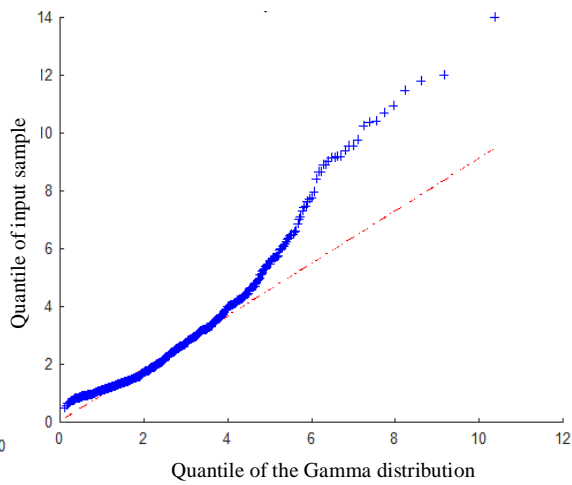
g) The Q-Q plot for the Weibull distribution.



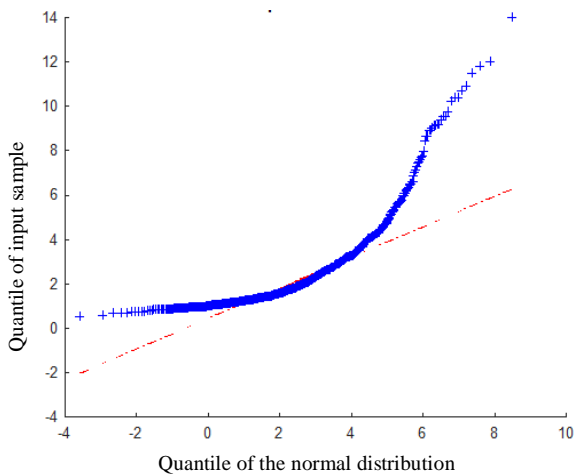
h) The Q-Q plot for the exponential distribution.



i) The Q-Q plot for the extreme value distribution.

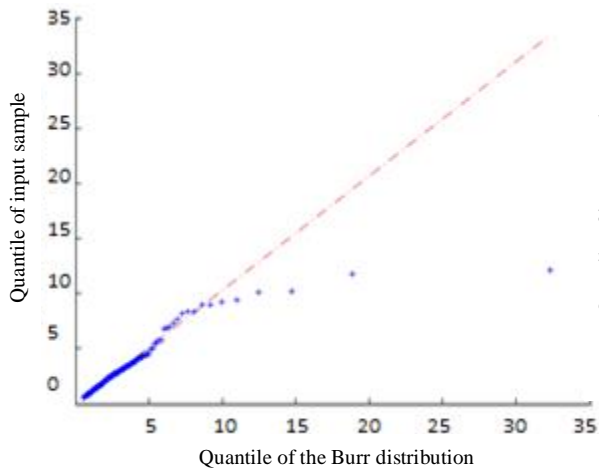


j) The Q-Q plot for the Gamma distribution.

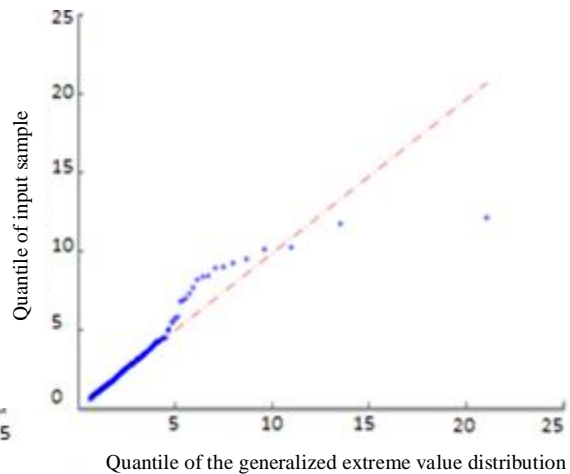


k) The Q-Q plot for the normal distribution.

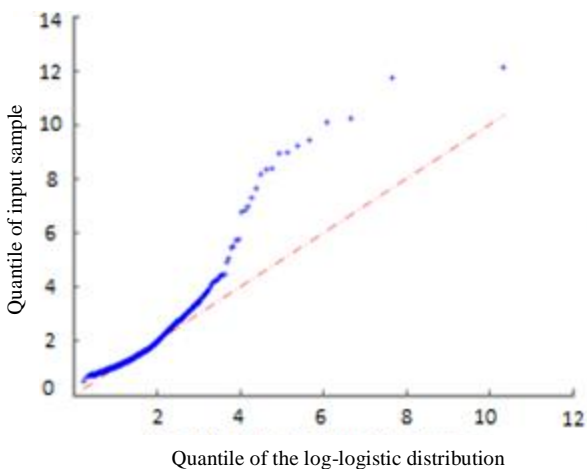
Figure 3.14: The Q-Q plots for the N3 dataset.



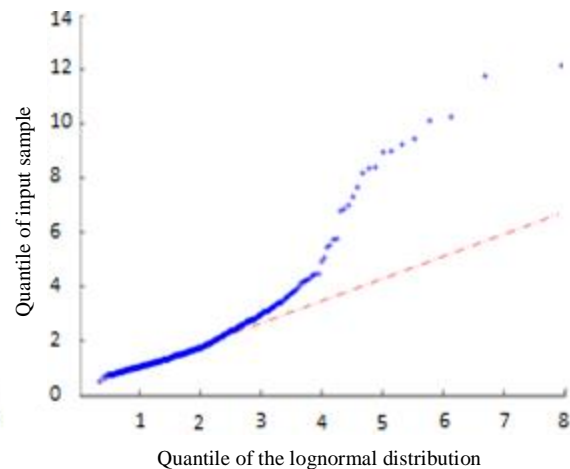
a) The Q-Q plot for the Burr distribution.



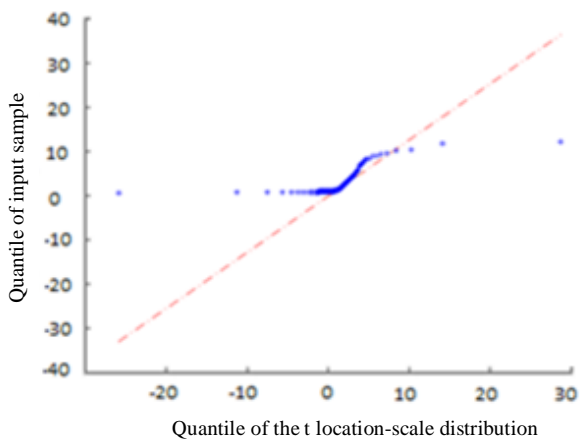
b) The Q-Q plot for the generalized extreme value distribution.



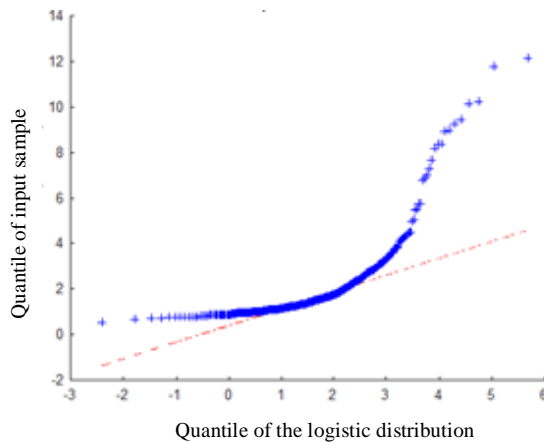
c) The Q-Q plot for the log-logistic distribution.



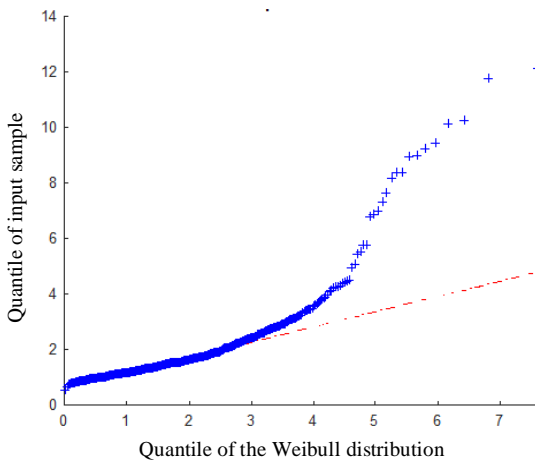
d) The Q-Q plot for the lognormal distribution.



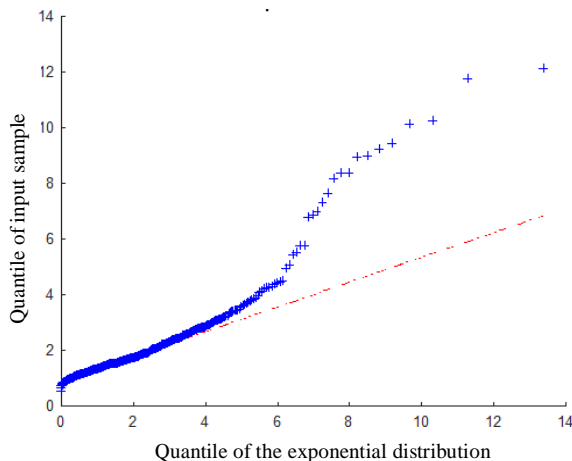
e) The Q-Q plot for the t location-scale distribution.



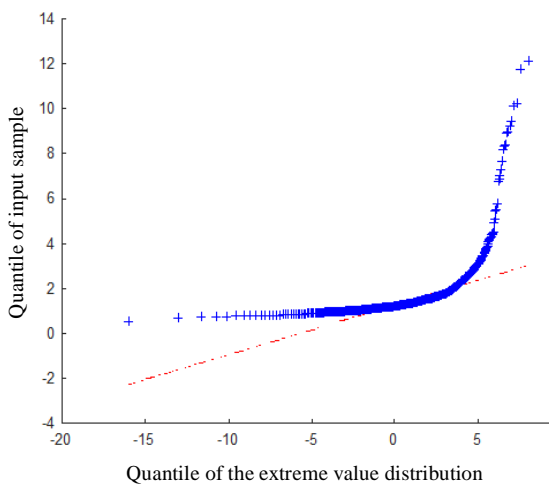
f) The Q-Q plot for the logistic distribution.



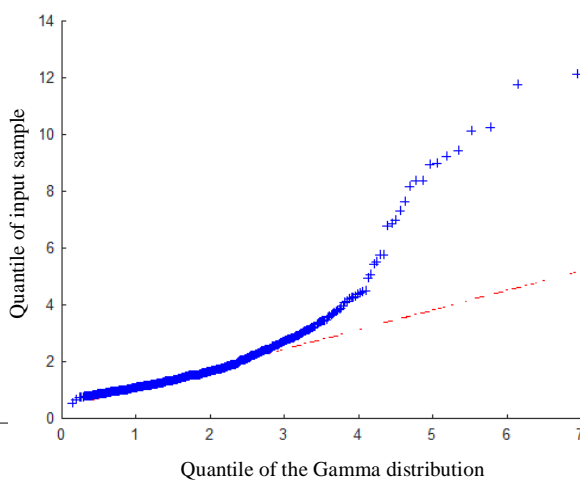
g) The Q-Q plot for the Weibull distribution.



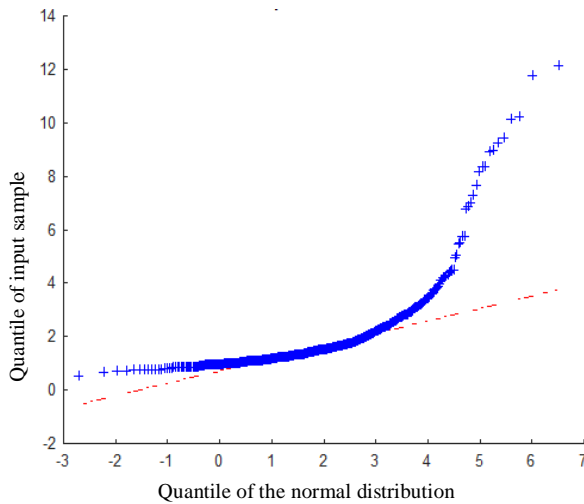
h) The Q-Q plot for the exponential distribution.



i) The Q-Q plot for the extreme value distribution.

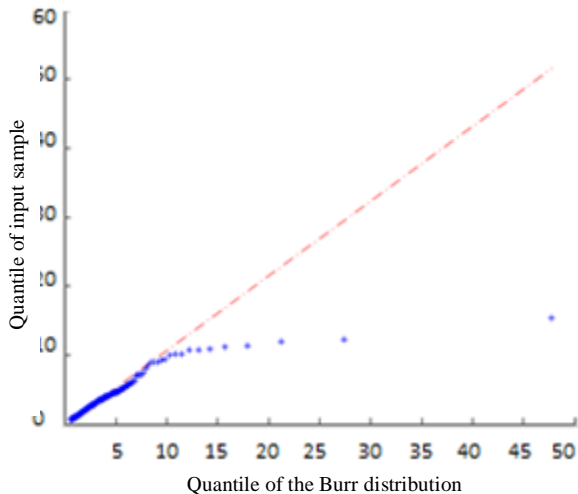


j) The Q-Q plot for the Gamma distribution.

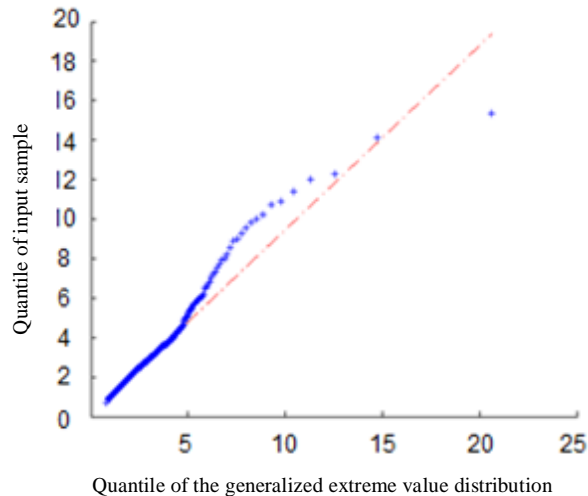


k) The Q-Q plot for the normal distribution.

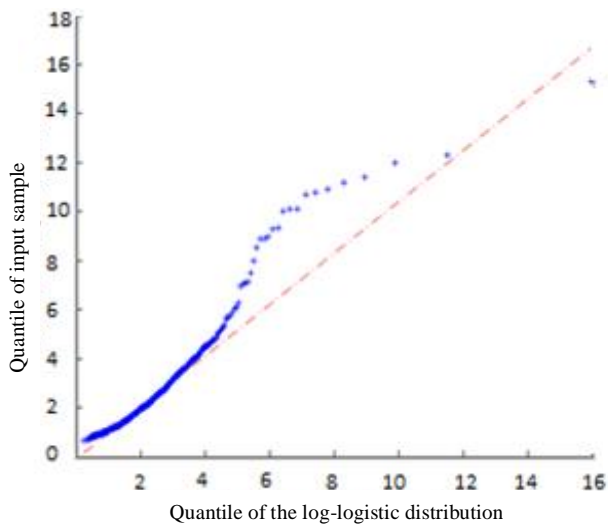
Figure 3.15: The Q-Q plots for the N4 dataset.



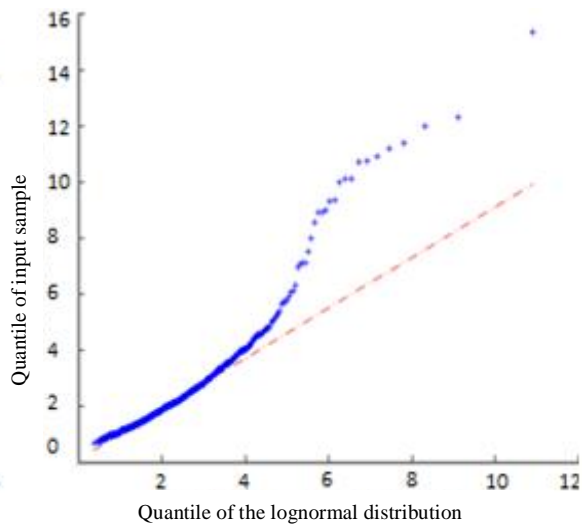
a) The Q-Q plot for the Burr distribution.



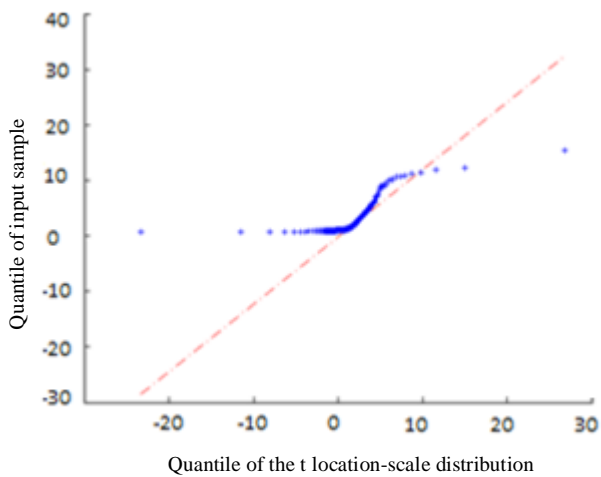
b) The Q-Q plot for the generalized extreme value distribution.



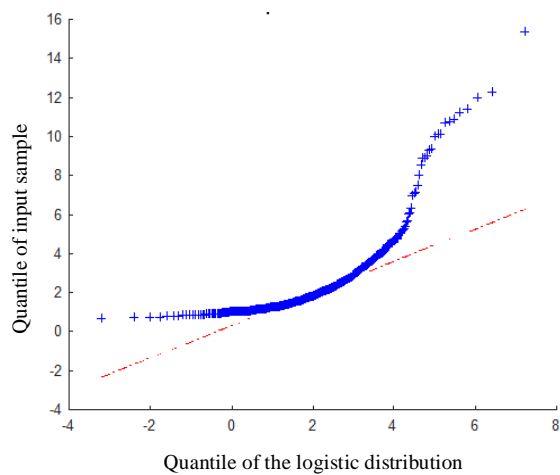
c) The Q-Q plot for the log-logistic distribution.



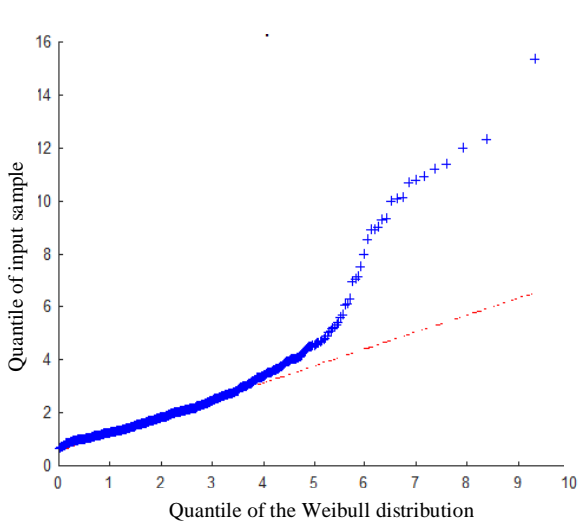
d) The Q-Q plot for the lognormal distribution.



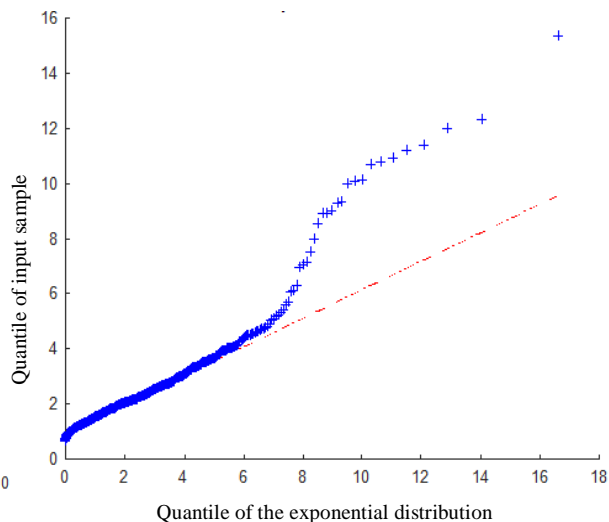
e) The Q-Q plot for the t location-scale distribution.



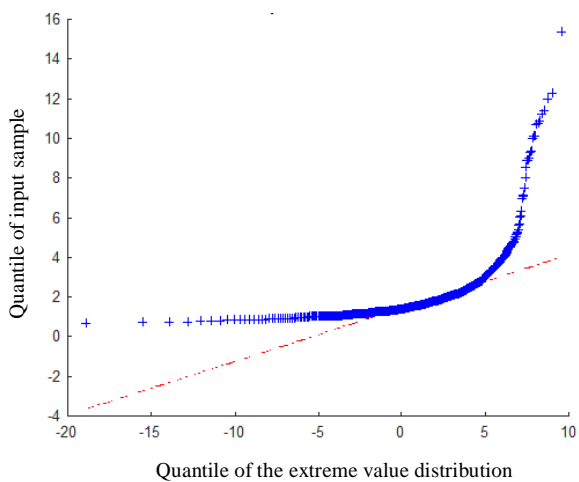
f) The Q-Q plot for the logistic distribution.



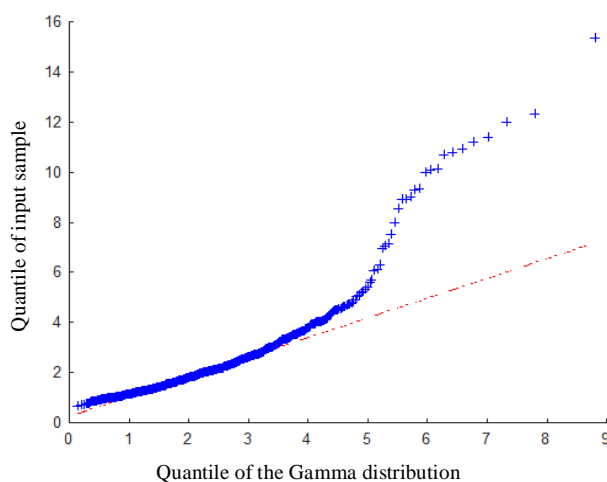
g) The Q-Q plot for the Weibull distribution.



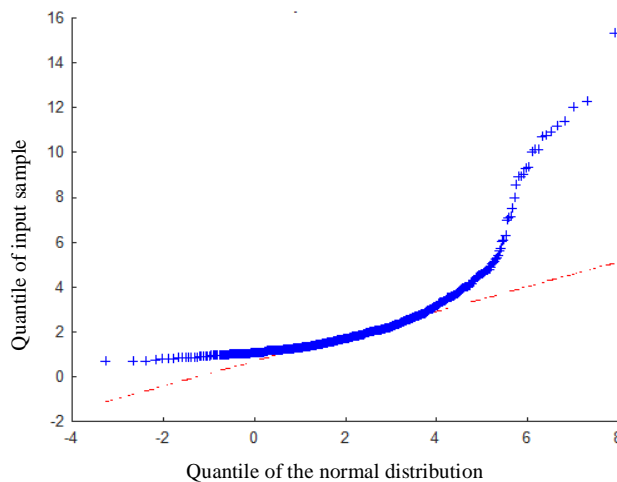
h) The Q-Q plot for the exponential distribution.



i) The Q-Q plot for the extreme value distribution.

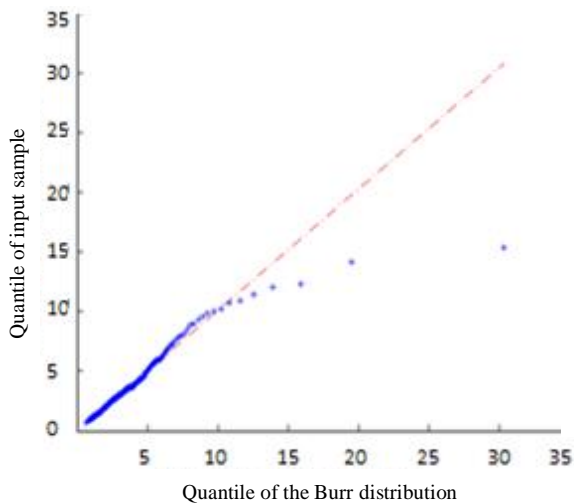


j) The Q-Q plot for the Gamma distribution.

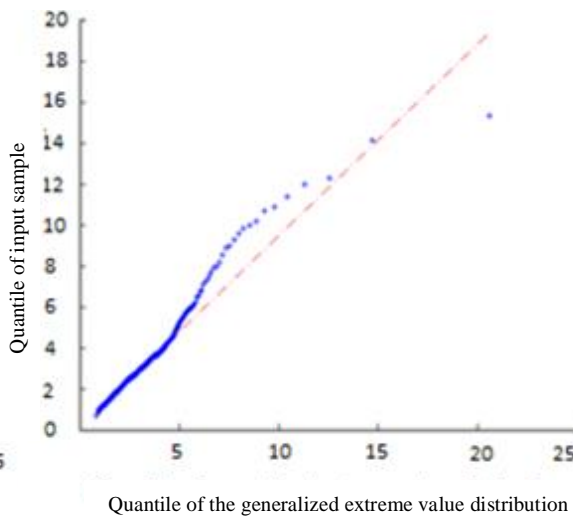


k) The Q-Q plot for the normal distribution.

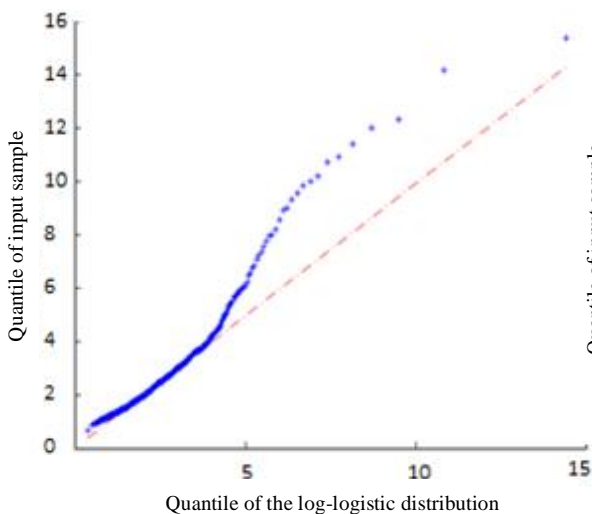
Figure 3.16: The Q-Q plots for the R1 dataset.



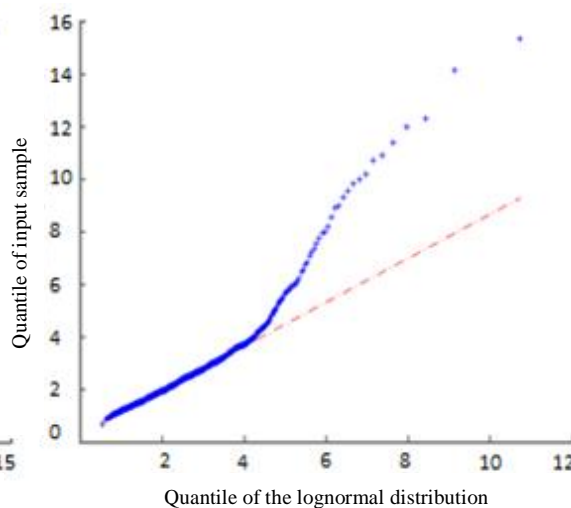
a) The Q-Q plot for the Burr distribution.



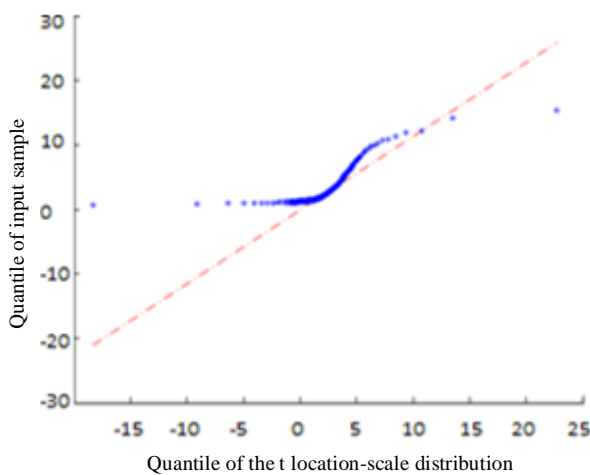
b) The Q-Q plot for the generalized extreme value distribution.



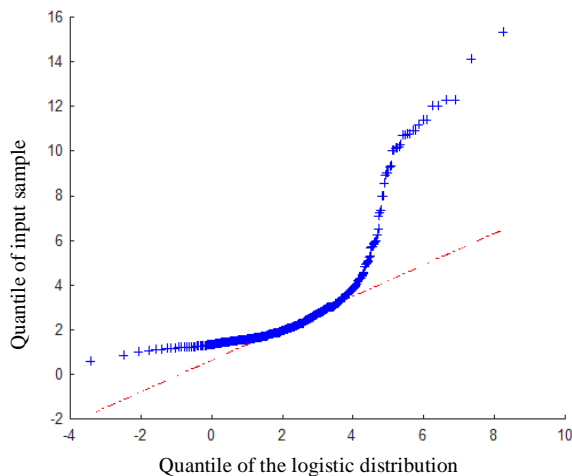
c) The Q-Q plot for the log-logistic distribution.



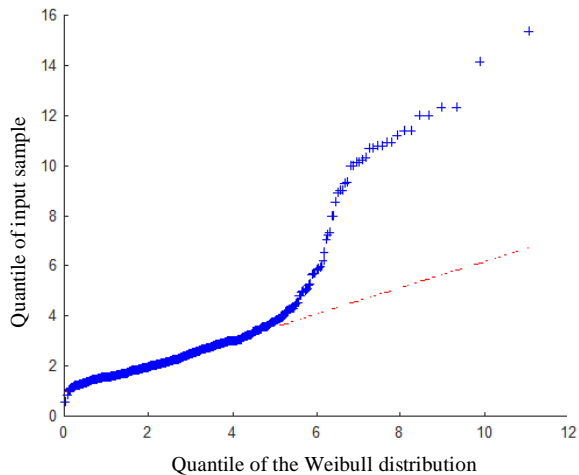
d) The Q-Q plot for the lognormal distribution.



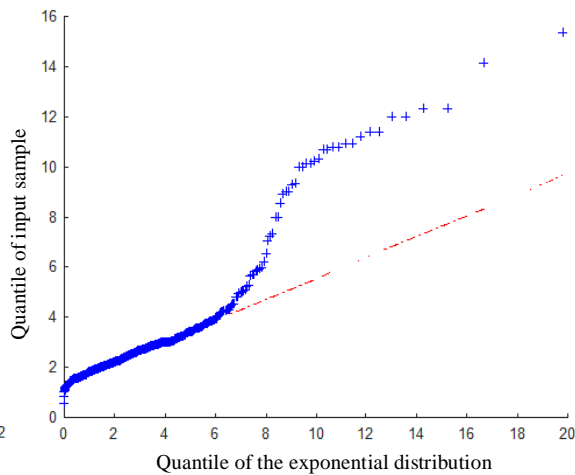
e) The Q-Q plot for the t location-scale distribution.



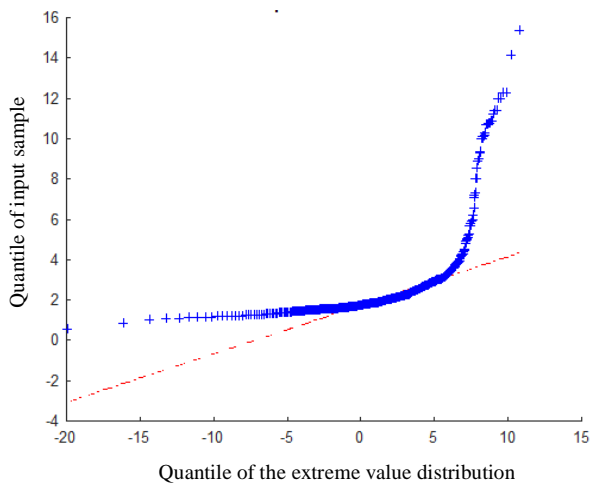
f) The Q-Q plot for the logistic distribution.



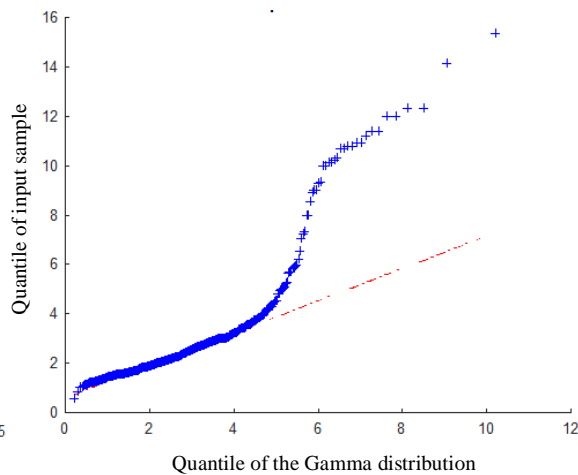
g) The Q-Q plot for the Weibull distribution.



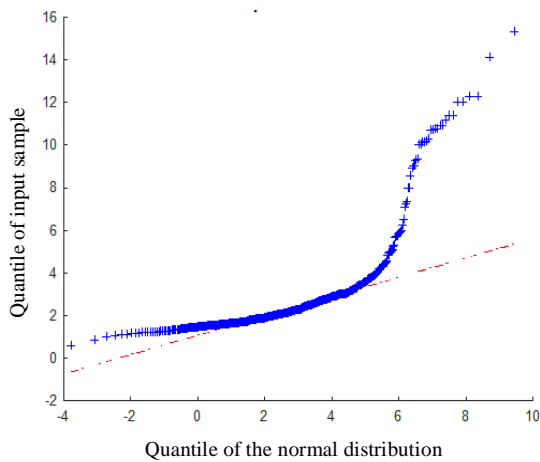
h) The Q-Q plot for the exponential distribution.



i) The Q-Q plot for the extreme value distribution.

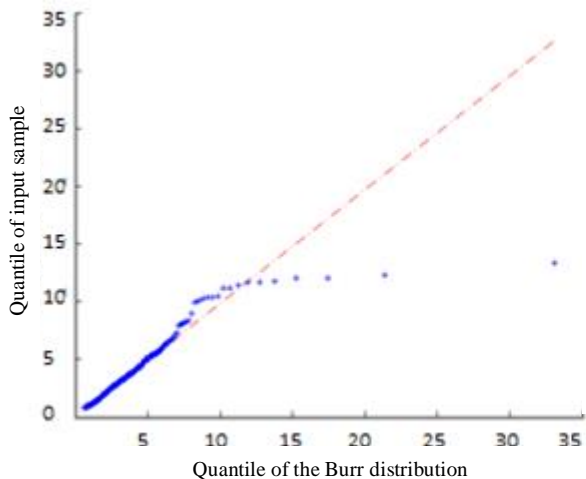


j) The Q-Q plot for the Gamma distribution.

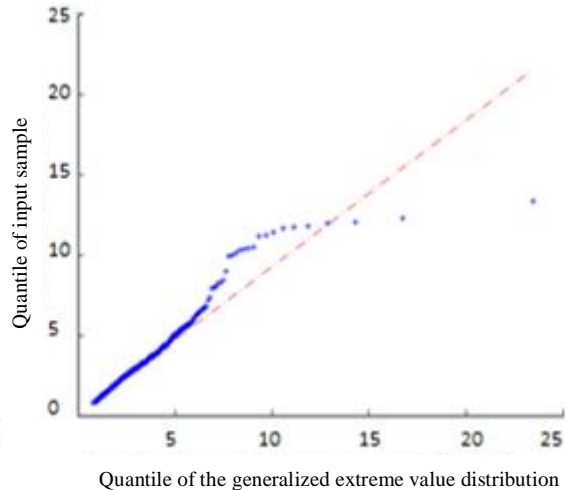


k) The Q-Q plot for the normal distribution.

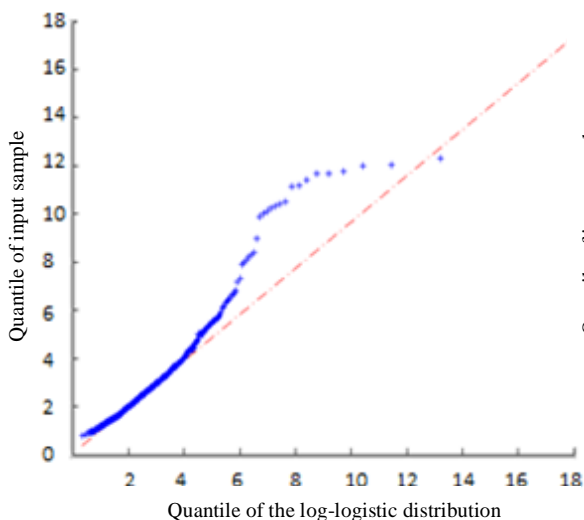
Figure 3.17: The Q-Q plots for the R2 dataset.



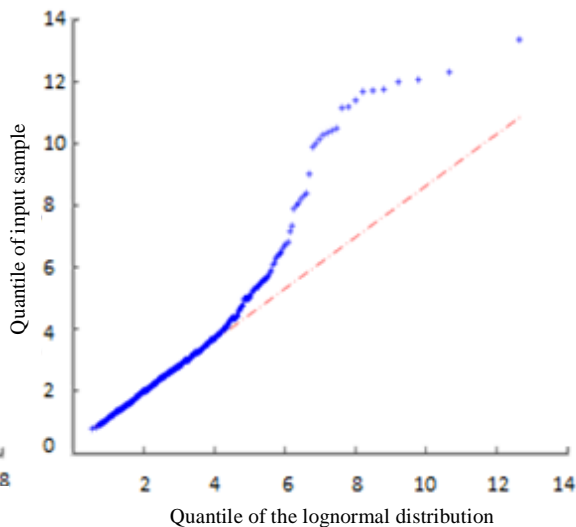
a) The Q-Q plot for the Burr distribution.



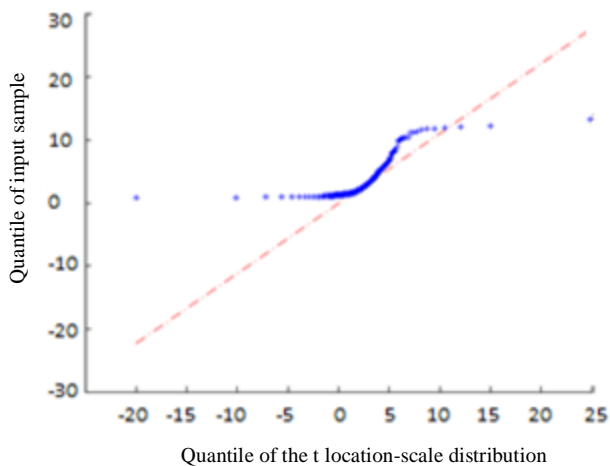
b) The Q-Q plot for the generalized extreme value distribution.



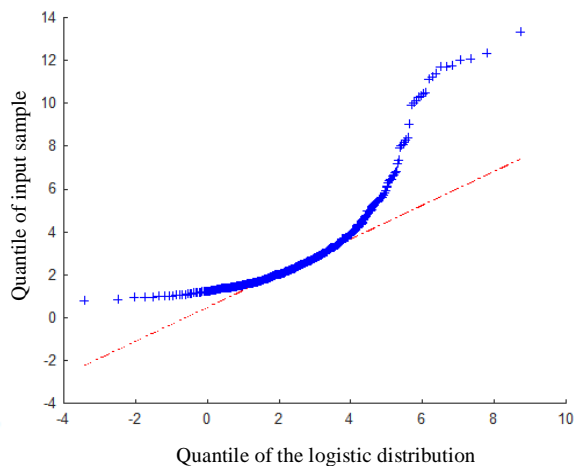
c) The Q-Q plot for the log-logistic distribution.



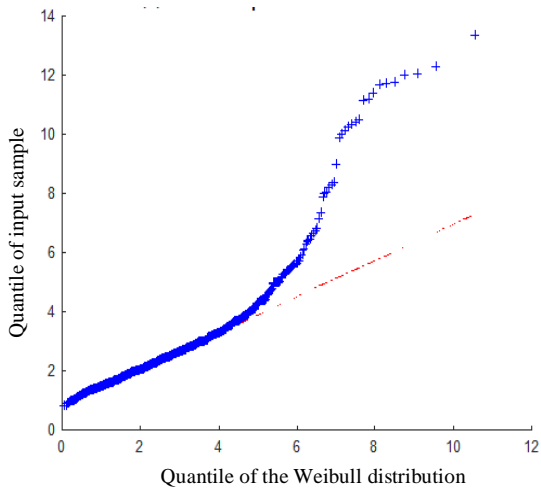
d) The Q-Q plot for the lognormal distribution.



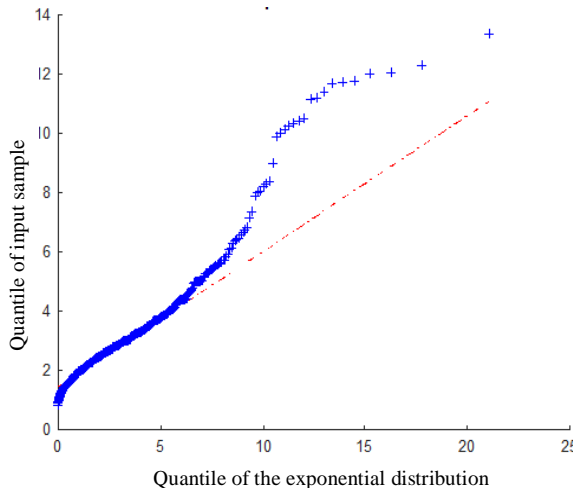
e) The Q-Q plot for the t location-scale distribution.



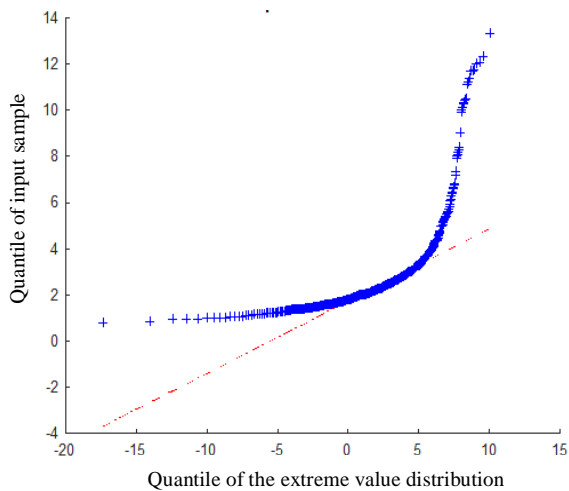
f) The Q-Q plot for the logistic distribution.



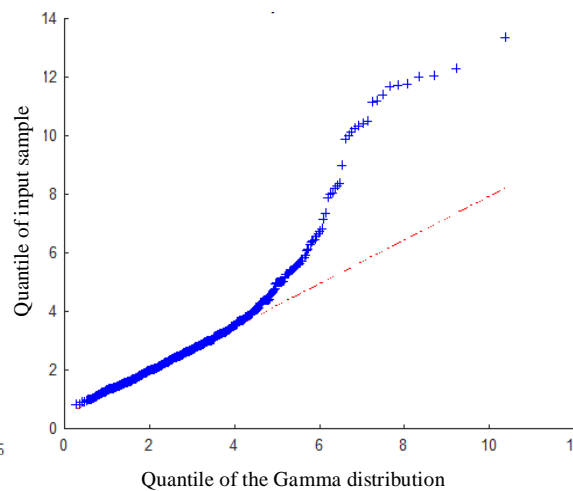
g) The Q-Q plot for the Weibull distribution.



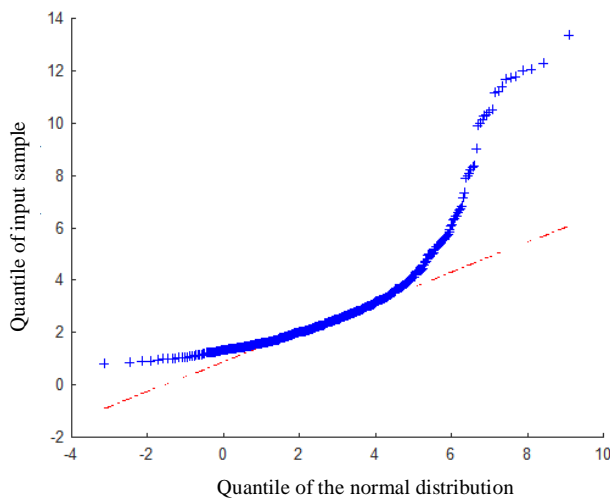
h) The Q-Q plot for the exponential distribution.



i) The Q-Q plot for the extreme value distribution.

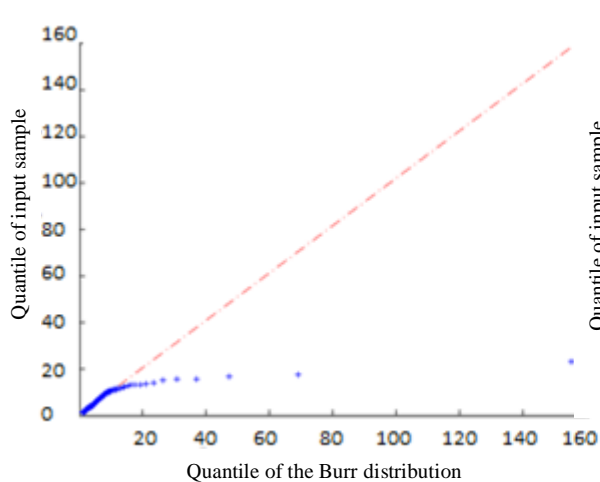


j) The Q-Q plot for the Gamma distribution.

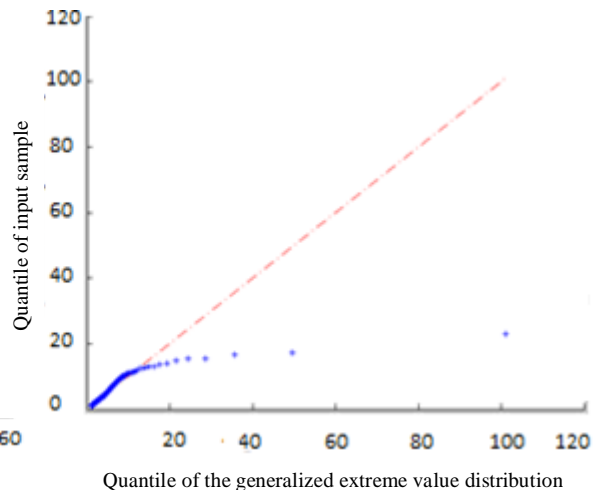


k) The Q-Q plot for the normal distribution.

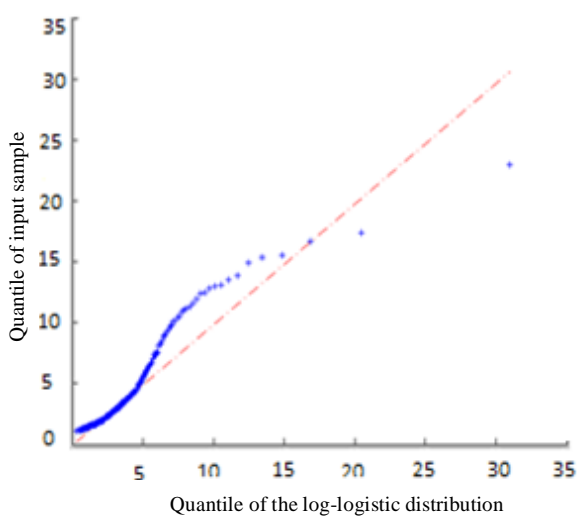
Figure 3.18: The Q-Q plots for the R3 dataset.



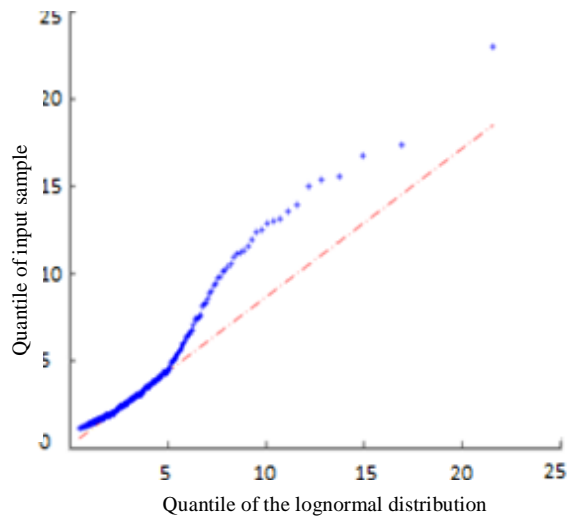
a) The Q-Q plot for the Burr distribution.



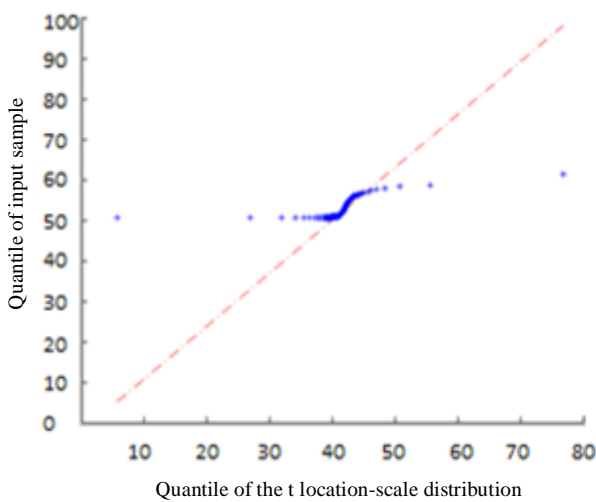
b) The Q-Q plot for the generalized extreme value distribution.



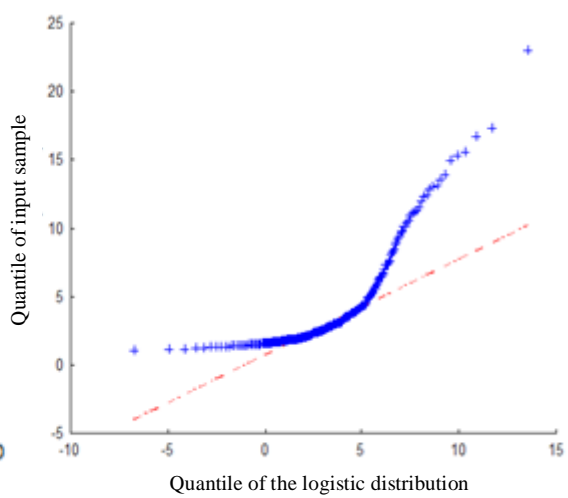
c) The Q-Q plot for the log-logistic distribution.



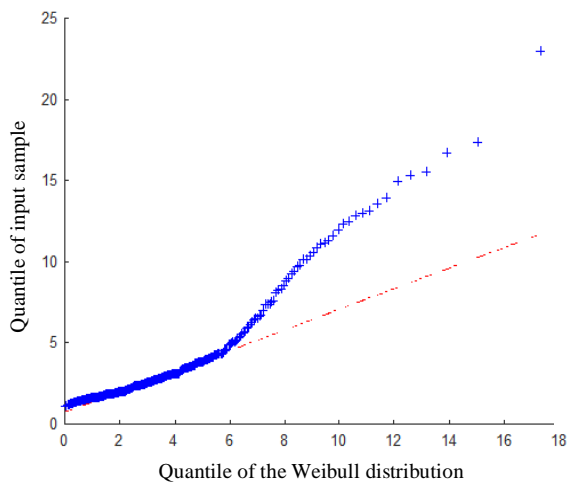
d) The Q-Q plot for the lognormal distribution.



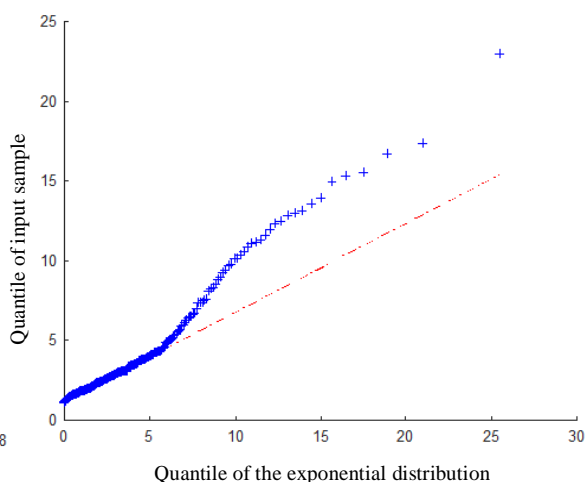
e) The Q-Q plot for the t location-scale distribution.



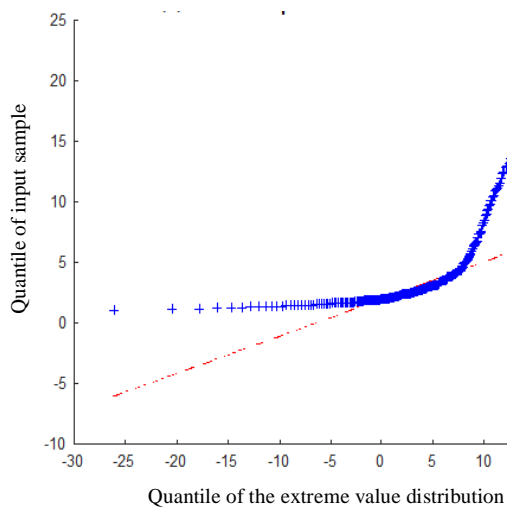
f) The Q-Q plot for the logistic distribution.



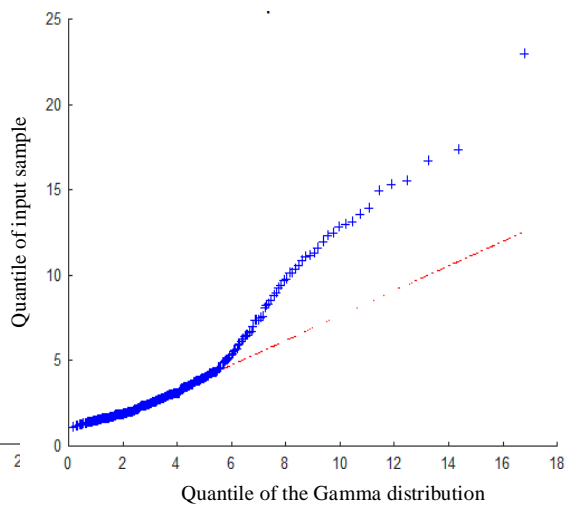
g) The Q-Q plot for the Weibull distribution.



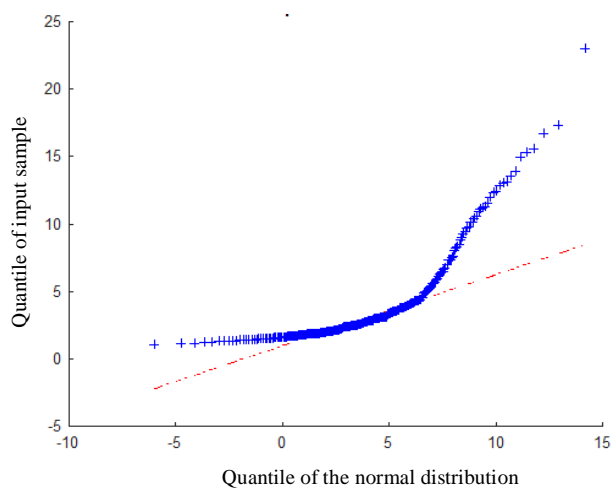
h) The Q-Q plot for the exponential distribution.



i) The Q-Q plot for the extreme value distribution.



j) The Q-Q plot for the Gamma distribution.



k) The Q-Q plot for the normal distribution.

Figure 3.19: The Q-Q plots for the R4 dataset.

In order to provide numerical results to support the visual evaluation, two statistical goodness of fit tests are used in this thesis. The Chi-Squared (C-S) and Kolmogorov-Smirnov (K-S) tests are the two most popular tests used to examine the goodness of fit of a distribution [6, 19].

### 3.2.1. The Kolmogorov-Smirnov (K-S) Test

For data,  $x_1, x_2, \dots, x_n$ , the empirical cumulative density function (ECDF) is defined as [32]

$$S_n(x) = k/n \quad (3.27)$$

where  $k$  is the number of observations less than or equal to  $x$ . The K-S test compares the ECDF of the data with the CDF of the distribution, so the K-S statistic is [32]

$$d = \max_x |F(x) - S_n(x)| \quad (3.28)$$

Then  $d$  is the maximum absolute difference between the CDF and ECDF over the entire dataset. A significance level of 5% is used to test the following hypotheses.

$H_0$ : the data comes from the distribution.

$H_1$ : the data does not come from the distribution.

To test the hypotheses, the  $p$  value associated with the K-S statistic is compared with the significance level. The  $p$  value is defined as [33]

$$p = d \sum_{j=1}^{\lfloor n(1-d) \rfloor} \binom{n}{j} \left(1 - d - \frac{j}{n}\right)^{n-j} \left(d + \frac{j}{n}\right)^{j-1} \quad (3.29)$$

If  $p$  is greater than 5%, and the distribution is accepted by the goodness of fit test and the distribution is rejected if  $p$  is smaller than 5%. Table 3.2 gives the K-S test results and  $p$  values for all distributions for each dataset.

### 3.2.1. The Chi-Squared (C-S) Test

Another statistical test for distributions is the Chi-Squared (C-S) goodness of fit test. The C-S test is used to determine if a dataset comes from a given probability distribution. This test applies to binned data so the data is divided into  $N$  bins.

The goodness of fit results are dependent on the bin size. The results obtained can be different if the bin size is too small. The Stat::fit [34] statistical software package recommends  $N = \sqrt[3]{2n}$  while in [35]  $N = 2n^{\frac{2}{5}}$  was determined to be optimum. Therefore, the latter value is used here. Then the number of data values that fall into each bin is compared to the expected number of values for that bin. An important feature of the C-S test is that it can be applied to any distribution that has a CDF. For data,  $x_1, x_2, \dots, x_n$  and PDF  $f(x)$ , the C-S test is used to determine whether there is a significant difference between the expected frequencies and the observed frequencies. The C-S test-statistic is [24]

$$\chi^2 = \sum_{i=1}^N \frac{(O_i - E_i)^2}{E_i} \quad (3.30)$$

where  $O_i$  is the observed frequency for the  $i$ th bin and  $E_i$  is the expected frequency for this bin defined as [34]

$$E_i = n \times P_i \quad (3.31)$$

$$\text{with } P_i = P[x_i < x < x_{i+1}] = F(x_{i+1}) - F(x_i) \quad (3.32)$$

where  $P_i$  is the probability that a value falls in the  $i$ th bin,  $F$  is the CDF of the distribution, and  $x_{i+1}$  and  $x_i$  are the boundaries of the  $i$ th bin. The hypotheses for this test are as follows.

$H_0$ : there is no significant difference between the expected and observed frequencies.

$H_1$ : the expected and observed frequencies are different.

The hypothesis test for a distribution was conducted with a 5% significance by comparing the test statistic to a critical value from the C-S distribution with  $N - k - 1$  degrees of freedom where  $k$  is the number of parameters in each distribution, for example for the Burr distribution,  $k = 3$ . The critical values for the C-S test statistic are given in Table A.1 in Appendix A. The  $p$  value associated with the C-S statistic ( $\chi^2$ ) is [36]

$$p = 1 - F_{\chi^2}(\chi_0^2; N - k - 1) \quad (3.33)$$

where  $F_{\chi^2}(\cdot; N - k - 1)$  is the CDF of a  $\chi^2$  distribution with  $N - k - 1$  degrees of freedom and  $\chi_0^2$  is the C-S statistics of the data. A larger  $p$  value represents a more compatible distribution for a dataset. Table 3.2 gives the C-S test results and  $p$  values for all distributions for each dataset.

Table 3.2: Goodness of fit test results for the headway data

| Probability distribution | Dataset | Chi-Squared test | $p$       | Kolmogorov-Smirnov test | $p$   |
|--------------------------|---------|------------------|-----------|-------------------------|-------|
| <b>Burr</b>              | N1      | accepted         | 0.069     | accepted                | 0.772 |
|                          | N2      | accepted         | 0.421     | accepted                | 0.156 |
|                          | N3      | rejected         | 9.455e-05 | accepted                | 0.222 |
|                          | N4      | accepted         | 0.324     | accepted                | 0.438 |
|                          | R1      | accepted         | 0.099     | accepted                | 0.170 |

|                                  |    |          |           |          |           |
|----------------------------------|----|----------|-----------|----------|-----------|
|                                  | R2 | accepted | 0.603     | accepted | 0.499     |
|                                  | R3 | accepted | 0.657     | accepted | 0.790     |
|                                  | R4 | rejected | 0.031     | accepted | 0.313     |
| <b>Generalized extreme value</b> | N1 | rejected | 2.656e-06 | rejected | 0.015     |
|                                  | N2 | rejected | 0.011     | accepted | 0.235     |
|                                  | N3 | accepted | 0.072     | accepted | 0.828     |
|                                  | N4 | rejected | 0.031     | accepted | 0.602     |
|                                  | R1 | accepted | 0.193     | accepted | 0.824     |
|                                  | R2 | accepted | 0.053     | accepted | 0.817     |
|                                  | R3 | accepted | 0.067     | accepted | 0.813     |
|                                  | R4 | accepted | 0.072     | accepted | 0.883     |
| <b>Log-logistic</b>              | N1 | rejected | 7.319e-14 | rejected | 9.051e-04 |
|                                  | N2 | rejected | 1.154e-08 | rejected | 0.002     |
|                                  | N3 | rejected | 1.228e-04 | rejected | 0.001     |
|                                  | N4 | rejected | 8.190e-10 | rejected | 0.014     |
|                                  | R1 | rejected | 1.538e-05 | rejected | 0.042     |
|                                  | R2 | rejected | 1.019e-13 | accepted | 0.239     |
|                                  | R3 | rejected | 4.617e-04 | accepted | 0.649     |
|                                  | R4 | rejected | 1.366e-05 | accepted | 0.099     |
| <b>Lognormal</b>                 | N1 | rejected | 4.091e-15 | rejected | 2.653e-10 |
|                                  | N2 | rejected | 5.171e-19 | rejected | 5.751e-05 |
|                                  | N3 | rejected | 8.161e-11 | rejected | 3.984e-05 |
|                                  | N4 | rejected | 1.924e-13 | rejected | 3.301e-06 |
|                                  | R1 | rejected | 1.316e-09 | rejected | 0.003     |
|                                  | R2 | rejected | 2.400e-32 | rejected | 6.096e-06 |
|                                  | R3 | rejected | 5.906e-09 | rejected | 0.014     |
|                                  | R4 | rejected | 1.252e-05 | rejected | 0.006     |
| <b>t location-scale</b>          | N1 | rejected | 1.325e-14 | rejected | 3.120e-10 |
|                                  | N2 | rejected | 3.289e-20 | rejected | 4.136e-17 |
|                                  | N3 | rejected | 6.400e-20 | rejected | 9.487e-20 |
|                                  | N4 | rejected | 4.983e-12 | rejected | 1.706e-10 |
|                                  | R1 | rejected | 9.778e-13 | rejected | 9.512e-11 |
|                                  | R2 | rejected | 2.621e-13 | rejected | 4.052e-09 |
|                                  | R3 | rejected | 3.081e-09 | rejected | 6.761e-06 |
|                                  | R4 | rejected | 2.034e-12 | rejected | 4.004e-08 |
| <b>Logistic</b>                  | N1 | rejected | 2.987e-34 | rejected | 9.334e-20 |
|                                  | N2 | rejected | 3.625e-36 | rejected | 2.712e-20 |
|                                  | N3 | rejected | 3.403e-36 | rejected | 2.123e-19 |
|                                  | N4 | rejected | 3.161e-23 | rejected | 1.716e-14 |
|                                  | R1 | rejected | 9.127e-19 | rejected | 2.465e-13 |
|                                  | R2 | rejected | 7.754e-12 | rejected | 2.785e-14 |
|                                  | R3 | rejected | 4.246e-15 | rejected | 1.857e-07 |
|                                  | R4 | rejected | 2.879e-10 | rejected | 4.014e-09 |
| <b>Weibull</b>                   | N1 | rejected | 2.586e-27 | rejected | 7.038e-28 |
|                                  | N2 | rejected | 1.106e-26 | rejected | 2.951e-19 |

|                      |    |          |            |          |           |
|----------------------|----|----------|------------|----------|-----------|
|                      | N3 | rejected | 8.503e-16  | rejected | 7.650e-13 |
|                      | N4 | rejected | 5.912e-14  | rejected | 4.139e-16 |
|                      | R1 | rejected | 5.708e-14  | rejected | 1.951e-13 |
|                      | R2 | rejected | 4.131e-31  | rejected | 8.772e-16 |
|                      | R3 | rejected | 2.236e-23  | rejected | 5.513e-09 |
|                      | R4 | rejected | 2.689e-07  | rejected | 5.060e-05 |
| <b>Exponential</b>   | N1 | rejected | 6.148e-16  | rejected | 2.913e-74 |
|                      | N2 | rejected | 5.215e-64  | rejected | 6.437e-76 |
|                      | N3 | rejected | 7.1730e-10 | rejected | 2.632e-52 |
|                      | N4 | rejected | 5.537e-08  | rejected | 2.214e-52 |
|                      | R1 | rejected | 6.911e-09  | rejected | 1.090e-50 |
|                      | R2 | rejected | 1.452e-36  | rejected | 7.359e-52 |
|                      | R3 | rejected | 2.081e-50  | rejected | 1.053e-44 |
|                      | R4 | rejected | 1.351e-23  | rejected | 3.135e-14 |
| <b>Extreme value</b> | N1 | rejected | 2.291e-50  | rejected | 6.182e-86 |
|                      | N2 | rejected | 9.411e-176 | rejected | 5.238e-89 |
|                      | N3 | rejected | 5.014e-43  | rejected | 2.728e-68 |
|                      | N4 | rejected | 9.591e-35  | rejected | 3.140e-65 |
|                      | R1 | rejected | 5.606e-38  | rejected | 1.669e-66 |
|                      | R2 | rejected | 3.825e-93  | rejected | 2.588e-53 |
|                      | R3 | rejected | 8.978e-88  | rejected | 7.780e-43 |
|                      | R4 | rejected | 8.398e-12  | rejected | 1.790e-22 |
| <b>Gamma</b>         | N1 | rejected | 4.105e-22  | rejected | 5.003e-19 |
|                      | N2 | rejected | 2.386e-24  | rejected | 1.890e-12 |
|                      | N3 | rejected | 7.170e-21  | rejected | 1.657e-08 |
|                      | N4 | rejected | 8.876e-21  | rejected | 4.516e-12 |
|                      | R1 | rejected | 1.877e-19  | rejected | 3.445e-08 |
|                      | R2 | rejected | 4.316e-18  | rejected | 1.313e-12 |
|                      | R3 | rejected | 8.117e-13  | rejected | 2.556e-06 |
|                      | R4 | rejected | 3.079e-09  | rejected | 1.532e-05 |
| <b>Normal</b>        | N1 | rejected | 7.007e-49  | rejected | 1.355e-37 |
|                      | N2 | rejected | 1.540e-62  | rejected | 4.897e-30 |
|                      | N3 | rejected | 2.272e-26  | rejected | 3.953e-22 |
|                      | N4 | rejected | 1.324e-28  | rejected | 9.790e-24 |
|                      | R1 | rejected | 8.201e-34  | rejected | 5.129e-20 |
|                      | R2 | rejected | 5.337e-31  | rejected | 7.160e-29 |
|                      | R3 | rejected | 4.203e-31  | rejected | 2.043e-17 |
|                      | R4 | rejected | 3.611e-14  | rejected | 2.389e-11 |

If a  $p$  value in Table 3.2 is less than 0.05, the distribution is rejected by the goodness of fit test, and if this value is larger than 0.05, the distribution is accepted. According to the K-S and C-

S test results, the Burr distribution was accepted by both tests for the N1, N2, and N4 datasets. Thus, it provides the best fit for the headway data in clear weather with higher traffic flow rates (more than 1600 vphpl). This distribution was also accepted by both tests for the R1, R2, and R3 datasets. Thus, the Burr distribution passed both goodness of fit tests at 5% significance for six of the eight datasets in clear weather with traffic flow rates higher than 1600 vphpl and in rainy weather with traffic flow rates higher than 1000 vphpl.

The generalized extreme value distribution provides the best fit for the R1 dataset as it has a higher  $p$  value for both tests than the Burr distribution. The C-S test  $p$  values for the Burr and generalized extreme value distributions are 0.099 and 0.193, respectively, and for the K-S test are 0.170 and 0.824. This distribution also provides the best fit for the N3 and R4 datasets. The generalized extreme value distribution passed both goodness of fit tests at 5% significance for the datasets with the lowest traffic flow rates (less than 1600 vphpl in clear weather and less than 1000 vphpl in rainy weather). These results indicate that the headway in clear weather conditions with low traffic flow rates and low average speeds has a similar distribution to the headway in rainy weather.

The generalized extreme value distribution also passed both goodness of fit tests at 5% significance for the R2 and R3 datasets. The generalized extreme value distribution was accepted by the C-S test at 5% significance for five of the datasets. The C-S test results at 5% significance indicate that the generalized extreme value distribution fits the headway data with lower traffic flow rates. Further, the K-S and C-S test results support the Q-Q plot results which indicated that the Burr and generalized extreme value distributions are the best for headway data.

The log-logistic distribution passed just the K-S test at 5% significance for the rainy weather datasets with low traffic flow rates less than 1300 vphpl (R2, R3, and R4). All the other distributions failed the goodness of fit tests at 5% significance for all eight datasets. Thus, the lognormal, t location-scale, logistic, Weibull, exponential, extreme value, Gamma, and normal distributions are not suitable to model headway data.

## Chapter 4

### Analysis of New Clear Weather Data

A highway similar to Emam Ali highway was considered to validate the results for traffic in clear weather in Chapter 3. The new headway data was obtained from the Sayad Shirazi Highway, Tehran, Iran. This is a six-lane highway that runs in a north-south direction. The speed limit on this highway is 90 kilometers per hour (km/h). The location of the data collection point on Google maps is shown in Figure 4.1, and Figure 4.2 shows the location of the new highway relative to the previous one.

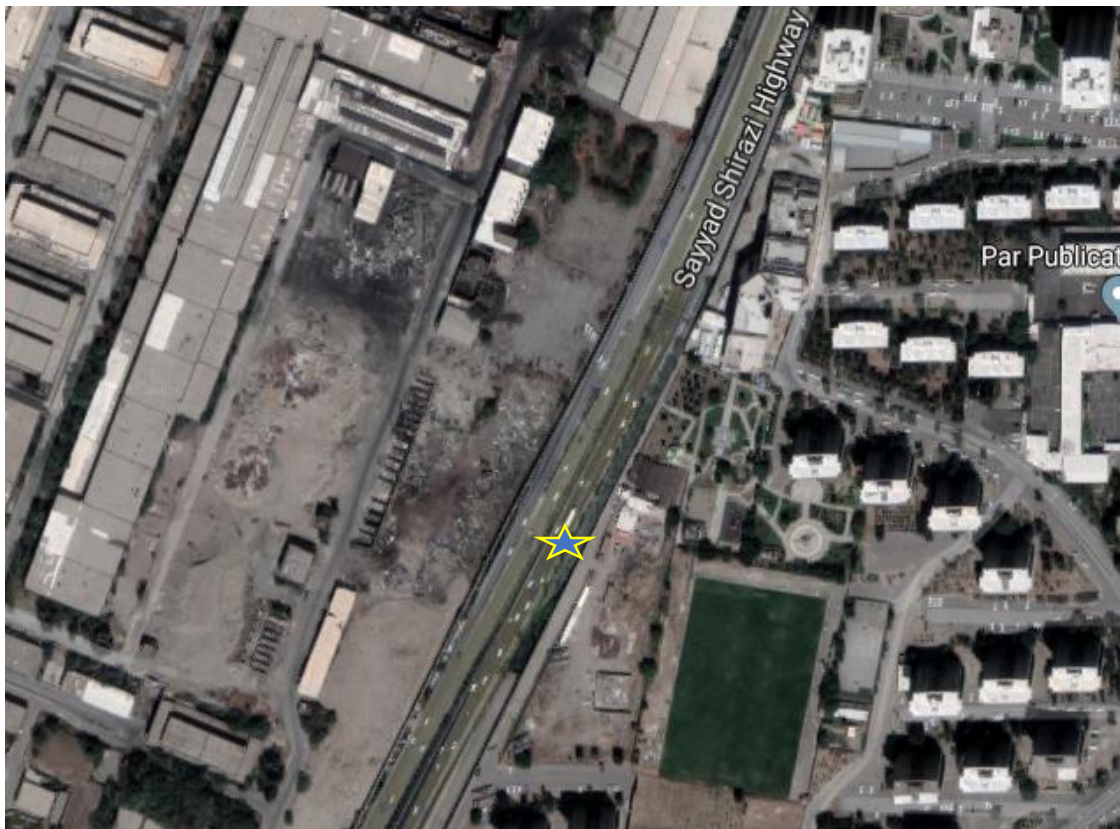


Figure 4.1: The location of the new data collection point on Google maps [37].

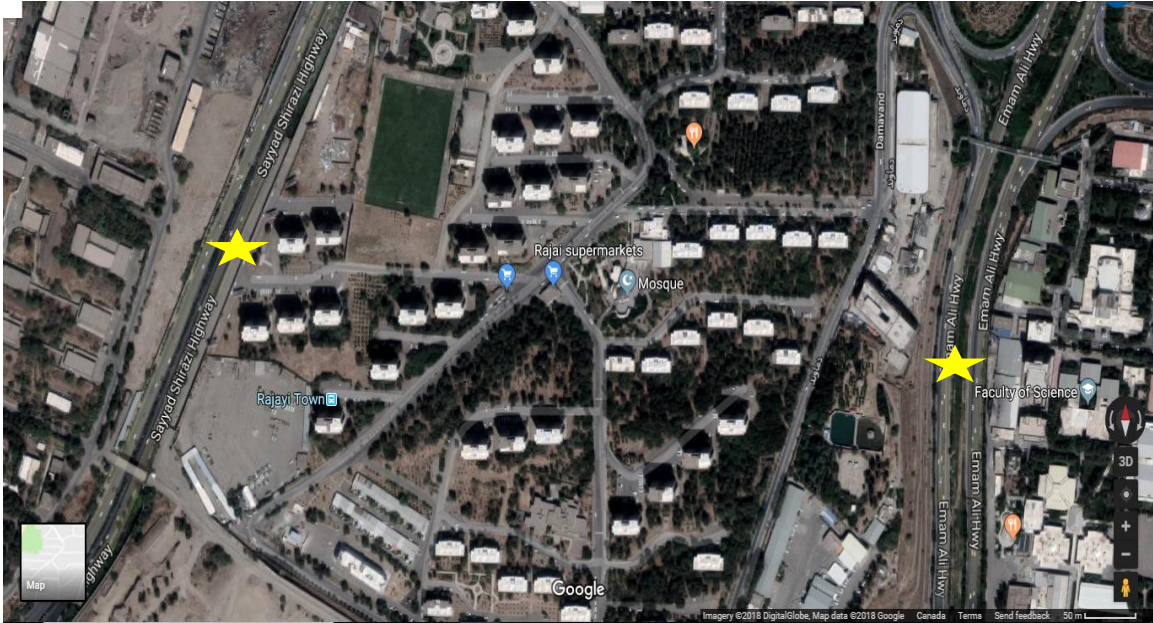


Figure 4.2: The locations on the two highways on Google maps where headway data was collected [38].



Figure 4.3: An image taken from the video data showing the reference point.

Video from a location on a 1.5 km straight highway section was used to obtain traffic data. The highway light pole shown in Figure 4.3 serves as a reference point to extract headway data. The video was recorded on a weekday in clear weather at the same time as on the Emam Ali highway, between 12:20 and 13:00, on July 30, 2018. To be consistent with the previous results, headway data was extracted only from the two left lanes in one direction. The video was divided into 5 min sections for analysis. Based on the traffic flow rate in each 5 min section, the data was divided into three datasets as indicated in Table 4.1. The number of data values is 2626.

Table 4.1: New dataset parameters

| <b>Dataset</b> | <b>Weather condition</b> | <b>Traffic flow rate<br/>(vphpl)</b> |
|----------------|--------------------------|--------------------------------------|
| T1             | Clear                    | >2100                                |
| T2             | Clear                    | Between 1700-2100                    |
| T3             | Clear                    | <1700                                |

The statistical parameters were obtained for each dataset and the results are shown in Table 4.2. The mean and median headway increase with a decrease in traffic flow rate as do the variance and standard deviation. This table shows that the median headway is smaller than the mean in all datasets and this difference decreases with increasing traffic flow rate. Thus, the median headway is smaller than the mean in all datasets and this difference decreases with increasing flow rate.

Compared to the Emam Ali highway data, the new data has a higher traffic flow rate, so there is a decrease in the mean headway in all datasets in clear weather. The lower variance and standard deviation also show that the headway in datasets T1 to T3 tends to be very close to the mean of headway. A comparison of the mean and median for all three datasets

shows that the mean is larger than the median which indicates that most vehicles have a headway less than the mean and there is a high concentration of short headways. Comparing of Tables 2.2 and 4.2 indicates that the max value, variance and standard deviation of headway in the T1, T2, and T3 datasets compared to those of the datasets obtained previously have decreased because of the increased traffic flow rates and the more constant headway.

Table 4.2: Statistical characteristics of the new headway data

| Dataset     | T1   | T2   | T3    |
|-------------|------|------|-------|
| Sample size | 1148 | 1099 | 379   |
| Mean        | 1.43 | 1.72 | 2.14  |
| Median      | 1.32 | 1.56 | 1.63  |
| Variance    | 0.35 | 0.52 | 3.32  |
| Std Dev     | 0.59 | 0.72 | 1.82  |
| Max         | 8.67 | 9.82 | 16.33 |
| Min         | 0.16 | 0.53 | 0.46  |

#### 4.1. Distribution Parameter Estimation

Based on the results obtained in Chapter 3, the  $p$  values for the goodness of fit tests for the extreme value and exponential distributions were near to zero, thus these distributions were not fit to the new datasets in this chapter. To find the best distributions, the Burr, generalized extreme value, log-logistic, lognormal, t location-scale, logistic, Weibull, Gamma, and normal distributions were fit to each headway dataset using the MLE technique. The PDFs of these distributions and the histogram of the new headway data for

each dataset are presented in Figures 4.4 to 4.6. The distribution parameters obtained for the three datasets are given in Table 4.3.

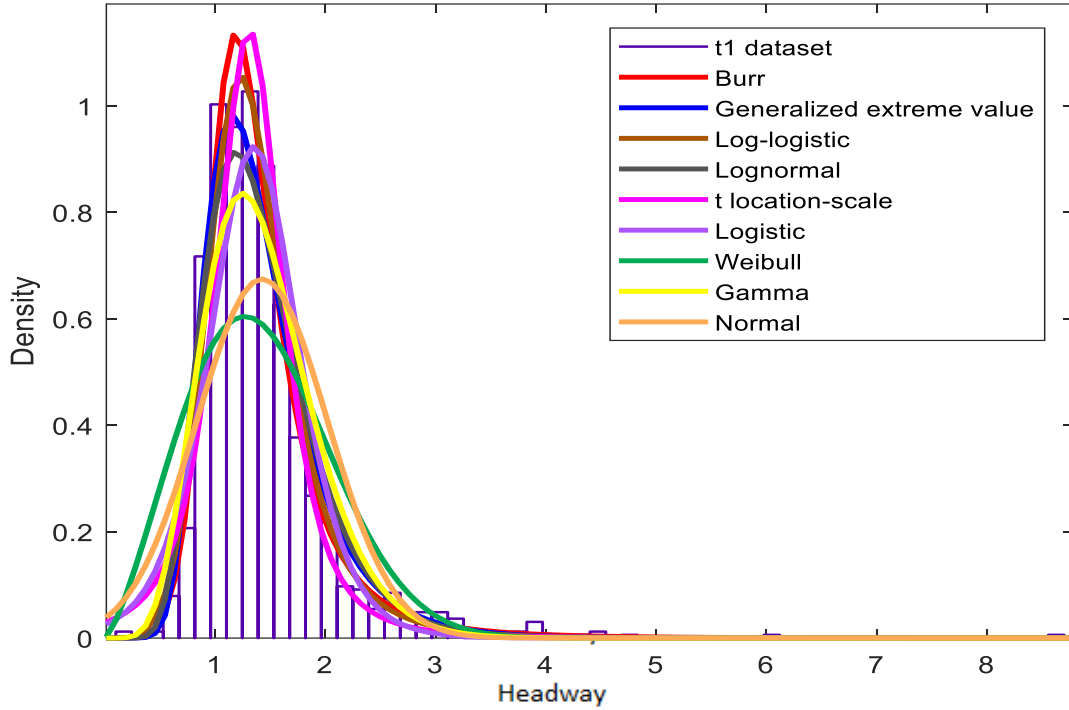


Figure 4.4: The PDFs of all distributions for the T1 dataset.

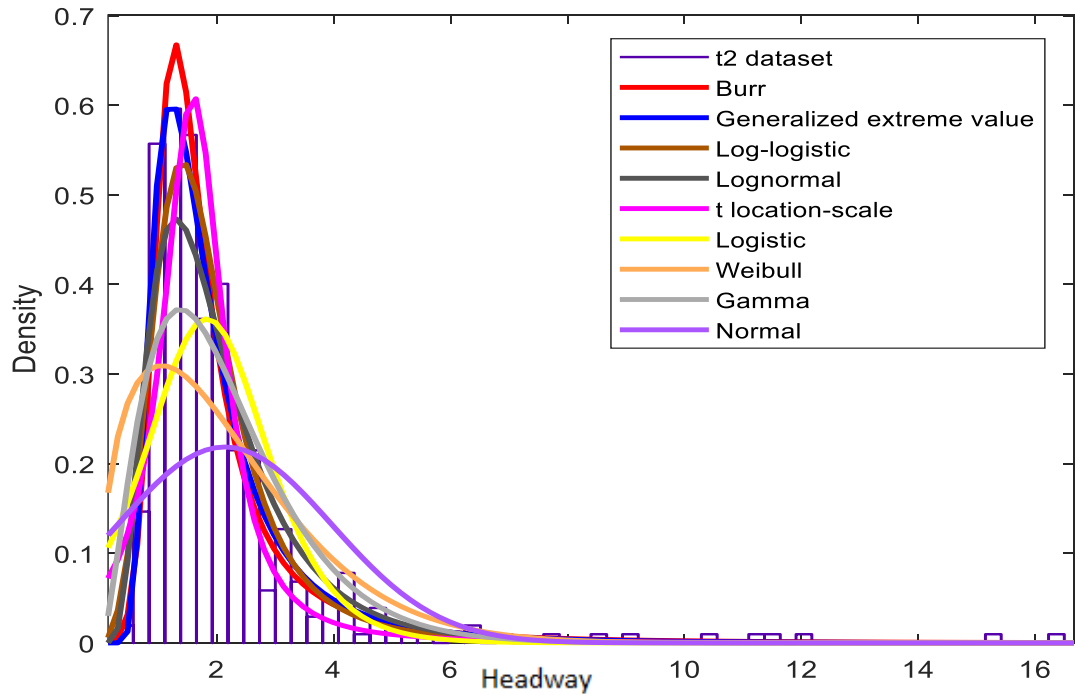


Figure 4.5: The PDFs of all distributions for the T2 dataset.

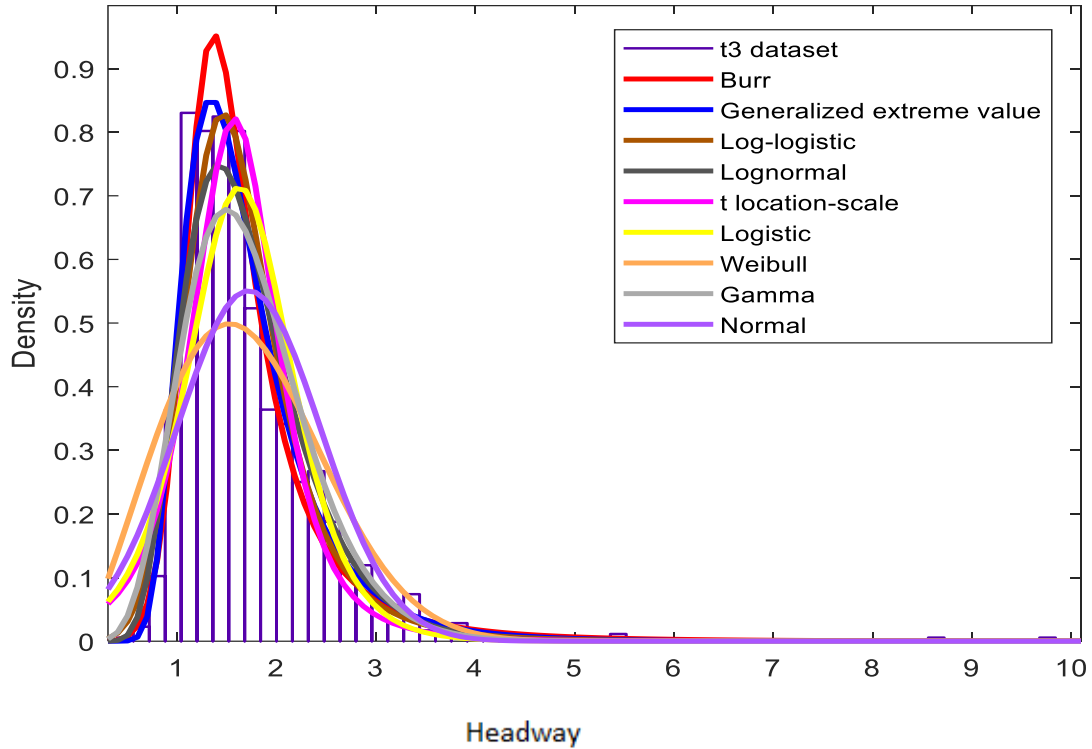


Figure 4.6: The PDFs of all distributions for the T3 dataset.

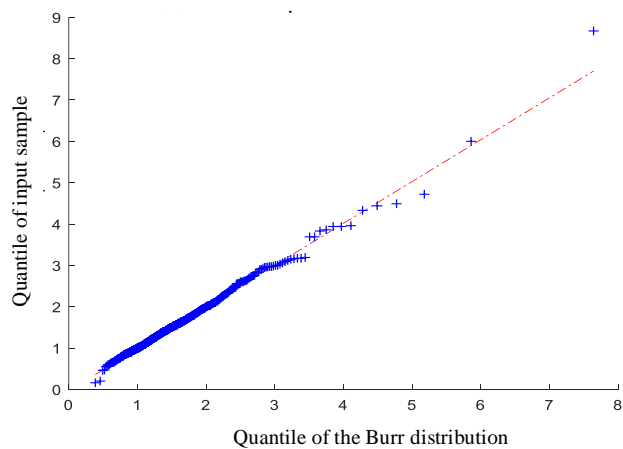
Table 4.3: Statistical distribution parameters for the new headway data

| Probability distribution         | Dataset | Shape Parameters      | Scale parameter | Location parameter |
|----------------------------------|---------|-----------------------|-----------------|--------------------|
| <b>Burr</b>                      | T1      | $k = 0.63$ $c = 6.55$ | $\alpha = 1.18$ | -                  |
|                                  | T2      | $k = 0.40$ $c = 5.24$ | $\alpha = 1.21$ | -                  |
|                                  | T3      | $k = 0.49$ $c = 7.05$ | $\alpha = 1.31$ | -                  |
| <b>Generalized extreme value</b> | T1      | $k = 0.06$            | $\alpha = 0.38$ | $\beta = 1.19$     |
|                                  | T2      | $k = 0.35$            | $\alpha = 0.64$ | $\beta = 1.40$     |
|                                  | T3      | $k = 0.14$            | $\alpha = 0.44$ | $\beta = 1.40$     |
| <b>Log-logistic</b>              | T1      | $k = 0.28$            | $\alpha = 0.19$ | -                  |
|                                  | T2      | $k = 0.53$            | $\alpha = 0.30$ | -                  |
|                                  | T3      | $k = 0.46$            | $\alpha = 0.20$ | -                  |
| <b>Lognormal</b>                 | T1      | $k = 0.29$            | $\alpha = 0.35$ | -                  |
|                                  | T2      | $k = 0.57$            | $\alpha = 0.56$ | -                  |
|                                  | T3      | $k = 0.48$            | $\alpha = 0.35$ | -                  |
| <b>t location-scale</b>          | T1      | $k = 1.31$            | $\alpha = 0.31$ | $\beta = 2.45$     |
|                                  | T2      | $k = 1.62$            | $\alpha = 0.60$ | $\beta = 2.36$     |
|                                  | T3      | $k = 1.58$            | $\alpha = 0.45$ | $\beta = 3.32$     |
| <b>Logistic</b>                  | T1      | -                     | $\alpha = 0.27$ | $\beta = 1.35$     |
|                                  | T2      | -                     | $\alpha = 0.69$ | $\beta = 1.83$     |
|                                  | T3      | -                     | $\alpha = 0.35$ | $\beta = 1.63$     |

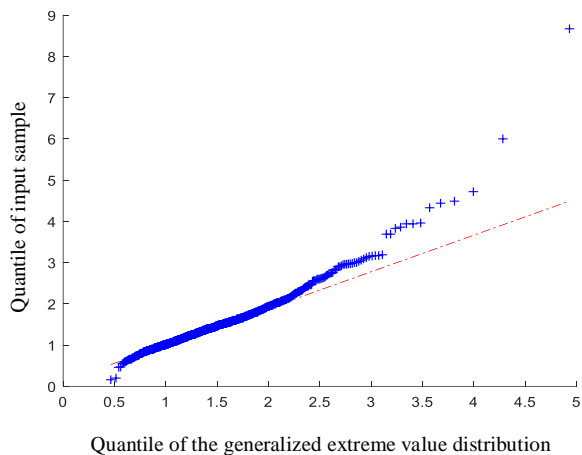
|                |    |            |                 |   |
|----------------|----|------------|-----------------|---|
| <b>Weibull</b> | T1 | $k = 2.37$ | $\alpha = 1.60$ | - |
|                | T2 | $k = 1.45$ | $\alpha = 2.39$ | - |
|                | T3 | $k = 2.36$ | $\alpha = 1.93$ | - |
| <b>Gamma</b>   | T1 | $k = 8.01$ | $\alpha = 0.18$ | - |
|                | T2 | $k = 2.80$ | $\alpha = 0.76$ | - |
|                | T3 | $k = 7.63$ | $\alpha = 0.23$ | - |
| <b>Normal</b>  | T1 | $k = 0.59$ | $\alpha = 1.43$ | - |
|                | T2 | $k = 1.82$ | $\alpha = 2.14$ | - |
|                | T3 | $k = 0.72$ | $\alpha = 1.72$ | - |

## 4.2. Goodness of Fit

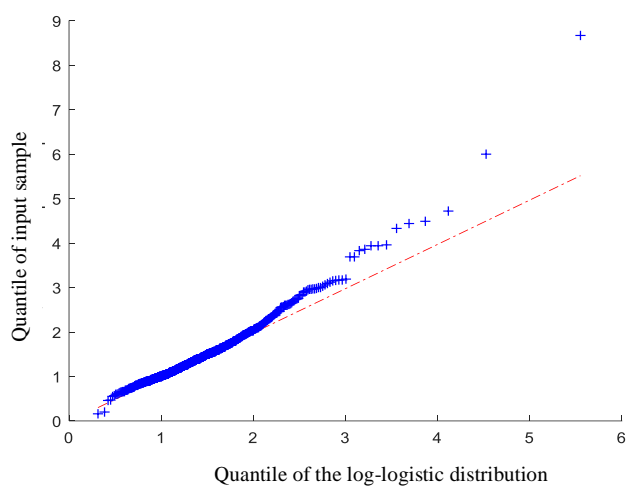
The best distributions for each dataset are determined using the goodness of fit tests. The Q-Q plots of the nine distributions for the three datasets are presented in Figures 4.7 to 4.9. Figure 4.7 gives the Q-Q plots of the nine distributions for the T1 dataset. This shows that the Burr distribution is a better fit for the T1 dataset compared to the other eight distributions. Figure 4.8 shows that the generalized extreme value distribution is closer to the 45 degree line compared to the Q-Q plots of the other eight distributions. Figure 4.9 shows that the Q-Q plots for the lognormal and generalized extreme value distributions are closer to this line than the other distributions. The nonlinear shape of the Q-Q plots for the t location-scale, logistic, Weibull, Gamma, and normal distributions indicates that they are not suitable for these headway datasets.



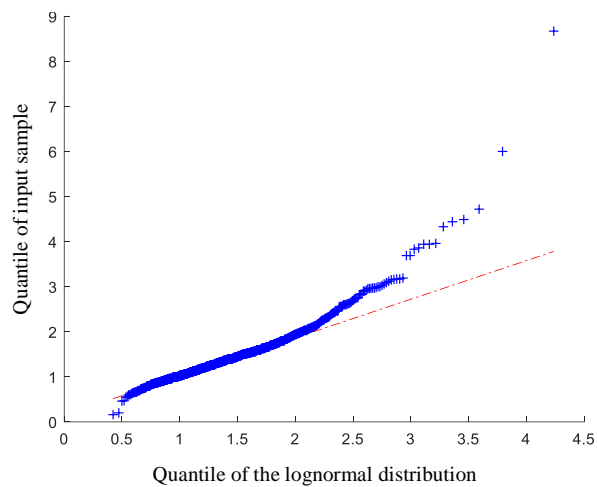
a) The Q-Q plot for the Burr distribution.



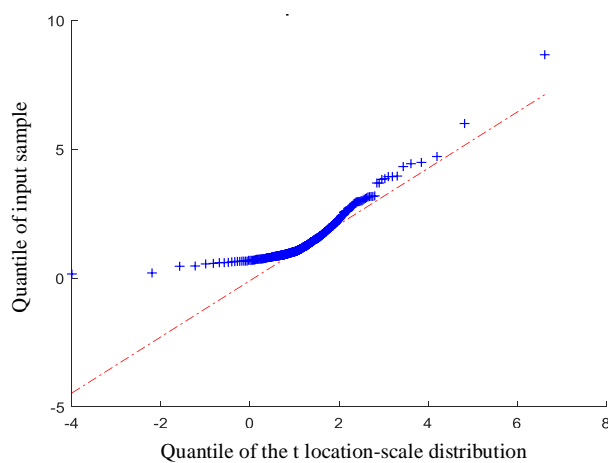
b) The Q-Q plot for the generalized extreme value distribution.



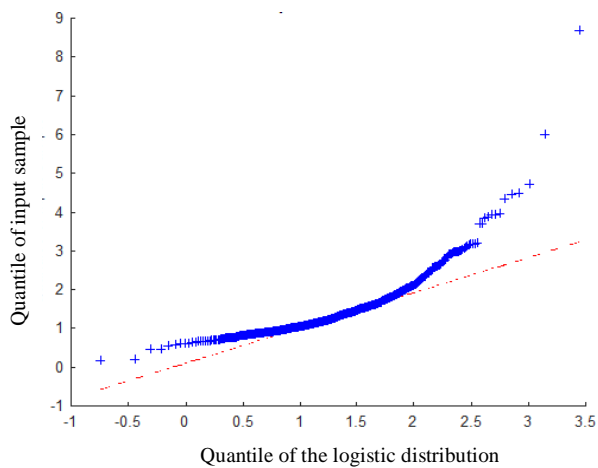
c) The Q-Q plot for the log-logistic distribution.



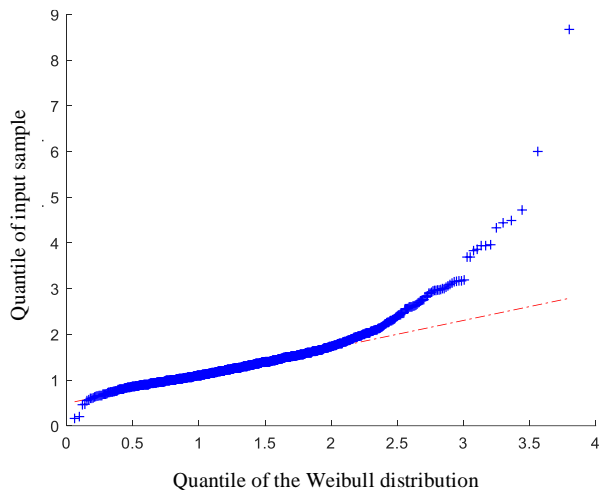
d) The Q-Q plot for the lognormal distribution.



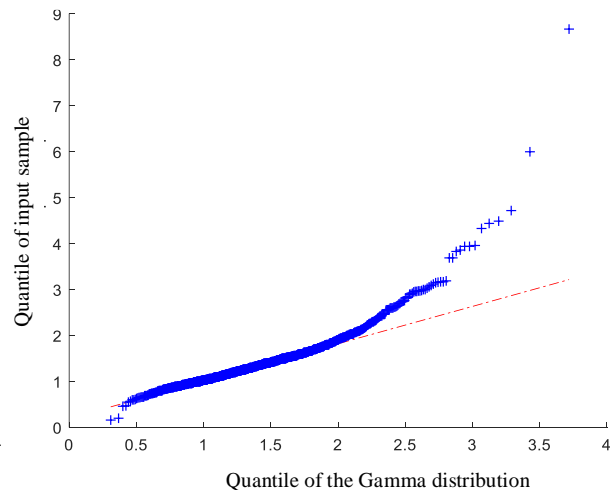
e) The Q-Q plot for the t location-scale distribution.



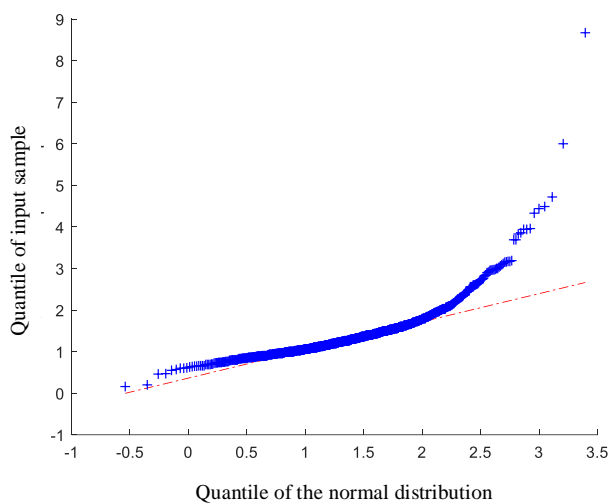
f) The Q-Q plot for the logistic distribution



g) The Q-Q plot for the Weibull distribution.

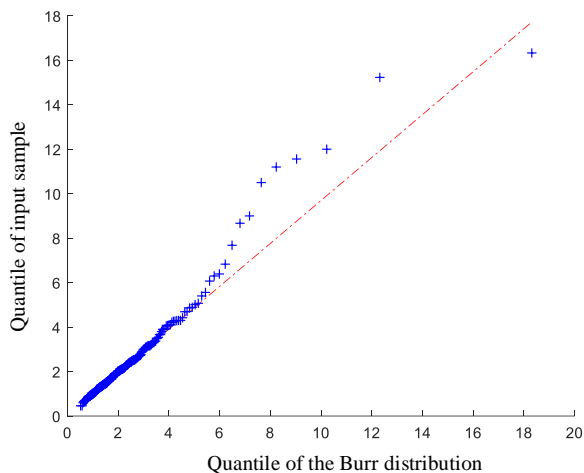


h) The Q-Q plot for the Gamma distribution.

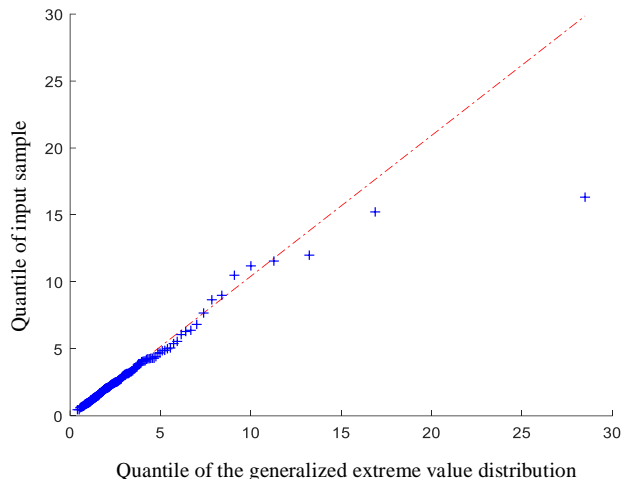


i) The Q-Q plot for the normal distribution.

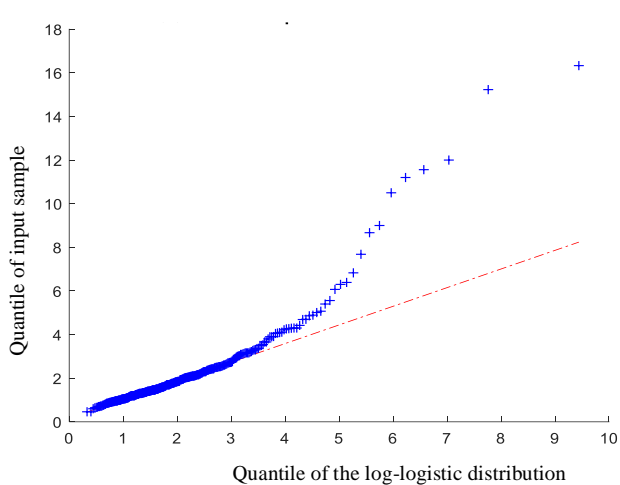
Figure 4.7: The Q-Q plots for the T1 dataset.



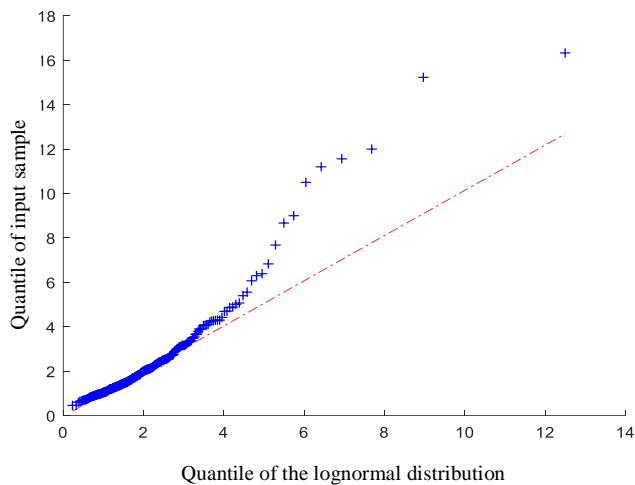
a) The Q-Q plot for the Burr distribution.



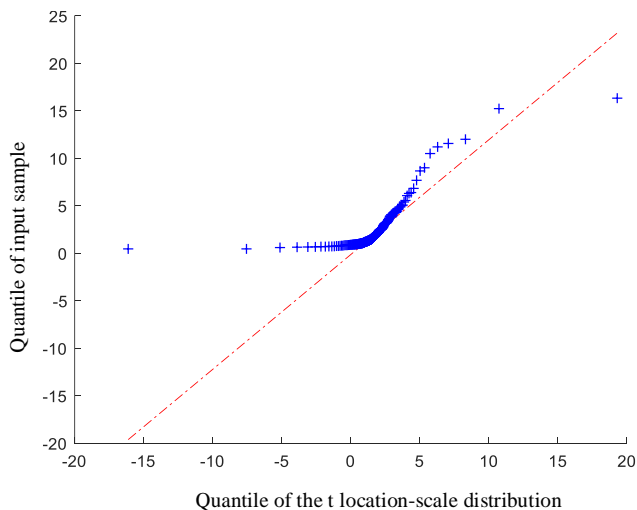
b) The Q-Q plot for the generalized extreme value distribution.



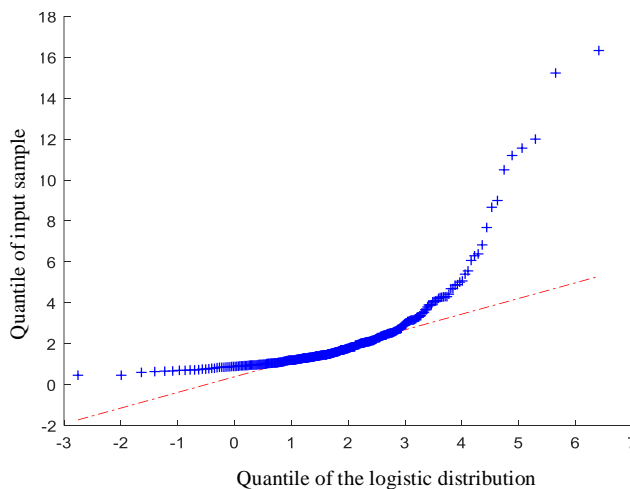
c) The Q-Q plot for the log-logistic distribution.



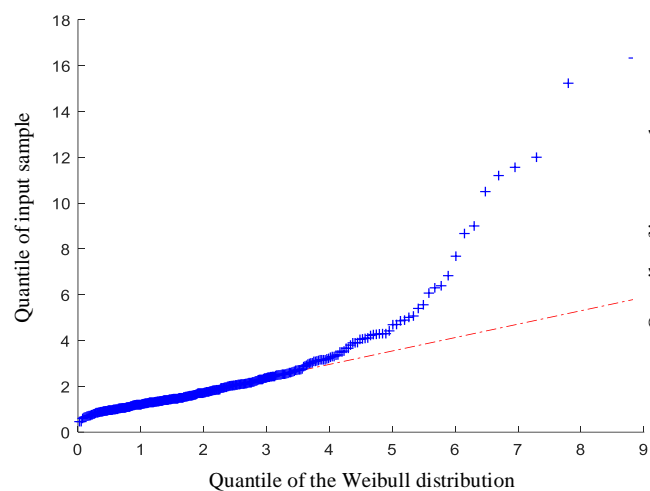
d) The Q-Q plot for the lognormal distribution.



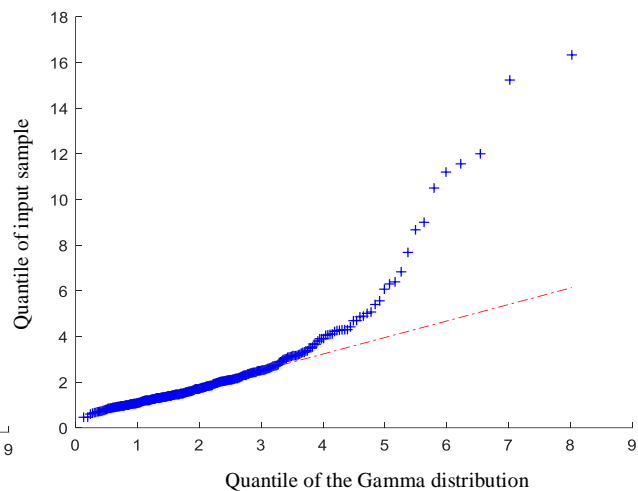
e) The Q-Q plot for the t location-scale distribution.



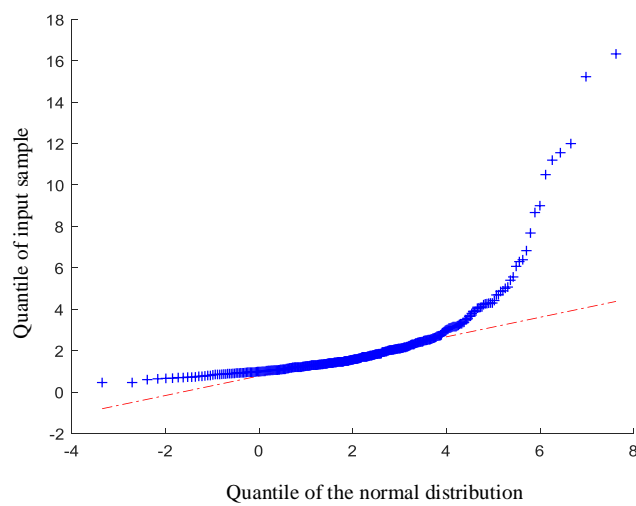
f) The Q-Q plot for the logistic distribution



g) The Q-Q plot for the Weibull distribution.

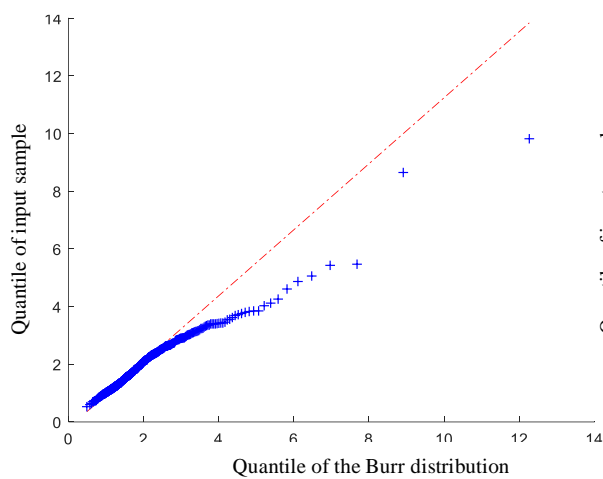


h) The Q-Q plot for the Gamma distribution.

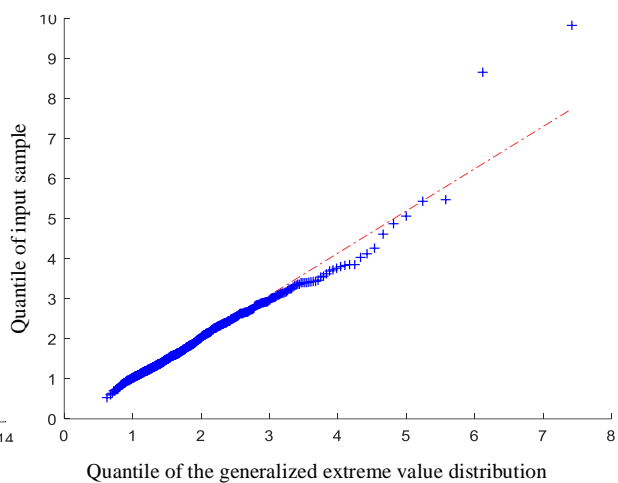


i) The Q-Q plot for the normal distribution.

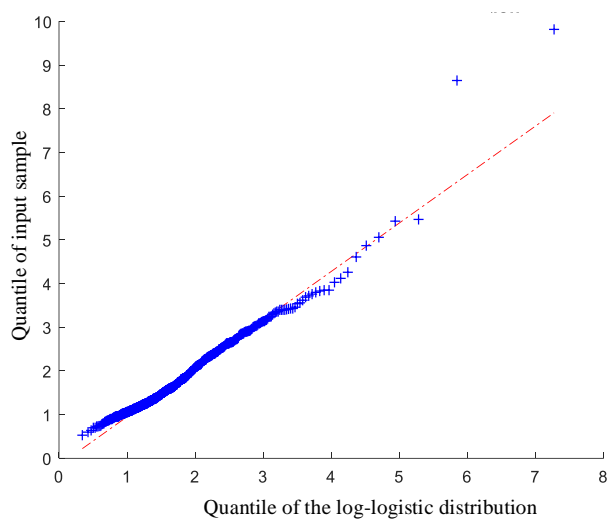
Figure 4.8: The Q-Q plots for the T2 dataset.



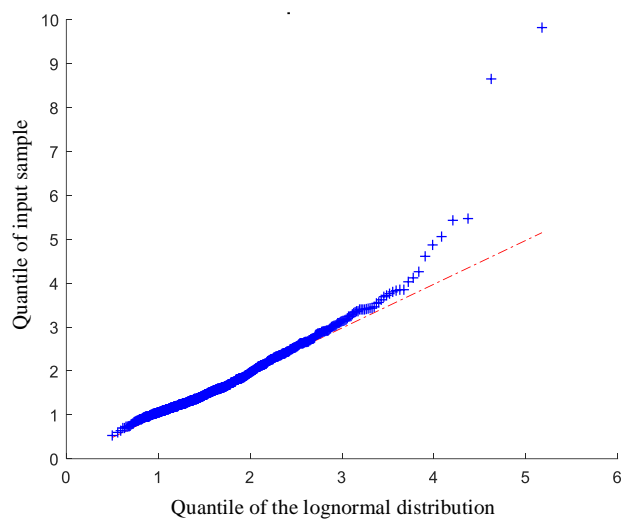
a) The Q-Q plot for the Burr distribution.



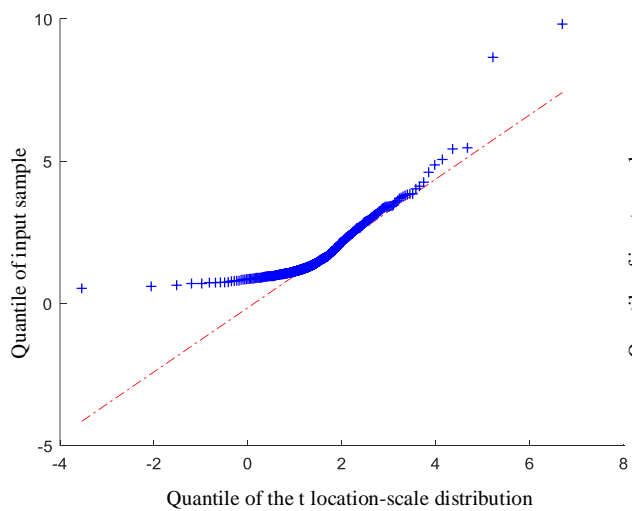
b) The Q-Q plot for the generalized extreme value distribution.



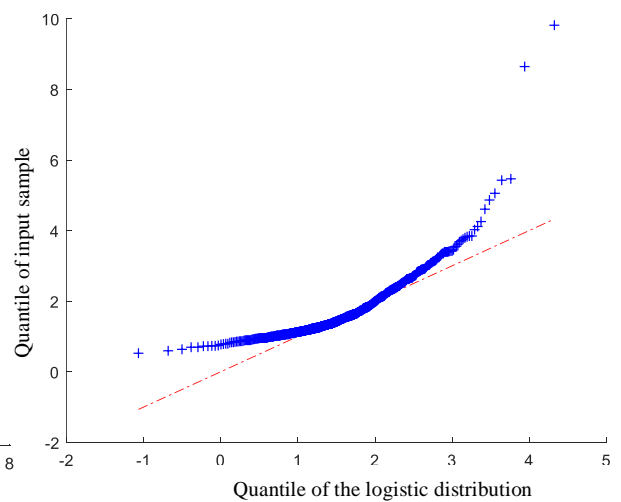
c) The Q-Q plot for the log-logistic distribution.



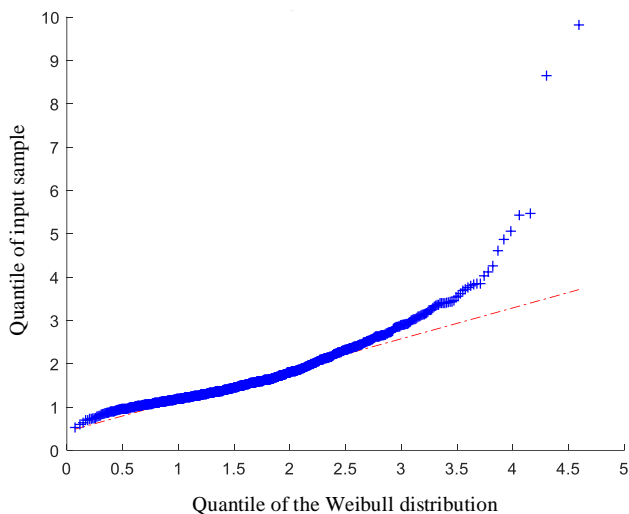
d) The Q-Q plot for the lognormal distribution.



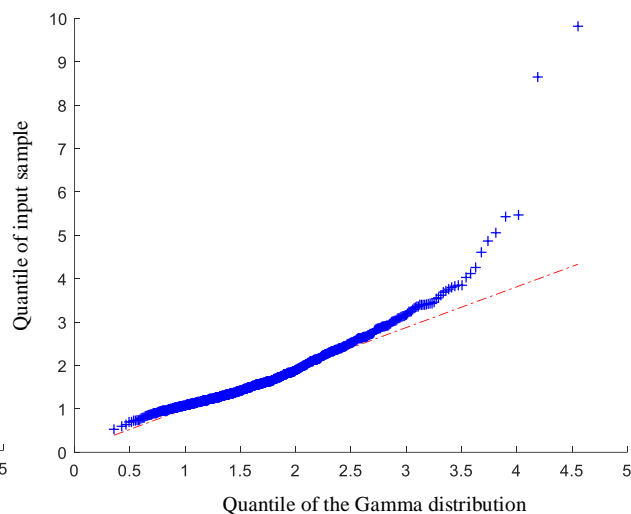
e) The Q-Q plot for the t location-scale distribution.



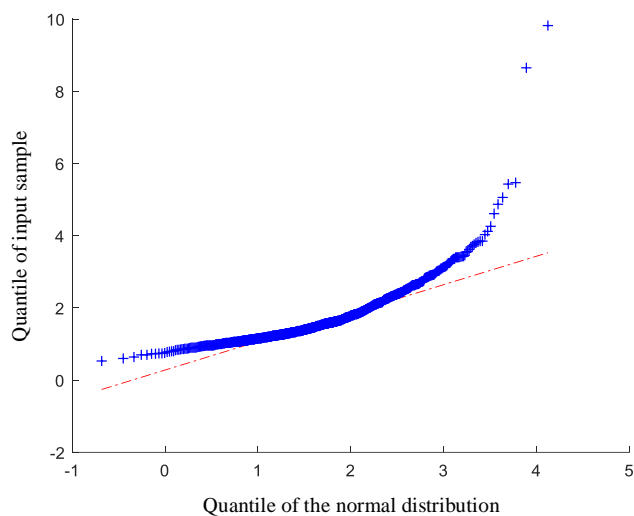
f) The Q-Q plot for the logistic distribution



g) The Q-Q plot for the Weibull distribution.



h) The Q-Q plot for the Gamma distribution.



i) The Q-Q plot for the normal distribution.

Figure 4.9: The Q-Q plots for the T3 dataset.

The C-S and K-S tests are used to determine the goodness of fit of each distribution. The results of the goodness of fit tests are presented in Table 4.4. The Burr distribution passed both goodness of fit tests at 5% significance for both the T1 and T2 datasets with higher traffic flow rates. The generalized extreme value distribution was accepted by both tests for the T2 and T3 datasets. The lognormal and log-logistic distributions passed just the K-S test for the T2 dataset. All the other distributions failed the goodness of fit tests for all

three datasets. Thus, the lognormal, t location-scale, logistic, Weibull, Gamma, and normal distributions are not suitable to model headway data.

Comparing the  $p$  values of both tests indicates that the generalized extreme value distribution has higher  $p$  values than the Burr distribution. Thus, the generalized extreme value distribution provides the best fit for the T2 and T3 datasets with lower traffic flow rates (less than 2100 vphpl). The Burr distribution is the second best distribution for the T2 dataset. In addition, the Burr distribution provides the best fit for the T1 dataset with higher traffic flow rates (more than 2100 vphpl). The C-S and K-S tests support the Q-Q test result that the Burr distribution is the best for the T1 dataset and the generalized extreme value distribution is the best for the T2 and T3 datasets.

Table 4.4: Goodness of fit test results for the new headway data

| <b>Probability distribution</b>  | <b>Dataset</b> | <b>Chi-Squared test</b> | <b><math>p</math></b> | <b>Kolmogorov-Smirnov</b> | <b><math>p</math></b> |
|----------------------------------|----------------|-------------------------|-----------------------|---------------------------|-----------------------|
| <b>Burr</b>                      | T1             | accepted                | 0.224                 | accepted                  | 0.584                 |
|                                  | T2             | accepted                | 0.150                 | accepted                  | 0.821                 |
|                                  | T3             | rejected                | 2.451e-04             | accepted                  | 0.099                 |
| <b>Generalized extreme value</b> | T1             | rejected                | 2.008e-06             | rejected                  | 0.012                 |
|                                  | T2             | accepted                | 0.185                 | accepted                  | 0.882                 |
|                                  | T3             | accepted                | 0.056                 | accepted                  | 0.350                 |
| <b>Log-logistic</b>              | T1             | rejected                | 3.030e-04             | rejected                  | 0.012                 |
|                                  | T2             | rejected                | 0.011                 | accepted                  | 0.281                 |
|                                  | T3             | rejected                | 1.200e-04             | rejected                  | 0.029                 |
| <b>Lognormal</b>                 | T1             | rejected                | 1.753e-08             | rejected                  | 0.012                 |
|                                  | T2             | rejected                | 0.003                 | accepted                  | 0.051                 |
|                                  | T3             | rejected                | 3.000e-04             | rejected                  | 2.975e-04             |
| <b>t location-scale</b>          | T1             | rejected                | 1.343e-06             | rejected                  | 0.012                 |
|                                  | T2             | rejected                | 2.231e-08             | rejected                  | 1.053e-05             |
|                                  | T3             | rejected                | 2.781e-04             | rejected                  | 3.011e-08             |
| <b>Logistic</b>                  | T1             | rejected                | 1.693e-34             | rejected                  | 6.787e-07             |
|                                  | T2             | rejected                | 2.310e-26             | rejected                  | 5.707e-08             |
|                                  | T3             | rejected                | 2.718e-21             | rejected                  | 2.358e-08             |

|                |    |          |           |          |           |
|----------------|----|----------|-----------|----------|-----------|
| <b>Weibull</b> | T1 | rejected | 1.573e-42 | rejected | 1.371e-21 |
|                | T2 | rejected | 6.365e-06 | rejected | 1.393e-08 |
|                | T3 | rejected | 3.350e-12 | rejected | 1.889e-18 |
| <b>Gamma</b>   | T1 | rejected | 4.131e-15 | rejected | 1.896e-07 |
|                | T2 | rejected | 1.603e-06 | rejected | 1.958e-05 |
|                | T3 | rejected | 8.681e-08 | rejected | 0         |
| <b>Normal</b>  | T1 | rejected | 1.283e-34 | rejected | 2.208e-20 |
|                | T2 | rejected | 1.736e-13 | rejected | 2.532e-15 |
|                | T3 | rejected | 1.516e-13 | rejected | 1.277e-16 |

## Chapter 5

### Conclusion

To understand the behaviour of drivers, the impact of weather on headway was examined in this thesis. Data obtained from the Emam Ali highway in Tehran was used to determine the statistical characteristics of headway and several distributions were considered. The headway data, traffic flow rates, and average speeds under rainy weather conditions differ from those in clear weather. The headway in clear weather with traffic flows greater than 1900 vphpl is the smallest and is the largest in rainy weather with less than 1000 vphpl traffic flow. This is consistent with previous results that rainy weather causes an increase in headway because of slippery roads, poorer braking efficiency, and lower visibility. The influence of rainy weather on the headway and average speeds suggests that drivers modify their behaviour under rainy conditions.

Finding appropriate headway distributions is essential for traffic simulation. In this thesis, various statistical distributions were examined and their fit to the headway data was determined using three goodness of fit tests, Q-Q plot, K-S, and C-S. While some distributions passed one or two tests, some were rejected by all three tests. In clear weather, the Burr distribution was found to provide the best fit for high traffic flows, i.e. for rainy weather with a high traffic flow. This confirms that the Burr distribution is suitable to model headway in clear and rainy weather conditions with a high traffic flow [7]. The generalized extreme value distribution was found to provide the best fit for headway in rainy weather or clear weather with a low traffic flow.

The headway distributions of the eight datasets show that the distribution patterns in high traffic flows (greater than 1300 vphpl) and high average speeds (between 80-90 km/h) are very similar. However, with low traffic flows and low average speeds, there is a similar pattern.

Comparing the results for the new headway data with the previous results confirms that the Burr distribution is the best for higher traffic rates. This means that the Burr distribution fits large headways better than small headways while the generalized extreme value distribution provides the best fit for lower traffic flow rates. These results show the relation between headway distributions and traffic conditions. They will be useful in selecting suitable headway distributions for traffic management and control under different weather conditions. The results of this thesis can be applied to other traffic applications such as estimating capacity and the level of service.

## **5.1 Future Work**

Future work can consider further vehicle headway analysis. The data obtained corresponds to medium traffic flow rates only. Thus, it would be interesting to collect data and find suitable distributions for rush hour and other situations to generalize the results to all traffic conditions. Further, the headway can be examined in night conditions.

## Bibliography

- [1] C. N. Bowman, J. A. Miller, Modeling traffic flow using simulation and big data analytics, Winter Simulation Conference, pp. 1206-1217, 2016.
- [2] Transportation Research Board (TRB), Highway Capacity Manual, National Research Council, Washington, DC, pp. 4-7, 2000.
- [3] S. Moridpour, Evaluating the time headway distributions in congested highways, Journal of Traffic and Logistics Engineering, vol. 2, no. 3, pp. 224-229, 2014.
- [4] A. F. Clitan, M. L. Dragomir, C. Madalina, I. M. Beca, G. Hoda, Improving traffic conditions on a set three intersections using microscopic simulation models, Procedia Engineering, vol. 181, pp. 139-145, 2017.
- [5] P. Yi, Y. Zhang, J. Lu, H. Lu, Safety-based capacity analysis for Chinese highways- A preliminary study, International Association of Traffic and Safety Sciences Research, vol. 28, no. 1, pp. 47-55, 2004.
- [6] S. Das, A. K. Maurya, Time headway analysis for four-lane and two-lane roads, Transportation in Developing Economies, vol. 3, no. 1, pp. 98-115, 2017.
- [7] H. M. Alhassan, J. Ben-Edigbe, Effect of rain on probability distributions fitted to vehicle time headways, Advanced Science Engineering Information Technology, vol. 2, no. 2, pp. 144-150, 2012.
- [8] S. Daisuke, O. Izumi, N. Fumihiko, On estimation of vehicular time headway distribution parameters, Traffic Engineering, vol. 34, no. 6, pp. 18-27, 1999.
- [9] T. L. Jackson, H. O. Sharif, Rainfall impacts on traffic safety: Rain-related fatal crashes in Texas, Geomatics, Natural Hazards and Risk, vol. 7, no. 2, pp. 843-860, 2016.

- [10] H. M. Alhassan, J. Ben-Edigbe, Effect of rainfall on microscopic traffic flow parameters, Malaysian Universities Transportation Research Forum and Conference, Kuala Lumpur, Malaysia, pp. 117-126, 2010.
- [11] S. A. Hassan, T.H. Xuan, N. Mashros, Characteristics of vehicle headway during rainfall at night in Pontian, Johor, Malaysian Journal of Civil Engineering, vol. 29, no. 1, pp. 59-74, 2017.
- [12] A. S. Ghamdi, Analysis of time headways on urban roads: A case study from Riyadh, Journal of Transportation Engineering, vol. 127, pp. 289-294, 2001.
- [13] S. Yin, Z. Li, Y. Zhang, D. Yao, Y. Su, L. Li, Headway distribution modeling with regard to traffic status, IEEE Intelligent Vehicles Symposium, Xi'an, China, pp. 1057-1062, 2009.
- [14] W. D. Ashton, Distributions for gaps in road traffic, IMA Journal of Applied Mathematics, vol. 7, no. 1, pp. 37-46, 1971.
- [15] G. H. Bham, S. R. P. Ancha, Statistical models for preferred time headway and time headway of drivers in steady state car-following, International Conference on Applications of Advanced Technology in Transportation, Chicago, IL, pp. 344-349, 2006.
- [16] J. D. Griffiths, J. Hunt, Vehicle headways in urban areas, Traffic Engineering and Control, vol. 32, No. 10, pp. 458-462, 1991.
- [17] G. Zhang, Y. Wang, A Gaussian kernel-based approach for modeling vehicle headway distributions, Transportation Science, vol. 48, no. 2, pp. 206–216, 2014.
- [18] G. Zhang, Y. Wang, H. Wei, Y. Chen, Examining headway distribution models with urban freeway loop event data, Transportation Research Record, vol. 1999, pp. 141-149, 2007.
- [19] Y. Wang, L. Jing, Study of rainfall impacts on freeway traffic flow characteristics, World Conference on Transport Research, Shanghai, China, pp. 1533-1543, 2016.

- [20] Tehran, Iran, Google Maps, Google, 15 March 2018.
- [21] R. Walpole, R. Myers, S. Myers, K. Ye, Probability and Statistics for Engineers and Scientists, 7th edition, Pearson, Upper Saddle River, NJ, pp. 237, 2002.
- [22] S. Mitra, K. Utsav, Car following under reduced visibility, International Conference on Road Safety and Simulation, Indianapolis, IN, pp. 65-76, 2011.
- [23] S. Hoogendoorn, P. Bovy, A new estimation technique for vehicle-type specific headway distributions, Journal of the Transportation Research Board, vol. 1646, National Research Council, Washington, DC, pp. 18-28, 1998.
- [24] S. Dong, H. Wang, D. Hurwitz, G. Zhang, and J. Shi, Nonparametric modeling of vehicle-type-specific headway distribution in freeway work zones, Journal of Transportation Engineering, vol. 141, no. 11, pp. 39-52, 2015.
- [25] S. Yin, Z. Li, Y. Zhang, D. Yao, Y. Su, L. Li, Headway distribution modeling with regard to traffic status, IEEE Intelligent Vehicles Symposium, Xi'an, China, pp. 1057-1062. 2009.
- [26] C. Forbes, M. Evans, N. Hastings, B. Peacock, Statistical Distributions, 4th edition, Wiley, pp. 131-134, 2011.
- [27] M. H. Tahir, M. Mansoor, M. Zubair, G. G. Hamedani, McDonald log-logistic distribution with an application to breast cancer data, Journal of Statistical Theory and Applications, vol. 13, no. 1, pp. 65-82, 2014.
- [28] R. T. Luttinen, Statistical Analysis of vehicle time headways, Ph.D. dissertation, Helsinki University of Technology, Lahti, Finland, pp. 33-34, 1996.
- [29] I. J. Myung, Tutorial on maximum likelihood estimation, Journal of Mathematical Psychology, vol. 47, no. 1, pp. 90-100, 2003.
- [30] R. J. Cowan, Useful headway models, Transportation Research, vol. 9, no. 6, pp. 371-375, 1975.

- [31] A. Loy, L. Follett, H. Hofmann, Variations of Q–Q plots: The power of our eyes, *The American Statistician*, vol. 70, no. 2, pp. 202-214, 2014.
- [32] F.J. Massey, Jr., The Kolmogorov-Smirnov test for goodness of fit, *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68-78, 2010.
- [33] W. J. Conover, *Practical Nonparametric Statistics*, 3rd edition, Wiley, New York, NY, pp. 428-433, 1999.
- [34] C. Harrell, B. K. Ghosh, R. Bowden, *Simulation Using ProModel*, 2nd edition, McGraw-Hill, New York, NY, pp. 72, 2004.
- [35] A. DasGupta, *Asymptotic Theory of Statistics and Probability*, Springer, Berlin, p. 445, 2008.
- [36] J. Kloeke, J. W. McKean, *Nonparametric Statistical Methods Using R*, Chapman and Hall/CRC, Madison, WI, pp. 34-38, 2016.
- [37] Tehran, Iran, Google Maps, Google, 18 August 2018.
- [38] Tehran, Iran, Google Maps, Google, 10 October 2018.

## Appendix A

Table A.1. Critical values for the C-S test statistic

| Degrees of freedom | 0.1     | 0.05    | 0.01    |
|--------------------|---------|---------|---------|
| 1                  | 2.706   | 3.841   | 6.635   |
| 2                  | 4.605   | 5.991   | 9.210   |
| 3                  | 6.251   | 7.815   | 11.345  |
| 4                  | 7.779   | 9.488   | 13.277  |
| 5                  | 9.236   | 11.070  | 15.086  |
| 6                  | 10.645  | 12.592  | 16.812  |
| 7                  | 12.017  | 14.067  | 18.475  |
| 8                  | 13.362  | 15.507  | 20.090  |
| 9                  | 14.684  | 16.919  | 21.666  |
| 10                 | 15.987  | 18.307  | 23.209  |
| 20                 | 28.412  | 31.410  | 37.566  |
| 30                 | 40.256  | 43.773  | 50.892  |
| 40                 | 51.805  | 55.759  | 63.691  |
| 50                 | 63.167  | 67.505  | 76.154  |
| 60                 | 74.397  | 79.082  | 88.380  |
| 70                 | 85.527  | 90.531  | 100.425 |
| 80                 | 96.578  | 101.880 | 112.329 |
| 90                 | 107.565 | 113.145 | 124.117 |
| 100                | 118.498 | 124.342 | 135.807 |