TECHNISCHE UNIVERSITEIT DELFT

CIVIL ENGINEERING

BACHELOR THESIS

Analysis of bicycle flow models at signalized intersections

Author:
M. Poppe, 4242025

Supervisors:
Ir. R. Koster
Dr. Ir. Y. Yuan

June 17, 2018

Abstract

It is of great importance to gain insights on bicycle traffic flow. Especially at signalized intersections as these are the main cause of delays. This report extends the research on bicycle traffic flow characteristics [Goñi-Ros et al., 2018] with a focus on headways from which the discharge flow and saturation flow can be derived. Cyclist data from a signalized intersection in Amsterdam for 57 periods is considered. To create more realistic headways, a model that divides the cyclist path into sublanes is used to calculate the headways [Botma and Papendrecht, 1991], because cyclists are able to form multiple queues. It is found that the widths of the sublanes influence the predicted headways. With maximum likelihood and the χ^2 goodness-of-fit test, it is found that the half normal distribution gives the best fit if the bicycle path is considered as one lane. Introduction of the sublanes causes the lognormal distribution to give the best fit. Finally, with the calculated headways per period a model for the saturation flow is derived [Raksuntorn and Khan, 2003], which states that the saturation flow is derived from the constant headways. Without the method of the sublanes, the headways are constant among the different cyclists in a period, but after applying the aforementioned method, it turns out that the headways get constant after the 8^{th} cyclist. Further research could focus on more profound research of the optimal sublane width, a renewed calculation of the saturation flow for sublanes and applying the same methods on other data sets as this will most likely lead to different results.

Contents

1 Introduction

Crossing traffic flows are controlled with signalized intersections, causing delays to its users who have to stop in case of a red light. Then all traffic comes to a halt and needs to accelerate when the traffic light turns green. Much research has been carried out about traffic flow at signalized intersection, trying to increase understanding in traffic flow characteristics. Various modes of transport can cross at a signalized intersection, such as motor vehicles, trams, cyclists and pedestrians. Behaviour of motor vehicles and pedestrians at signalized intersections has been under more attention than cyclists [Wang et al., 2011], the latter being the subject of this report. Cyclist behaviour at signalized intersections is different than behaviour of motor vehicles as cyclists may have the possibility to form multiple queues [Raksuntorn and Khan, 2003] or may have a tendency to wait at the right of the bicycle path where they can put their right foot on the pavement.

To generate insights in bicycle flow characteristics at signalized intersections, the macroscopic variables of traffic flow theory are used, namely density, discharge flow and speed [Lebacque and Khoshyaran, 2002]. This research focuses mainly on discharge flow, which is the number of cyclists passing a cross-section of a bicycle path in a unit of time, usually measured in hours. Discharge flow is often defined as the reciprocal of average headway, the time between two cyclists passing a cross-section of a bicycle path. Many headway distribution models have been developed in order to be able to simulate and forecast the traffic flows, such as [Adams, 1936], [Tolle, 1971], [Cowan, 1975] and [Zhang et al., 2007]. They are all tested with motor vehicles, but as stated earlier, cyclists may behave differently and therefore it is interesting to investigate and compare these models with cyclist data.

The average headway is directly related to the discharge flow, but there are other ways of calculating this macroscopic variable. A popular one is Edie's definition for discharge flow [Edie, 1961]. Moreover, the saturation flow can be of importance for the design of a signalized intersection. Saturation flow is defined as the number of cyclists in a dense flow of traffic for a specific intersection in case of a ongoing green phase per unit of time [Chand et al., 2017] and is therefore immediately related to the capacity of the bicycle lane. The method of [Raksuntorn and Khan, 2003] is used for calculation of the saturation flow. An extension is the introduction of sublanes [Botma and Papendrecht, 1991] as cyclist headways are not always influenced by a cyclist in front but at the opposite side of the cyclist path. The headway models and flow models are compared and evaluated with empirical data from a signalized intersection in Amsterdam for the method without sublanes and the method with sublanes.

The aim of this report is to provide useful insights in bicycle traffic flow. First, a comparison will be made between different headway distribution models with cyclist data as the outcome may be different than when using motor vehicle data. Another aim is to look into different flow methods, namely discharge flow calculated with headways, discharge flow according to Edie's definition and saturation flow. The sublane method is applied to the headway models and the saturation flow.
Therefore the following main question with 3 sub questions will be answered:
What useful bicycle headway models and bicycle flow characteristics at a signalized intersection can be derived from data containing cyclist trajectories at a signalized intersection in Amsterdam on June 6, 2016?
1. Which of the proposed headway distribution models fits the data best in terms of maximum likelihood and goodness-of-fit?
2. What are the bicycle flows at the signalized intersection in Amsterdam?
3. What happens to the headway and the saturation flow if sublanes are introduced?

The outline of the report is designed as follows. Chapter 2 gives insight into the different flow models and its associated headway distribution models. Chapter 3 deals with the analyses of the models by explaining the data and showing how the different results can be compared. Then, the results of the analyses are given in chapter 4. Finally, the results are discussed in chapter 5 and chapter 6 concludes.

2 Models

This framework of this paper is traffic flow theory, which will be explained briefly in section 2.1. The relation between discharge flow and headway will be clarified, before moving on to various headway distribution models in section 2.1.1. Next, in section 2.2 Edie's definition for discharge flow [Goñi-Ros et al., 2018] will be described. Moving on to saturation flow in section 2.3, its definition and a method to calculate the saturation headway are given. Finally, section 2.4 gives a summary of the whole chapter.

2.1 Traffic flow theory

Traffic flow theory has various components, but three macroscopic variables are widely used [Hoogendoorn and Knoop, 2013], namely density k, average speed v and flow q. Density k is defined as the number of cyclists N per distance unit X, which is usually measured in meters. The average speed v is in this case the mean speed of the cyclists when crossing the stop line of the signalized intersection. This report focuses mainly on the third macroscopic variable: flow q.

The discharge flow q_d is described as the number of cyclists N that pass the signalized intersection during a unit of time T, which is usually given in hours for traffic flow and translates to 3600 seconds [Hoogendoorn and Knoop, 2013]. This gives us equation (1).

$$q_d = \frac{3600 \times N}{T} = \frac{3600 \times N}{\sum_{j=1}^{N} h_j} = \frac{3600}{\bar{h}} \tag{1}$$

in which h is the average headway, defined in equation (2) as the time in seconds between two successive cyclists when they cross a predefined line:

$$h_j = t_j - t_{j-1} \tag{2}$$

in which index j ranges from 2 to the total number of cyclists N.

2.1.1 Headway distribution models

Much research has been carried out to come up with a headway distribution model $p(h)$ in order to be able to simulate the discharge flow [Ye and Zhang, 2009]. Such a model is basically a probability density function (PDF), from which random samples can be generated. A PDF describes how likely it is that a value is retrieved according to its parameter(s). Mathematically, this is written as

$$P(h = x) \quad \forall x \in \mathbb{R} \tag{3}$$

In words, equation (3) describes the probability that the headway h is equal to x, which can be calculated for all x that are a real number. Doing so results in a graph and the shape of this graph is different for every headway distribution model.

Table 1 shows different proposed distribution models with their parameters, PDF and mean. They will be compared in terms of statistical measures. Most research with these models has been performed for motor vehicle data and it is therefore interesting to find out whether the same models are applicable for bicycle traffic. It is shown that cyclists quickly accelerate towards the stop line and often pass this line with a constant velocity [Goñi-Ros et al., 2018], implying that the acceleration is equal to zero.

Table 1: Different headway distribution models

#	Distribution	Parameters	PDF	Mean
1	Normal	μ, σ	$\frac{1}{\sqrt{2\pi}\sigma} exp(\frac{-(h-\mu)^2}{2\sigma^2})$	μ
2	Lognormal	μ, σ	$\frac{1}{h\sigma\sqrt{2\pi}} exp(-\frac{1}{2}(\frac{ln(h)-\mu}{\sigma})^2)$	$exp(\mu + \sigma^2/2)$
3	Exponential	λ	$\lambda exp(-\lambda h)$	$1/\lambda$
4	Gamma	k, θ	$\frac{1}{\theta^k \Gamma(k)} h^{k-1} exp(-\frac{h}{\theta})$	$k\theta$
5	Half Normal	μ, σ	$\frac{1}{\sqrt{2\pi}\sigma} exp(\frac{-(h-\mu)^2}{2\sigma^2})$ with $h \geq 0$	$\sigma\sqrt{2}/\sqrt{\pi}$

The first proposed headway model [Adams, 1936] is a pioneer in headway distribution models and is equal to the normal (Gaussian) distribution. In the PDF, μ is defined as the expected headway $E[h]$ and σ^2 as the variance of the headway $Var[h]$. The normal distribution implies symmetry around the mean μ, an assumption which was later on dropped when more complex models were introduced.

A lognormal distribution was then introduced to model headway distributions [Tolle, 1971]. The characteristic of the lognormal distribution is that the logarithm of the headway is normally distributed. So if X is lognormally distributed, then $Y = ln(X)$ has a normal distribution. From table 1 it is concluded that this model is more advanced as the calculation of the mean is already a bit more difficult.

Another straightforward model, the exponential distribution [Cowan, 1975], is directly derived from the Poisson distribution. The Poisson distribution is widely used for calculating the occurrence of a given number of independent events, or arrivals of cyclists in this situation, in case of a constant arrival rate λ (in cyclists per unit of time). Assuming the Poisson distribution for the number of cyclists, then the inter-arrival time, or headway h in this situation, is given by an exponential distribution as in table 1.

The fourth distribution model to be considered is the gamma distribution [Zhang et al., 2007]. Although this distribution model is less intuitive than the former, it is frequently used to model waiting times as it is related to the exponential distribution. The gamma distribution has two parameters, namely k and . Shape parameter k defines the shape of the PDF, which can take on a variety of shapes. Scale parameter θ determines how spread out the distribution is. If θ is larger than 1, the distribution is called stretched and if θ is smaller than 1, then the distribution is compressed.

The half normal distribution is fairly unknown, but not uncommon to use. This distribution is similar to the PDF of the normal distribution, as shown in table 1. The only difference is the restriction that h cannot be negative. This results in a folded normal distribution. It may be an interesting distribution as the distance between cyclist can indeed not be negative.

2.2 Edie's definitions

Edie's definitions [Edie, 1961] about traffic flow theory are a common way to deal with discharge flows. This definition takes into account a specified line a, which is defined as a line after the official stop line. This line is used next to stop line as some cyclists may stop to form a queue after the official stop line. With this information, the definition for discharge flow $q_{d,E}$ is given by

$$q_{d,E} = \frac{\sum_{j=2}^{N} \chi_j}{\Delta x \times \Delta t} \tag{4}$$

in which χ_j is the distance travelled by cyclists j during period Δt. The time Δt is the time between the moment that the first cyclist, $j = 1$, crosses the line a and the moment that the last cyclist, $j = N$, crosses the line a. Δx is the distance between the stop line and line a. More insights in the data will

be given in section 3.1, in which figure 1 shows bicycle trajectories based on the data set to visualize how Edie's definition for discharge flow is calculated.

2.3 Saturation flow

Saturation flow q_s is measured as the number of cyclists in a dense flow of traffic for a specific intersection in case of a ongoing green phase per unit of time, which is usually measured in hours [Chand et al., 2017]. It can be defined in a similar way as discharge flow q_d in equation (1), namely as a function of saturation headway h_s, such that

$$q_s = \frac{3600}{h_s} \tag{5}$$

in which h_b is calculated as

$$h_s = \frac{\sum_{j=n}^{N} h_j}{N - n} \tag{6}$$

in which the saturation flow starts at cyclist n and N is the number of the last cyclist in the queue [Rahman et al., 2005]. Saturation headway is defined as the distance (in seconds) between two consecutive cyclists when their speed is constant and their acceleration therefore equal to zero. When cyclists have stopped at a signalized intersection, they have some start-up time in which they accelerate after the traffic light has turned green. The headway between cyclists is at first assumed not to be constant [Raksuntorn and Khan, 2003] but gets constant after cyclist n. The saturation headway will be empirically obtained with data described in section 3.1.

2.4 Summary of the proposed models

Chapter 2 states different models with a link to flow. Discharge flow is often defined as the reciprocal of average headway. There exist many headway distribution models and some of them are explained in this chapter. Then, saturation flow is calculated by focusing on when the headways in a period get constant.

3 Methodology

Having introduced all the different models and flows in chapter 2, this chapter explains the different methodology for the proposed models using empirical data. Section 3.1 explains the data set used for the application of the methods that are proposed in this chapter. Then, section 3.2 shows how to compare the five different proposed headway distribution models and choose the relatively best one. Moving on to flow models, it is explained how to compare discharge flow calculated with headways and discharge flow according to Edie's definition in section 3.3. Furthermore, a method to calculate the saturation flow is given in section 3.4. Then, section 3.5 explains the method of the sublanes. Finally, section 3.6 gives a summary of the whole chapter.

3.1 Data Set

Data in the form of images has been gathered from signalized intersection Weteringlaan in Amsterdam as explained in [Goñi-Ros et al., 2018]. After cleaning the data, converting it and adding some criteria, 57 different periods are left. Examples of the criteria are the absence of pedestrians and the fact that all cyclists have to cross the stop line before the light turns red again. All these 57 periods are unique measurements in time on June 6, 2016 when cyclists were waiting for the light to turn green at this specific signalized intersection. The data consists of (x,y)-coordinates for each cyclist for various time stamps, so the movements of all cyclists are known from the moment they arrive in the queue until the moment they cross the stop line. From the data several macroscopic variables can be derived, such as density and average speed and most importantly for our research: headway.

In order to extract useful variables, one needs to know the exact situation at signalized intersection. When the signal is red, cyclists dismount and wait until the signal turns green. The red phase is a queueing process, in which it is assumed that the number of cyclists waiting is monotonically increasing, so no cyclists leave the area before the light turns green. All the cyclists are standing in an area $A = L \times W$ with length L and width W. L is not fixed and depends on the number of cyclists, whereas the width of the bicycle path W is fixed, namely 2 meters. It is possible though that cyclists wait outside of this area. During the green phase, all cyclists cross the intersection. From this process, the necessary variables can be extracted. To provide a better understanding of the data, figure 1 is provided. This figure shows the position of all cyclists in a single period at all moments in time. Cyclists farther from the stop line start moving at a later moment than cyclists closer to the stop line. Moreover, line a is indicated after the stop line. Especially the passing times, the times of crossing the stop line, are important, as the headways can be derived from that data.

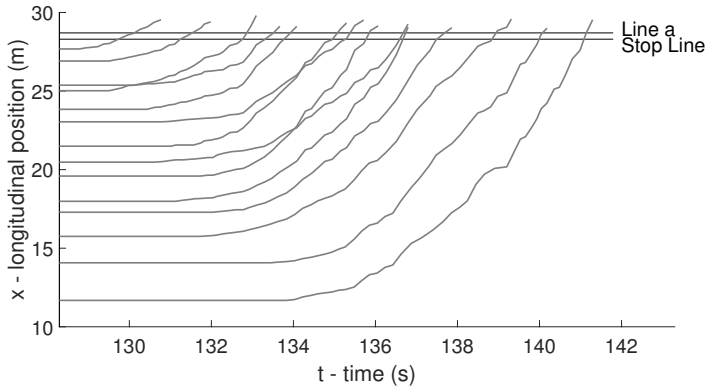


Figure 1: Cyclist trajectories for a single period with 14 cyclists for a signalized intersection in Amsterdam, retrieved from [Goñi-Ros et al., 2018].

3.2 Comparison of different headway distribution models

Analyses are carried out with empirically obtained data discussed in section 3.1. As noted, the headways h can be obtained with the data for the different periods as the passing times are known. A comparison between the different headway distribution models is first made with the maximum likelihood method and then with the χ^2-test, also known as the goodness-of-fit test. The distribution that describes the data best is chosen by taking into account the outcomes of both methods.

From the likelihood function, the parameters can be fitted according to the data. Maximum likelihood states how likely it is that the fitted parameters are correct. Different likelihood values can be compared to get an idea which distribution suits the data best. The results are also visualised in order to obtain insights from a different angle by fitting the proposed models in a histogram constructed with all the headways. A histogram is basically a plot with shows the frequencies of the each value and can be used to derive the shape or distribution of the data.

Furthermore, the goodness-of-fit test is carried out to compare the different headway distribution models. After fitting the parameters for each distribution with maximum likelihood, a random sample equal to the size of the data set is drawn. If the drawn headways from a distribution and the observed headway are ranked, the goodness-of-fit test can be performed:

$$\chi^2_{\alpha,d} = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{O_i} \tag{7}$$

in which O_i and E_i are the observed and the expected or drawn headway for observation i, respectively. The critical values of the test depend on the chosen significance level α and the degrees of freedom d equal to the number of parameters in the distribution - 1. The goodness-of-fit test is used

to find out whether the null hypothesis is correct, namely that the right distribution for the headway was assumed. This test is performed for all of the proposed distributions. If the null hypothesis is rejected, it is concluded that the data does not follow such a distribution. Different p-values are compared next to the likelihoods to conclude which headway distribution model fits the data best. A p-value indicates the probability that the null hypothesis is correct. A small p-value gives the indication of strong evidence against the null hypothesis, so it is rejected. As critical p-value one usually takes 0.05.

3.3 Comparison of different flow definitions

In chapter 2 three different flow definitions were defined, namely discharge flow q_d, discharge flow according to Edie's definition $q_{d,E}$ and saturation flow q_s. The first two definitions differ per period, whereas the latter definition is an asymptotic result and takes into account every period. It is therefore a single number. First, the discharge flow and the flow according to Edie's definition will be visualised by means of a figure. The linear relationship between the flow variables can be identified in a statistical way by means of correlation. The sample correlation coefficient between two variables X and Y is calculated as

$$r_{X,Y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{8}$$

In this case, X and Y are replaced by q_d and $q_{d,E}$, respectively. A correlation coefficient varies between -1 and +1, indicating a negative and a positive relationship, respectively. From a correlation coefficient with the value 0 it can be concluded that there exists no relationship between the two variables.

3.4 Calculation of saturation flow

The method of [Raksuntorn and Khan, 2003] is followed to obtain the saturation headway when taking into account all the different headways per period. First, it is analysed for which cyclist in the queue the headways remain constant. The hypothesis here is that headways in the front of the queue are larger and then slowly decrease until they get constant. Having derived the cyclist number of the saturation headway, the saturation headway is determined as in equation (6) and therefore also the saturation flow according to equation (5). Equation (6) might need some adjustment as there are very few data points for large periods. Then the saturation headway is not anymore calculated up to the last cyclist, but up to an arbitrary cyclist number.

3.5 Sublanes

As cyclists are able to form multiple queues if the width of the bicycle path is large enough, it is desirable to take this into account when calculating the headways. If two cyclists are pairing, so cycling next to each other, it may happen that they cross the stop line at almost exactly the same time. For the saturation flow, that calculated headway does not add any value as the paired cyclists are not following each other. To overcome this issue, a method of dividing the cyclist path into sublanes is proposed [Botma and Papendrecht, 1991]. From now on, to this method will be referred as 'sublane method', whereas the former method is referred to as the 'one lane method'. The sublane method requires an adjusted definition of headways as bicycle traffic does not follow fixed sublanes and is therefore different from motorized vehicles. Thus, the method is as follows. The bicycle path width is divided in a number of sublanes with fixed width. Every cyclist is positioned in one of the sublanes. Then, the headway is calculated as the cyclist closest ahead who is positioned in one of the 3 adjacent lanes on the left as well as the right of the cyclist. This condition is stated equation (9):

$$|pos_i - pos_j| < 3 \times w_{sub} \quad \forall i,j \quad if \quad i < j \tag{9}$$

for which i and j are the indices of the cyclists in the order in which they pass the stop line. pos_i and pos_j are the positions of the cyclists perpendicular to the direction of the flow. Moreover, w_{sub} is the width of the sublane. This parameter is varied to obtain different results before choosing an optimal sublane width. The restriction $i < j$ is added because the headway of cyclist j solely depends on the cyclists in front of him. For example, for calculating the headway of cyclist 4, only cyclists 1, 2 and 3 are considered.

The method is clarified with an example in figure 2. All cyclists in period 55 are indicated with a red cross. First of all, it is noticed that cyclist C is officially out of the range of the bicycle path width, so it is needed to take into account that it is possible that cyclists are positioned outside of the official path. When taking a look at cyclist A, according to the one lane method, his headway is defined as the difference in time with the closest cyclist in front of him. For cyclist A, this is cyclist B. This immediately shows the added value of the sublane method. As cyclist B is more than 3 sublanes away, now cyclist C is obtained as the cyclist in front of cyclist A. So a different headway is calculated. A second example from figure 2 shows that the headway for cyclist C is defined as the difference in time with cyclist E and not cyclist D as in the one lane method.

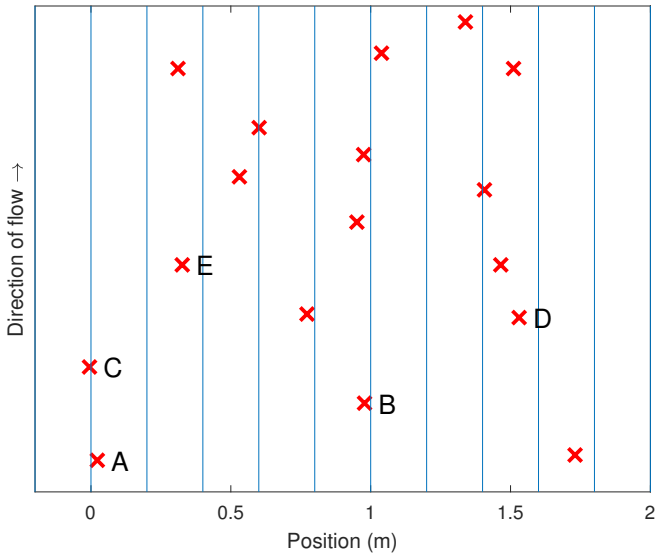


Figure 2: An example for the sublane method for period 55 and with sublane width $w_{sub} = 0.20$ m. The red crosses are the positions of the cyclists when passing the stop line. Five cyclists have been given a name, namely A, B, C, D and E. The blue lines show how the sublanes are divided. This period has in total 11 sublanes. The cycling direction is vertically upwards.

As w_{sub} is varied, the total sublane, which is equal to 3 times the sublane, also varies. One needs to keep in mind that if the total sublane gets wider, the capacity per sublane increases. The bicycle path width is fixed, so the total capacity does not necessarily increase. Thus, in order to create a useful comparison, the saturation flows needs to be adjusted according to the width of the sublanes as shown in equation (10) to obtain the total saturation flow $q_{s,total}$:

$$q_{s,total} = q_s \times L_{total} \tag{10}$$

in which the total number of lanes L_{total} is calculated as

$$L_{total} = \frac{W}{3 \times w_{sub}} \tag{11}$$

with fixed bicycle path with $W = 2$ meter.

3.6 Summary of the proposed methodology

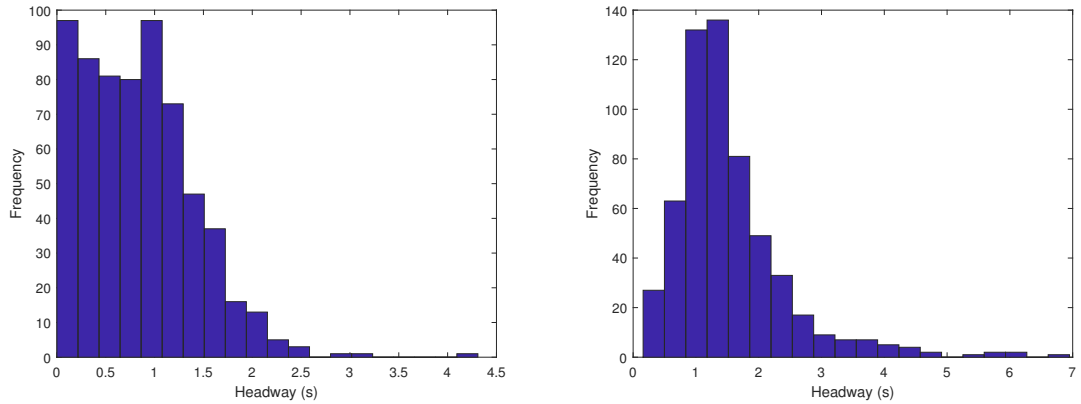
Chapter 3 gives insight into the data set, from which the cyclist behaviour in terms of movement for each cyclist in the 57 periods can be derived. The five proposed headway distribution models will be compared with maximum likelihood and the χ^2 goodness-of-fit test. Thereafter, the two methods for discharge flow will be compared by calculating their correlation coefficient. At last, saturation flow is derived from a figure which shows the headways for the cyclists in each period. The research on headway distribution models and saturation flow is extended by dividing the bicycle path into sublanes.

4 Results

Chapter 4 is organised in the following way. Section 4.1 deals with the different headway distribution models for the one land method as well as the sublane method after first having taken a look at its summary statistics. Having picked the relatively best distribution for the headway data, the headways are used for the flow models. First for a comparison between the two different discharge flows in section 4.2, with one flow based on headway and the other on Edie's definitions. Thereafter, the saturation flow is calculated for two different situations in section 4.3. Finally, section 4.4 gives a summary of the whole chapter.

4.1 Headway distribution models

First of all, the empirical headway data is analysed before using it to derive the fitted parameters of the different models. The headways are calculated by subtracting the different passing times as in equation (2) for each period. To provide useful insight in the summary statistics, the histograms of the headways for the one lane method and the sublane method are given in figure 3 and the summary statistics in table 2.



(a) One lane method

(b) Sublane method

Figure 3: The empirically obtained bicycle headways for the signalized intersection in Amsterdam (June 6, 2016).

A few interesting things can be concluded from figure 3. First, the one lane method in figure 3a is taken a look at. There is one outlier with a value which is higher than 4. This means that there was a gap in one period between two cyclists larger than 4 seconds with no cyclists in between. There can be various reasons for this, such as not paying attention to the green light or simply an error in the data. This outlier causes that most observations are on the left part of the figure. Nonetheless, it can be concluded that the headway distribution is left-tailed. This means that most of the observations are in the left part of the data set where the highest frequencies are seen. The frequencies are at first very similar, but decrease after approximately 1.3 seconds. What might seem curious, is

that there are some observations very close to zero. The underlying reason is that cyclists can form multiple queues and are therefore able to pass the stop line at almost exactly the same moments due to the fact that they are overtaking each other or are cycling in pairs. This problem is solved with the introduction of sublanes, as can be seen in figure 3b. The headways are now differently distributed, with a big peak close to a headway of 1 second. For this set of data, it is also observed that the headway distribution is left-tailed. There are a few outliers in this case, as can be concluded from the right part of figure 3b, where a few headways larger than 6 seconds are found.

Table 2: Summary statistics for headways retrieved from the data set for the signalized intersection in Amsterdam (June 6, 2016). This table reports shape and spread of the data.

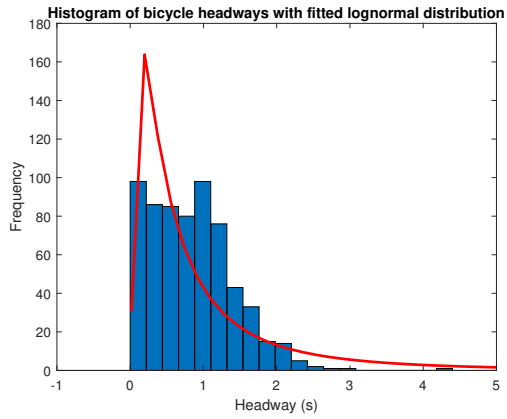
Statistics	One lane	Sublane
Observations	638	578
Minimum	0.0003	0.1590
Maximum	4.3099	6.9490
Mean	0.8305	1.5412
Median	0.8035	1.3275
Standard deviation	0.5725	0.9061
Kurtosis	4.9054	9.4278
Skewness	0.8900	2.0532

The results extracted from figure 3 can also be quantified with the results from table 2. For the one lane method, the data consists of 638 headways. This is 60 more than for the sublane method, mainly due to the fact that for the sublane method the second or even third cyclists in a queue has no headway. The minimum headway for the one lane method is equal to 0.0003 seconds and the maximum is equal to 4.31 seconds. For the sublane method these statistics are larger, namely 0.1590 and 6.9490, respectively. Moreover, the headway distribution is as expected left-tailed for both situations as from table 2 can be concluded that the median is smaller than the mean. This is also underlined by the positive value for the skewness (0.8900 and 2.0532), which defines the measure of asymmetry. A perfectly symmetric data set would have a skewness equal to zero. The sublane method gives a more left-tailed distribution as its value is higher. For the one lane method, a standard deviation equal to 0.5725 is obtained. With that information, it is calculated that the largest value is two standard deviations larger than the second largest value and can therefore definitely be considered as a outlier. The sublane method gives no official outlier as there are a few very large headways. Finally, the kurtosis indicates the peak form of the distribution. The normal distribution has a kurtosis equal to 3 and is often used as a comparison. Because its values is larger than 3 (4.9054 and 9.4278), the distribution is so to say leptokurtic, which means that it produces more outliers than the normal distribution. It must be noted for the one lane method that leaving out the outlier leads to a kurtosis of 3.1306, which is almost similar to the normal distribution.

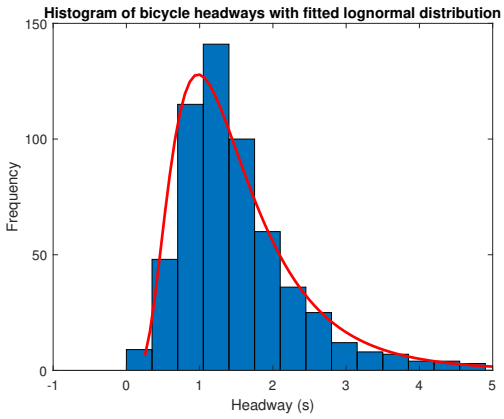
Next, the distributions proposed in section 2.1.1 are fitted to the headway data by means of maximum likelihood. The results of applying the normal, lognormal, exponential, gamma and the half normal distribution to the data are first shown in a graphical way in figure 4 below and in figure 8 in Appendix A. Figure 4 provides the distributions that fit the data the best and figure 8 shows the fits for the other distributions.

From figure 4 below and figure 8 in Appendix A it is easily shown that different distributions can take on a variety of shapes. Figure 8a and 8b shows the fitted symmetric normal distribution. This distribution is lacking a bit because of the fact that a headway is always positive and the fitted normal distribution takes on negative values. In order to overcome this problem, the half normal distribution has been introduced as shown in figures 4c and 4d. For the one lane method, this seems to give a slightly better fit, but the red line is unable to catch the bins of similar size up to 1.3 seconds. Two relatively comparable distributions in terms of shape are the lognormal and the gamma distribution in figures 4a and 4b and in figures 8e and 8f, respectively. The main difference

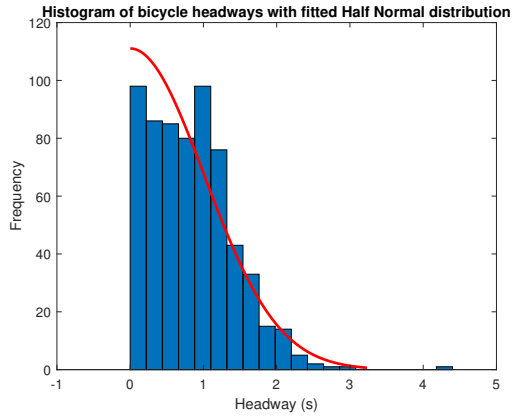
is that the gamma distribution is more smoothly, as the lognormal distribution has a larger peak. This is why the lognormal distribution seems to give a slightly better fit for the sublane method. The exponential distribution in figures 8c and 8d and the half normal distribution in figures 4c and 4d differ from the others as their distributions are monotonically downwards sloping. Their slopes are very different as the slope of the exponential distribution is the steepest at 0 seconds, whereas the slope of the half normal distribution is equal to zero for a headway of 0 seconds.



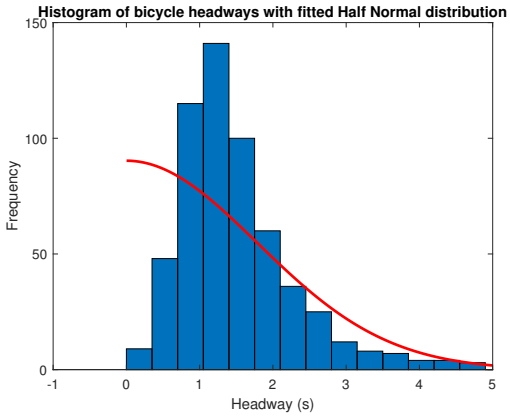
(a) Lognormal (one lane method)



(b) Lognormal (sublane method)



(c) Half normal (one lane method)



(d) Half normal (sublane method)

Figure 4: Histograms of the cyclist headway data at the signalized intersection in Amsterdam (June 6, 2016) with two different fitted distribution using maximum likelihood, namely the lognormal and the half normal distribution for the one lane method and the sublane method.

Table 3: Statistical results of the different headway distribution models for the one lane method for the signalized intersection in Amsterdam (June 6, 2016).

		Maximum likelihood Fitted parameter values			Goodness-of-fit test	
#	Distribution	First parameter	Second parameter	Negative logl.	Test value	p-value
1	Normal	$\hat{\mu} = 0.831$	$\hat{\sigma} = 0.572$	548.987	11.323	0.010
2	Lognormal	$\hat{\mu} = -0.583$	$\hat{\sigma} = 1.153$	624.524	233.960	0.000
3	Exponential	$\hat{\lambda} = 0.831$		519.522	135.486	0.000
4	Gamma	$\hat{k} = 1.403$	$\hat{\theta} = 0.592$	499.195	69.025	0.000
5	Half normal	$\hat{\mu} = 0$	$\hat{\sigma} = 1.008$	468.449	30.806	0.000

Table 4: Statistical results of the different headway distribution models for the sublane method for the signalized intersection in Amsterdam (June 6, 2016).

		Maximum likelihood Fitted parameter values			Goodness-of-fit test	
#	Distribution	First parameter	Second parameter	Negative logl.	Test value	p-value
1	Normal	$\hat{\mu} = 1.5412$	$\hat{\sigma} = 0.9053$	762.662	161.214	0.0000
2	Lognormal	$\hat{\mu} = 0.2849$	$\hat{\sigma} = 0.5522$	641.613	18.281	0.0026
3	Exponential	$\hat{\lambda} = 1.5412$		828.006	347.024	0.0000
4	Gamma	$\hat{k} = 3.5449$	$\hat{\theta} = 0.4348$	646.129	40.481	0.0000
5	Half normal	$\hat{\mu} = 0$	$\hat{\sigma} = 1.7874$	755.188	205.881	0.0000

Moving on to a more quantitative way of interpreting the graphs, tables 3 and 4 are used. These tables shows the results of the maximum likelihood test and the goodness-of-fit test for both the one lane method and the sublane method. For completeness, the parameter estimates with which the graphs in figure 4 and 8 are drawn are also given. First, a focus on the maximum likelihood values. These are given as a negative log likelihood, which needs some explanation. The negative log likelihood is one of the many ways to measure which model or distribution is able to describe the data best. It tries to minimize the error, so in this case the lower the value of the negative log likelihood, the better the fit of the distribution. This is very intuitive and can also be extracted from the headway distribution figures. The more the histogram looks like the distribution, the better its fit. From table 3 it is concluded that the half normal distribution creates the best fit in terms of maximum likelihood for the one lane method as it has the smallest value (468.449), followed by the gamma distribution (499.195). For the sublane method in table 4 on the other hand, the lognormal distribution (641.613) gives the best fit as its maximum likelihood value is slightly small than the gamma distribution (646.129).

The two right columns of tables 3 and 4 show the results of the goodness-of-fit test. The goodness-of-fit test gives a value, which can then be translated to a p-value. The p-value and the test value are inversely related, so the lower the test value, the higher the p-value. For the one lane method in table 3, for four out of five distributions, the p-value is equal to 0.000. This means that the null hypothesis that the headways are distributed from one of these four distributions is strongly rejected. These four distributions are the lognormal, exponential, gamma and half normal distribution. In order to be able to compare the results with the same p-value, the test values are given. For the four aforementioned distribution, it is shown that the half normal distribution (30.806) fits best to the data, relatively, as it has the lowest test value. Performing the goodness-of-fit test for the normal distribution leads to a p-value of 0.010. As this result is lower than the critical value of 0.05, the null hypothesis is still rejected in favour of the alternative hypothesis, which states that the headways of the one lane method are derived from another distribution. Nevertheless, it is concluded that the normal distribution is the distribution that fits the headway data best, relatively, according to the goodness-of-fit test. Then, for the sublane method in table 4, there is a similar situation as the p-value for four out five distributions is equal to 0.000, namely for the normal, exponential, gamma and half normal distribution. If only those distributions are considered, the gamma distribution gives the best fit as it has the lowest test value (40.481). The highest p-value as a result of the goodness-of-fit test is obtained for the lognormal distribution, namely 0.0026. The critical value is equal to 0.05, and therefore the null hypothesis is still rejected in favour of the alternative hypothesis, which states that the headways of the sublane method are derived from another distribution. Nevertheless, according to the goodness-of-fit test, it is concluded that the lognormal distribution is the distribution that fits the data best, relatively, .

The maximum likelihood and the goodness-of-fit test lead to different conclusions. This is a perfect example of why it is always clever to include multiple tests as they may lead to different results. In conclusion, the headway data for both situations comes from none of the examined distributions and overall, the half normal distribution performs best for the one lane method as it has the lowest negative log likelihood and the second smallest test value as shown in table 3. For the sublane

method, the lognormal distribution gives the best fit, for the maximum likelihood test as well as the goodness-of-fit test.

4.2 Discharge flow definitions

Two widely used methods for calculating discharge flow are given in equations (1) and (4), namely with headways or with Edie's definitions, respectively. It is interesting to compare the two methods for discharge flow as they both describe the same flow characteristic, but calculated in different ways. Figure 5 shows the results of the two different ways of calculating the discharge flow. It is clear that there is often no difference between the two methods. In 35 (out of 57) cases the calculated discharge flows are exactly the same. In the 22 periods in which differences occur, the discharge flow calculated with the headway almost always generates a higher flow. The differences are always relatively small, namely not more than 10% of the discharge flow in that period.

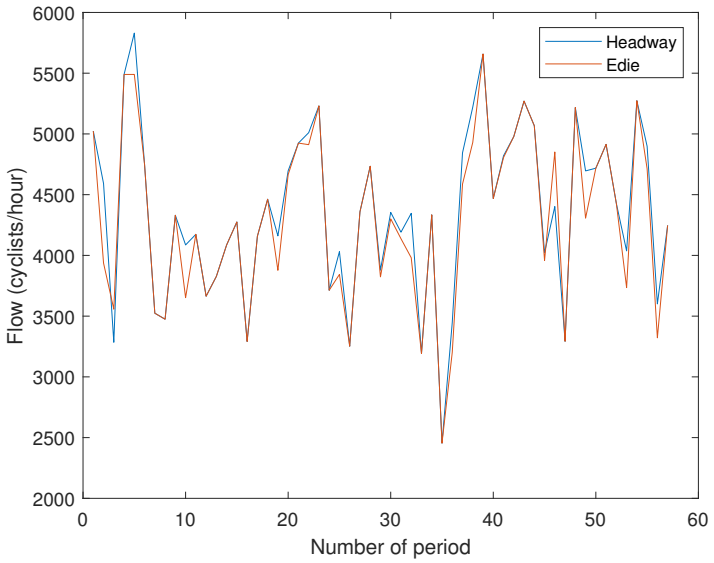


Figure 5: The results of the two methods of calculating discharge flow, namely with headway (in blue) and according to Edie's definitions (in red) for 57 periods of the signalized intersection in Amsterdam (June 6, 2016)..

Furthermore, their correlation is calculated with equation (8) and this results in a value of 0.971. The maximum correlation is equal to 1, so the positive linear relationship between the two methods for discharge flow is very strong.

4.3 Saturation flow

Saturation flow depends on saturation headway as stated in equation (5). Saturation headway is described as the headway in a queue when the time between two consecutive cyclists remains constant. It is assumed that this headway is not constant at first when cyclists are accelerating. It is also assumed that when the cyclists pass the stop line, their speed is constant and therefore also their headways. Two different situations are identified, namely one in which the bicycle path is one lane and the other one which introduces the concept of sublanes as explained in section 3.5. First, the results for the one lane method are given in subsection 4.3.1 and then the results for the sublane method are given in subsection 4.3.2.

4.3.1 Saturation flow for the one lane method

Figure 6 shows the headways for the one lane method for each cyclist in the data set. The horizontal axis shows the number of the headway per period, so as there are 57 periods, the leftmost column shows 57 circles. This is because in every period there is a first headway, namely the difference in seconds between the first and second cyclist. The number of dots is slowly decreasing and the 14^{th} column (the headway between the 14^{th} and 15^{th} cyclist) for example has 9 circles, as there are only 9 periods with 15 or more cyclists. The vertical axis shows the length of the headway. The upper circle is recognized as the already mentioned outlier in section 4.1, with a headway of more than 4 seconds. When calculating the saturation headway, only headways up to the 12^{th} headway are taken into account as after that there a too few observations. The blue line indicates the mean headway per number of headway in each period. This line is relatively constant except for the right part where less data points are observed. As there is no downward sloping line as expected according to 3.4 it is inferred that the method for saturation flow as described in [Raksuntorn and Khan, 2003] does not account for this set of data in case of the one lane method. The headway mean for the first 12 cyclists is equal to 0.8342 seconds and is very close to the mean (0.8305) of the whole data set as calculated in table 2. Converting this to a saturation flow results in

$$q_s = \frac{3600}{h_s} = \frac{3600}{0.8342} = 4316 \text{ cyclists/hour} \tag{12}$$

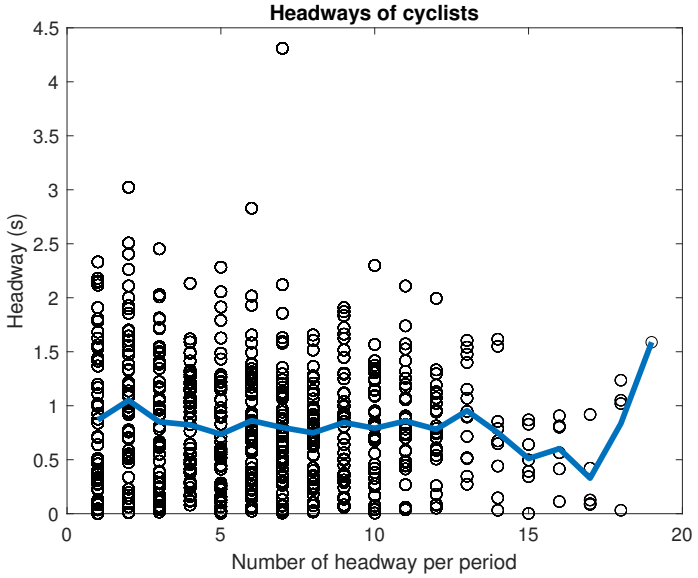


Figure 6: Headways of cyclists for each period for the one lane method for the signalized intersection in Amsterdam (June 6, 2016). The black circles show the headway per cyclists and the blue line is the average headway per number of headway per period.

4.3.2 Saturation flow for the sublane method

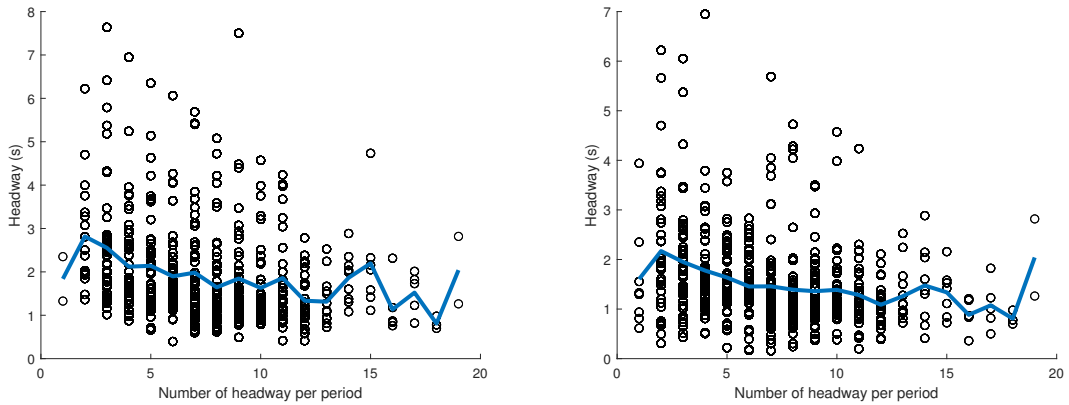
The results of the one lane method in subsection 4.3.1 rejected the null hypothesis that average headways across cyclists would decrease and then get constant. As an extension, the sublane method is introduced [Botma and Papendrecht, 1991]. This method has an extra parameter, namely the width of the sublanes w_{sub}, so by varying this parameter, different results can be obtained. Table 5 shows the results of an analysis in which the sublane width is varied. First of all, as a result it is noted that the width of the sublanes is negatively correlated with the saturation headway h_s in the third column. An increase in the width means a decrease in the saturation headway. This can be explained as with a larger width, a bigger range is taken into account for the calculation of the headway. This is accompanied by a larger probability that a closer cyclist and as a result a smaller

headway is found. The reason why this range of sublanes is chosen is the following. For sublane widths smaller than 0.15 m, there is no saturation headway anymore and for sublane widths larger than 0.50 m, the result is close to the result of only one lane as shown in figure 6. The decrease in headways gets very small when the width gets very large.

Table 5: The different results for the sublane method when the width of the sublanes is varied for the signalized intersection in Amsterdam (June 6, 2016). The headway number is the number after which the headways get constant and the saturation is achieved. The other variables are the sublane width w_{sub}, the saturation headway h_s, the saturation flow q_s, the number of total lanes L_{total} and the total saturation flow $q_{s,total}$.

w_{sub} (m)	Headway (-)	h_s (m)	q_s (cyc/h)	L_{total} (-)	$q_{s,total}$ (cyc/h)
0.15	8	1.79	2009	4.4	8840
0.20	9	1.55	2330	3.3	7689
0.25	8	1.38	2616	2.7	7063
0.30	8	1.21	2979	2.2	6554
0.35	8	1.05	3424	1.9	6506
0.40	8	0.98	3680	1.7	6256
0.45	8	0.93	3868	1.5	5802
0.50	8	0.91	3962	1.3	5151

For most cases, the saturation headway is obtained at the 8^{th} headway, so the 9^{th} cyclist. It has to be stressed that the there is very few data for long periods, so the part of the figure after the 12^{th} headway is not taken into account. The same reasoning is also applied to the second cyclist, as there are very few second cyclists per period with a headway. In section 3.4 is described that a monotonically downward sloping headway mean is expected across the cyclists in the queue. This is not the case for all the examined sublane widths. For example, this accounts for sublane width $w_{sub} = 0.15$ m as shown in figure 7a, which shows some so called 'zigzag' movements. This situation is therefore not monotonically downward sloping. Of all examined models, it is found that the model with sublane width $w_{sub} = 0.25$ m provides the best results as shown in figure 7b. Figure 7b shows a smooth curve, from the 3^{th} until the 12^{th} headway. Moreover, an average Dutch cyclist needs a width of 0.75 m [Boersma et al.]. This is perfectly in line with $3 \times 0.25 = 0.75$ as in equation (9).



(a) Width = 0.15 m

(b) Width = 0.25 m

Figure 7: Headways of cyclists for each period for the method with the sublanes for the signalized intersection in Amsterdam (June 6, 2016). The black circles show the headway per cyclists and the blue line is the average headway per number of cyclist per period.

The rightmost column of table 5 shows the total saturation flow $q_{s,total}$ as calculated with equation

(10). The total saturation flow is defined as the number of cyclists that can pass the signalized intersection in case they optimally use the cyclist path. With an increasing sublane width, the total saturation flow decreases. From this can be concluded that the total number of sublanes is decisive for the differences in total saturation flow for different sublane widths. The calculated total saturation flows for small sublane widths are in line with other findings when using the sublane method [Botma and Papendrecht, 1991], but are larger than the findings in other papers about bicycle saturation flow [Raksuntorn and Khan, 2003] [Rahman et al., 2005] and much larger than the calculated saturation flow for the one lane method in equation (12).

4.4 Summary of the results

The distributions that give the best fit for the one lane method and the sublane method are the half normal distribution and the lognormal distribution, respectively. This is checked with maximum likelihood and the goodness-of-fit test. The latter test uses the null hypothesis that the headway data comes from the proposed distribution. This hypothesis is rejected in all cases. Moreover, the two methods of calculating discharge flow, namely with headway and with Edie's definition provide similar results. Next to that, saturation flow is calculated for the one lane method as well as the sublane method. The one lane method shows no signs of saturation flow, but for the sublane method this is achieved at the 9^{th} cyclist in a queue.

5 Discussion

The research about bicycle traffic flow at signalized intersections provides some interesting outcomes. This chapter goes into the depths about the added value of the results and how they can be improved and extended.

First of all, the research about headway distribution models and saturation flow has been carried out for two different situations. Namely a relatively simple situation in which the bicycle path is considered as one lane and a somewhat more advanced situation which the bicycle path is split up in different sublanes. The sublanes have been created with an intuitive method [Botma and Papendrecht, 1991] which states that the headway is calculated with the cyclist in front within a range of 3 adjacent lanes. This method seems solid, but it is difficult to derive the optimal sublane width. Table 5 compares some results, and then based on the smoothness of the saturation headway curve and logical reasoning an optimal sublane width is chosen. This research could be more profound by quantifying the analysis better.

Next to the sublane method, five different headway distribution models are taken a look at. All the proposed models are frequently used in such circumstances, as a queue or waiting times. Except for the normal and the half normal distribution, which were added after a first look at the data. It is probably not useful to expand this part of the research by adding more distributions. The main problem is the lack of significance, which is expressed by a insignificant p-value (smaller than 0.05). This means that the headways for the one lane method as well as the sublane method are not distributed from one of the five proposed models. There can be various reasons for this, and further research could look into possibilities to solve this problem. Possible solutions are splitting up the data in periods with few cyclists and periods with many cyclists, only looking at periods during rush hour or examining the outliers in order to find out what their nature is.

The methodology about saturation flow [Raksuntorn and Khan, 2003] is quite straightforward, but did not work work out at first for the one lane method. The saturation flow theory was extended by the introduction of sublanes and this led to the conclusion that the saturation headway is achieved after the 9^{th} cyclist. The report also defines how to calculate the total saturation flow once the saturation flow of a sublane is known by multiplying it with the total number of lanes in the bicycle path. This part can be improved by adjusting the equation about total saturation flow as these values in table 5 seems to be too large when compared to other research.

Cyclist behaviour can be very stochastic and therefore the outcomes of the signalized intersection in Amsterdam cannot be expanded to other signalized intersection, although is seems fair to assume that cyclists across Amsterdam behave the same. To substantiate these findings, two examples are given. [Raksuntorn and Khan, 2003] shows that the saturation headway for intersections in Colorado and California is obtained at the 5^{th} cyclist in a queue, but that method does not take into account sublanes. Moreover, research by [Rahman et al., 2005] about intersections in Yokohama, Japan and Dhaka, Bangladesh shows that the saturation headway of cyclists in these cities is larger than the saturation headway for the signalized intersection in Amsterdam. So when using the same methodology for a set of cyclist data at another signalized intersection, it is very likely that different values for the flow characteristics will be derived. Therefore, it is very interesting to find out what the differences are between signalized intersection across the world.

Finally, the goal is to translate the results to actual actions in terms of redesigning the signalized intersection or the bicycle path. There are various possibilities, but it should be tested which solutions are effective. An example of an design solution is the marking of the different sublanes in order to try to prevent cyclists to cycle in the middle of the road.

6 Conclusion

This report takes different approaches in deriving bicycle flow characteristics for a signalized intersection at Weteringlaan, Amsterdam. The focus is on headway characteristics, from which several flow characteristics can be calculated. The data consists of 57 periods with in total 695 cyclists, whose coordinates are known at several time stamps, so their movement can be tracked. If the bicycle path is considered as one lane, the headway of a cyclist is calculated by looking only at the cyclist in front him. It is unlikely that the headway of this cyclist is always directly connected to the cyclist in front of him as the latter may be cycling at the opposite side of the road. Therefore, two different situations for obtaining the headways are defined. The first situation considers the bicycle path as one lane, whereas the second situation splits up the bicycle path in sublanes and the headway is defined by the closest cyclist in front within the range of 3 adjacent sublanes left and 3 adjacent sublanes right of the cyclist.

As mentioned, two approaches are used to derive characteristics. The first approach is a more statistical one, in which different distribution models are compared with maximum likelihood and the χ^2 goodness-of-fit test to find out which distribution fits the headway data best. The main finding is that the headways do officially not come from one of the proposed distribution models, due to the fact that the goodness-of-fit test rejects the null hypothesis that the headways come from the considered distribution. Nevertheless, for the one lane method the half normal distribution is able to fit the data best. For the sublane method, with the lognormal method the best results are obtained.

The second approach entails the derivation of saturation flow, which is directly linked to saturation headway. For the one lane method, it is found that the headways are constant among all the cyclists in a period. For the sublane method on the other hand is derived that the headways decrease among the cyclists in a period and get constant for the 9^{th} cyclist. Due to the fact that a cyclist takes up 75 cm of the bicycle path width and the smoothness of the saturation headway curve, it is chosen that a sublane width of 25 cm gives the best impression of reality.

Finally, two methods of discharge flow are compared. Discharge flow can be calculated with Edie's definitions and with headway. Both methods lead to similar results and have a very strong positive correlation.

References

W. F. Adams. Road traffic considered as a random series. *Institution Civil Engineers J/UK/*, 1936.

R. Boersma, B. van Arem, and F. Rieck. Casestudy appelscha: een onderzoek naar de implementatie van automatisch vervoer in het buitengebied.

H. Botma and H. Papendrecht. Traffic operation of bicycle traffic. *Transportation Research Record*, (1320), 1991.

S. Chand, N. J. Gupta, and S. Velmurugan. Development of saturation flow model at signalized intersection for heterogeneous traffic. *Transportation research procedia*, 25:1662–1671, 2017.

R. J. Cowan. Useful headway models. *Transportation Research*, 9(6):371–375, 1975.

R. W. Edie. Devonian limestone reef reservoir, swan hills oil field, alberta. *Trans. Canad. Inst. Min. and Metall*, 54:278–285, 1961.

B. Goñi-Ros, Y. Yuan, W. Daamen, and S. P. Hoogendoorn. Empirical analysis of the macroscopic characteristics of bicycle flow during the queue discharge process at a signalized intersection. Technical report, 2018.

S. Hoogendoorn and V. Knoop. Traffic flow theory and modelling. *The Transport System and Transport Policy: An Introduction*, pages 125–159, 2013.

J. Lebacque and M. Khoshyaran. First order macroscopic traffic flow models for networks in the context of dynamic assignment. In *Transportation Planning*, pages 119–140. Springer, 2002.

M. M. Rahman, S. Ahmed, and T. Hassan. Comparison of saturation flow rate at signalized intersections in yokohama and dhaka. In *Proceedings of the Eastern Asia Society for Transportation Studies*, volume 5, pages 959–966, 2005.

W. Raksuntorn and S. Khan. Saturation flow rate, start-up lost time, and capacity for bicycles at signalized intersections. *Transportation Research Record: Journal of the Transportation Research Board*, (1852):105–113, 2003.

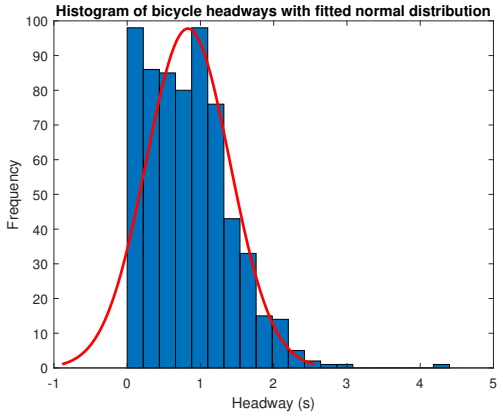
J. Tolle. The lognormal headway distribution model. *Traffic Engineering & Control*, 8(8), 1971.

Y.-G. Wang, G. Wei, X. Zhu, and Y.-L. Pei. Capacity of bicycle platoon flow at two-phase signalized intersection: a case analysis of xi'an city. *PROMET-Traffic&Transportation*, 23(3):177–186, 2011.

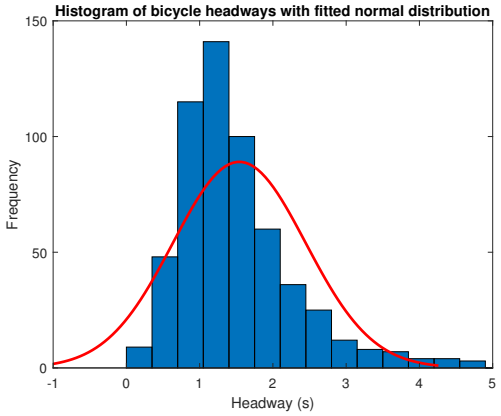
F. Ye and Y. Zhang. Vehicle type-specific headway analysis using freeway traffic data. *Transportation Research Record: Journal of the Transportation Research Board*, (2124):222–230, 2009.

G. Zhang, Y. Wang, H. Wei, and Y. Chen. Examining headway distribution models with urban freeway loop event data. *Transportation Research Record: Journal of the Transportation Research Board*, (1999):141–149, 2007.

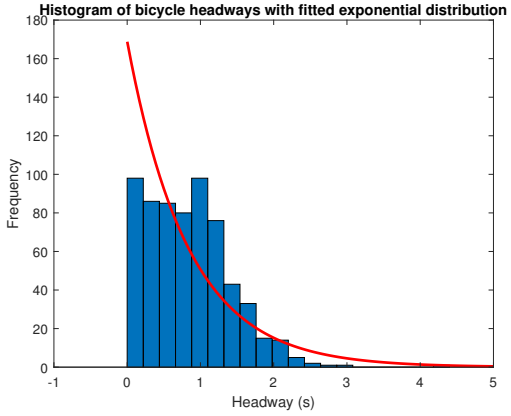
Appendix A



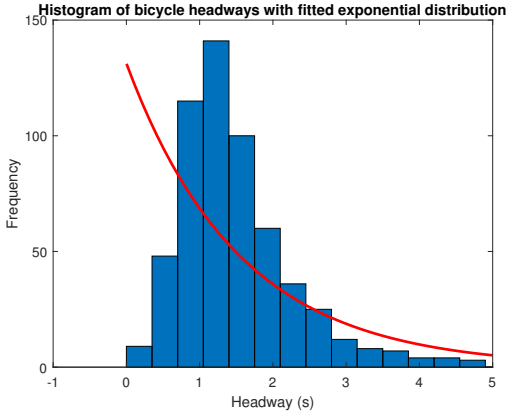
(a) Normal (one lane method)



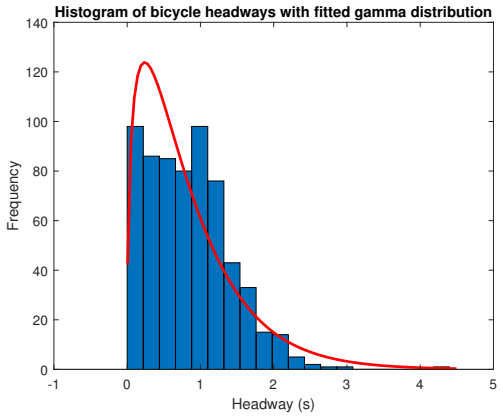
(b) Normal (sublane method)



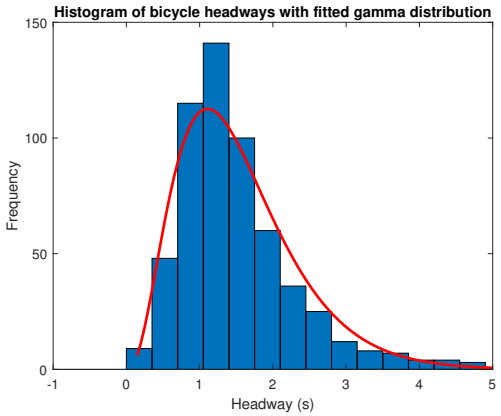
(c) Exponential (one lane method)



(d) Exponential (sublane method)



(e) Gamma (one lane method)



(f) Gamma (sublane method)

Figure 8: Histograms of the cyclist headway data at the signalized intersection in Amsterdam (June 6, 2016) with three different fitted distributions (normal, exponential and gamma) using maximum likelihood for the one lane method and the sublane method.