Technische Universiteit
Eindhoven
University of Technology

Department of Mathematics and Computer Science

Postbus 513, 5600 MB Eindhoven
The Netherlands
www.tue.nl

Author
Mehrsa Nejatianrad (1650009)
Alexandra Murariu (1769839)

Responsible Lecturer
Marko Boon
Céline Comte

Date
October 25, 2021

A Study on Traffic Jams on A2 Highway, The Netherlands

Subtitle

Mehrsa Nejatianrad (1650009)
m.nejatianrad@student.tue.nl

Alexandra Murariu (1769839)
a.murariu@student.tue.nl

Technische Universiteit
Eindhoven
University of Technology

Table of contents

Title
A Study on Traffic Jams on A2 Highway, The Netherlands

1 Introduction

Traffic jams are getting more frequent every day as the number of cars on Dutch motorways grows. In order to acquire a better understanding of the economic impact of lost hours from drivers delayed in traffic jams, a research was done on the dataset given by Rijkswaterstaat. The dataset includes 26 columns providing information on each traffic jam on the Dutch motorway network, from January 1, 2015. For the purpose of this study the data from September, 2016 and A2 highway was chosen. A2 highway is one of the most heavily traveled roads in the Netherlands. The route connects the Belgian A25 road and the city of Amsterdam, near the Amstel interchange, with the Belgian border, near Maastricht (NL) and Liège. (B)[1].
The purpose of this study is finding out the main causes and frequency of traffic jams in order to give solutions for improving the current situation.

Outline

- Chapter 2. Descriptive statistics - overview on traffic jams situation in the Netherlands and A2
- Chapter 3. Distribution fitting - distributions are fitted to the time between traffic jams and the duration of traffic jams.
- Chapter 4. Simulation - stochastic simulation of the cumulative traffic jam duration
- Chapter 5. Conclusions and recommandations - ideas on how to improve the traffic jams situation

2 Descriptive statistics

This section provides an overall insight on the traffic jam situation on all highways. It also concentrates on specific parts of data from highway A2 and gives more details about the problem. The results will later motivate the choice of stochastic process division, since traffic jams' frequency and durations are heavily influenced by their causes and the day of the week in which they take place. In the appendix some general data statistics are shown:A.14

2.1 Number of traffic jams per highway

In order to get a better understanding of on which highways in The Netherlands traffic jams are more likely to occur, a bar plot is shown. As can be seen in figure **??**, A1 and A2 highways have the highest number of traffic jams which are 2058 and 1559. On the other hand, A65, N50, N261 and N65 have only 1 traffic jam each.

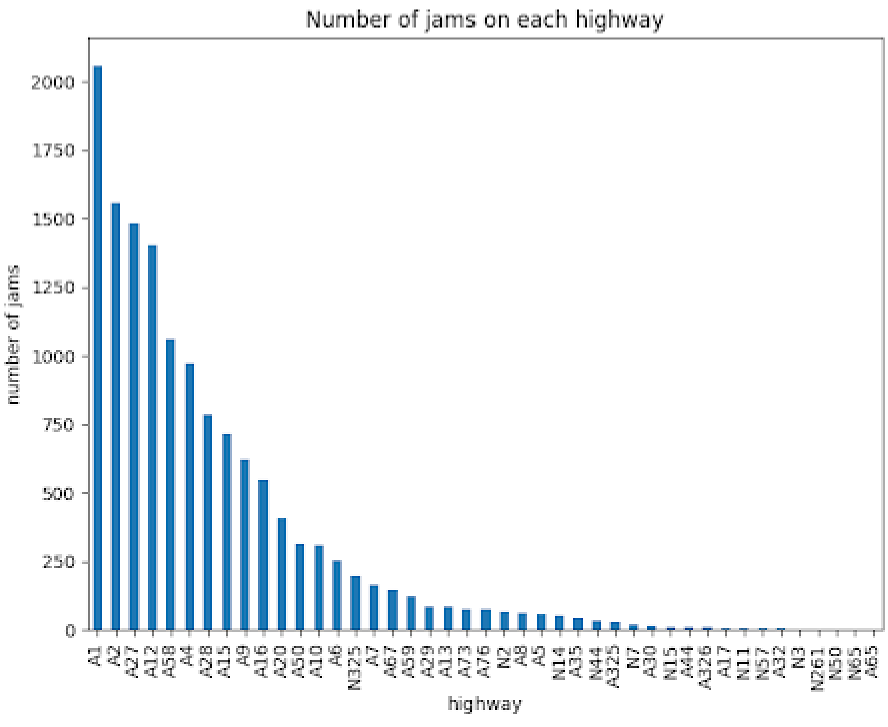


Figure 2.1: Histogram of the number of traffic jams per weekday

2.2 A study on traffic jams on A2 per weekdays

2.2.1 Number of traffic jams per day

In order to determine on which days of the week more traffic jams occur, a bar plot was made. Figure 2.2 illustrates that during the working days, especially on Tuesday, Thursday and Friday, more traffic jams take place than during the weekend.

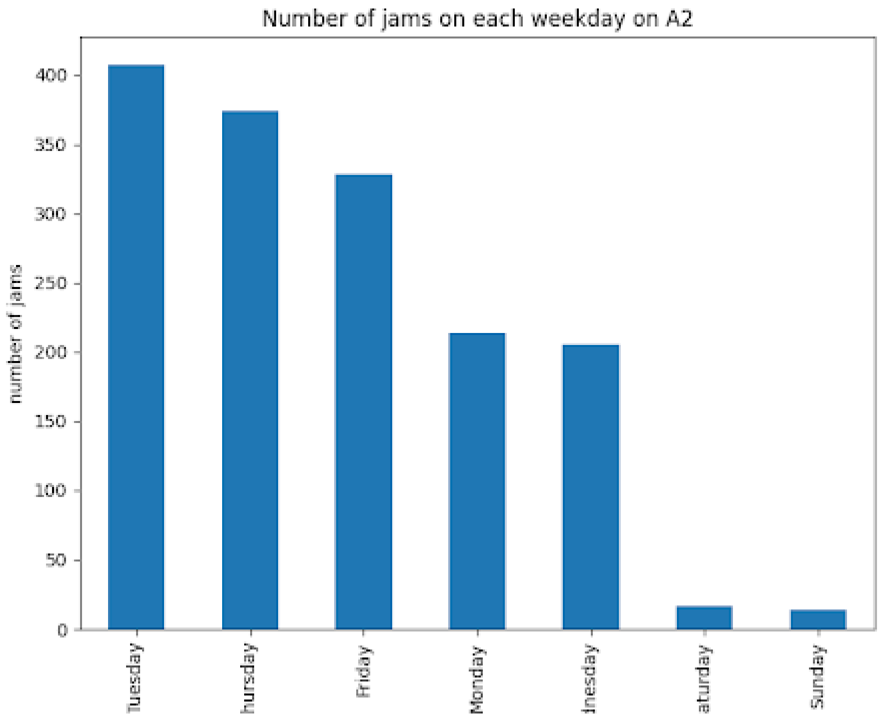


Figure 2.2: Bar plot of the number of traffic jams per day

2.2.2 Main traffic jam causes for the whole month

The top 10 main causes of traffic jams were plotted. As can be seen in Figure 2.3, rush hour traffic jams are the most frequent ones. Also, traffic jams that occur outside rush hours and the ones that are caused by accidents are the second and third most frequent traffic jams. In figures A.1 top causes for each day of the week are plotted.

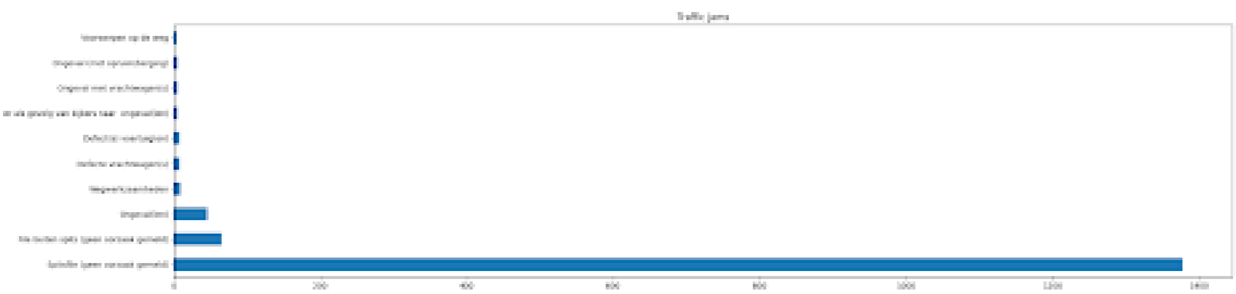


Figure 2.3: Bar plot of the main traffic jam causes for the whole month of September

3 Distribution fitting

This section yields a better understanding of the duration of traffic jams and time between them. Several distributions are plotted against the histograms and ECDFs of datasets, which are separated later by their categories that influence most the resulting distribution: day of the week and cause (as seen in the previous section). After testing different distributions, the best are assigned to fit into the later presented simulation process.

3.1 Duration of traffic jams

In this section data is divided by cause of the traffic jam. Only the top 3 causes of traffic jams are taken into consideration, since the other traffic jams occur very rarely and a proper distribution cannot be concluded from the sample data. However, in the final simulation a margin of error will be added in order to compensate for the ignored traffic jams.

3.1.1 All traffic jams' duration

Here are presented histograms of the durations of all traffic jams on A2 the whole month against several known distributions.

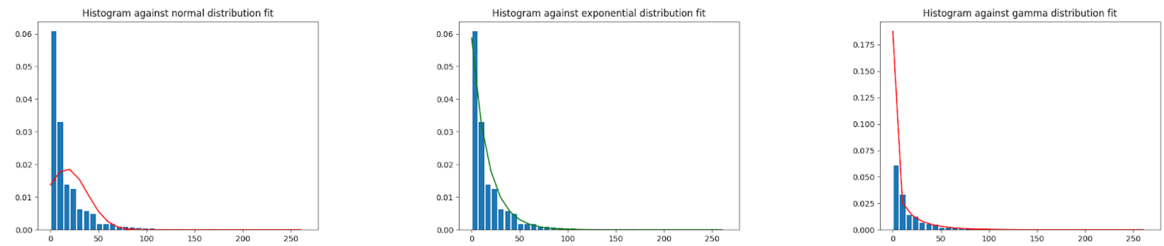


Figure 3.1: Histograms with their distribution fits

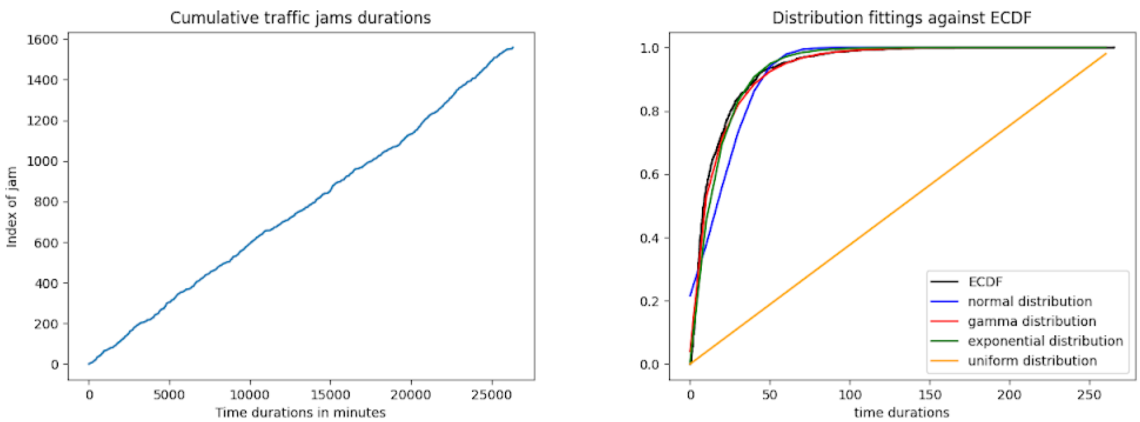


Figure 3.2: CDF and ECDF of jams' durations against several distributions

From the graphs and K-tests, the best fit for this data is either a gamma or an exponential distribution.

3.1.2 Duration of rush hour jams (Spitsfile)

An important thing to note is that rush hour jams don't take place during weekends. Later, while constructing the simulation, this will also be taken into consideration. In the appendix (figure A.5) are presented histograms of rush hour jams' durations, plotted against several well-known distributions. Here the ECDF against distributions fits is shown.

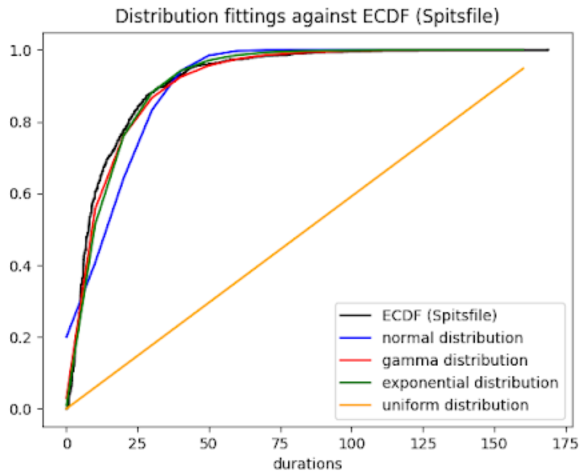


Figure 3.3: ECDF with several known distributions

From the graphs and K-tests, the best fit for this data is either a gamma or an exponential distribution.

3.1.3 Duration of outside rush hour jams (File buiten spits)

In the appendix are presented histograms of outside rush hour jams' durations, plotted against several well-known distributions. In the appendix (figure A.5) are presented histograms of those jams plotted against several well-known distributions. Here the ECDF against distributions fits is shown.

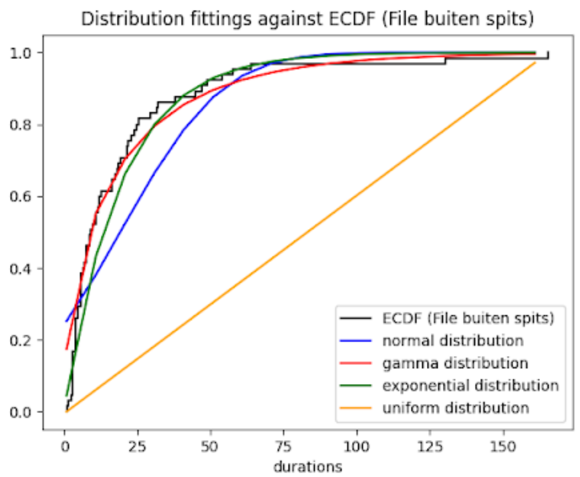


Figure 3.4: ECDF with several known distributions

From the graphs and K-tests, the best fit for this data is either a gamma or an exponential distribution.

3.1.4 Duration of jams caused by accidents ("Ongeval" jams)

In the appendix (figure A.7) are presented histograms of durations of jams caused by accidents, plotted against several well-known distributions. Here the ECDF against distributions fits is shown.

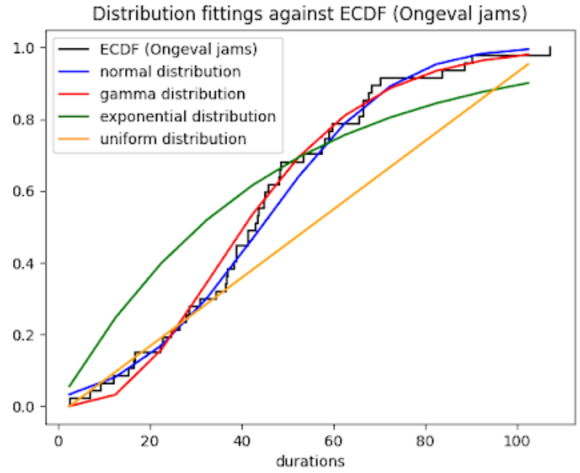


Figure 3.5: ECDF with several known distributions

From the graphs and K-tests, the best fit for this data is either a gamma or a normal distribution.

3.2 Time between traffic jams

As seen in section 2.2.1, traffic jams are more likely to occur on working days. Thus, in this section data is divided into two parts: jams that took place during the weekends and jams that took place during working days. In September 2016, there were 9 weekend days and 21 working days. In the final simulation this time difference will be taken into consideration when compounding the Poisson processes. In figure A.2 are plotted time intervals between traffic jams for the whole month.

3.2.1 Time intervals between all jams

Here are presented histograms of the time intervals between all traffic jams on A2 the whole month against several known distributions.

From the graphs and K-tests, the best fit for this data is the gamma or exponential distribution

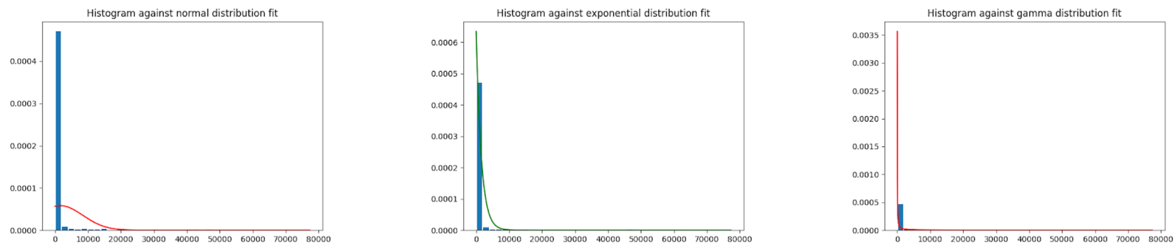


Figure 3.6: Histograms with their distribution fits

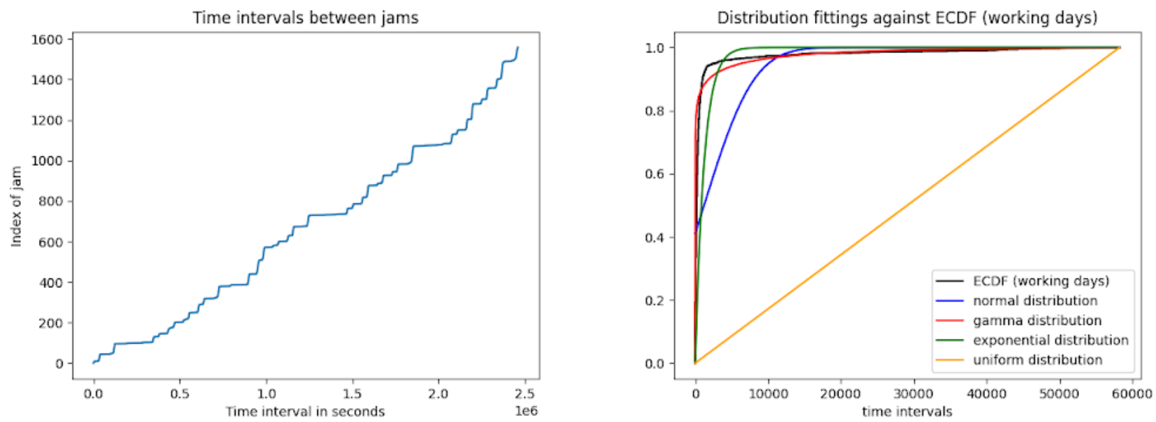


Figure 3.7: CDF and ECDF of time intervals between jams against several distributions

3.2.2 Time intervals between jams that take place during working days

An important thing to note is that during working days traffic jams are more likely to occur. In the appendix (figure A.3) are presented histograms of time intervals during working days, plotted against several well-known distributions. Here the ECDF against distributions fits is shown.

From the graphs and K-tests, the best fit for this data is the exponential distribution.

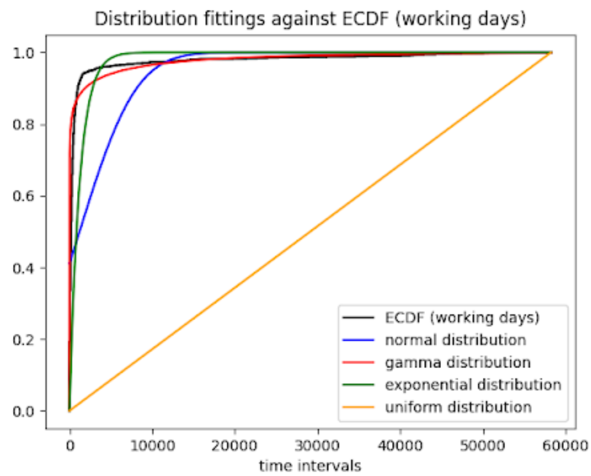


Figure 3.8: ECDF with several known distributions

3.2.3 Time intervals between jams that take place during weekends

An important thing to note is that during weekends traffic jams are less likely to occur. In the appendix(figure A.4) are presented histograms of time intervals during weekends, plotted against several well-known distributions. Here the ECDF against distributions fits is shown.

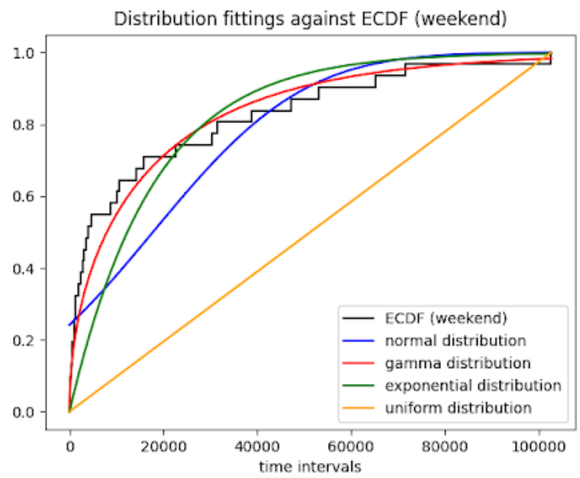


Figure 3.9: ECDF with several known distributions

From the graphs and K-tests, the best fit for this data is the gamma distribution.

3.3 Time between traffic jams by causes and time of the week

In this section data is divided even more, being separated by day of occurence and cause at the same time. This will become very useful later, for creating a much more realistic model of the data, as more appropriate results are concluded from those separate distributions that are presented below.
As mentioned above, rush hour jams occur only on working days. Below are presented several ECDFs. They reveal the best distribution fittings for all separate cases.

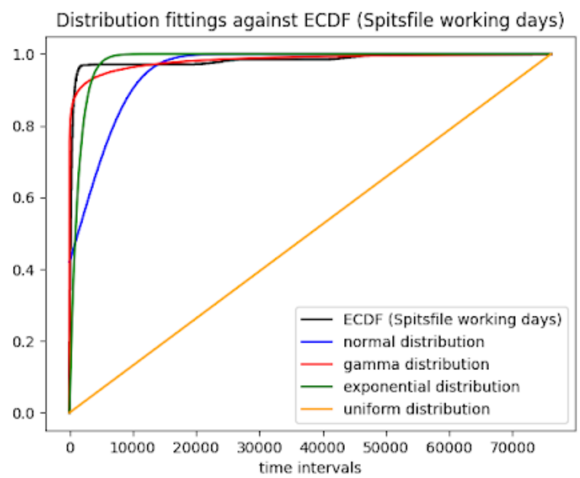


Figure 3.10: Time intervals between rush hour traffic jams

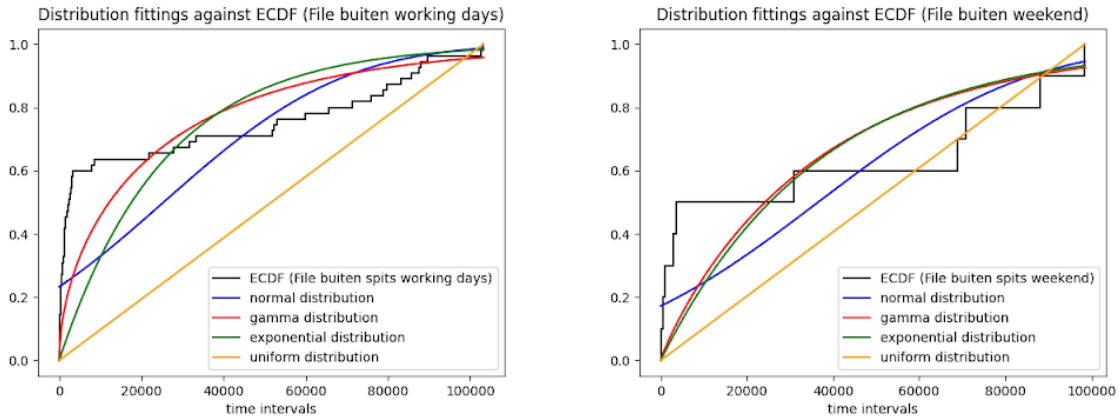


Figure 3.11: Time intervals between outside rush hour traffic jams

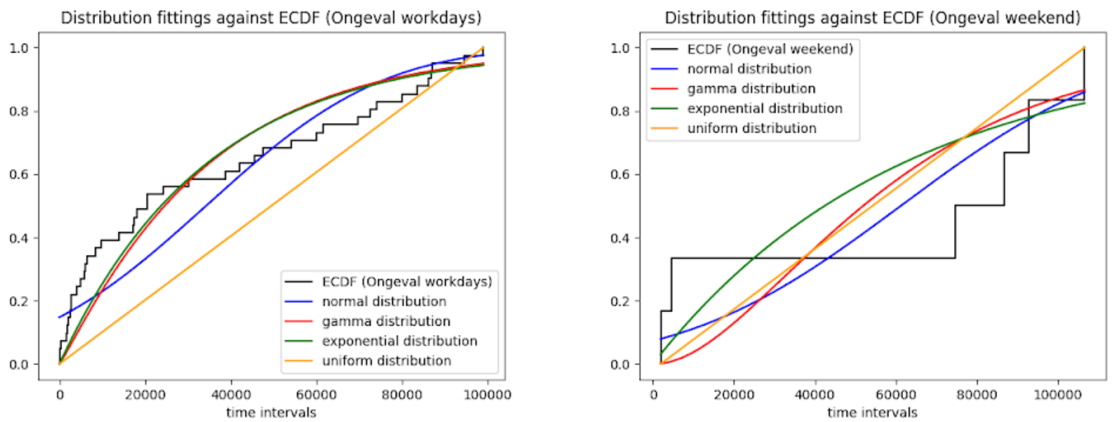


Figure 3.12: Time intervals between accident caused traffic jams

4 Simulation

4.1 Model description

In this section a stochastic simulation of the traffic jams on A2 is presented. This mathematical model gives insight into the cumulative duration of all traffic jams on A2, during September 2016.
Method 1: A compound Poisson process was modelled using the rate and distributions determined in the previous chapter. More precisely, the rate used is $\frac{1}{\lambda}$, where λ is the mean of time distances between all traffic jams (regardless of their cause or time of occurence). The distribution chosen for traffic jams durations is the exponential one. The total time is 43200 (number of minutes in one month). After this, all traffic jams' durations were summed up in order to obtain the total duration of traffic jams from a month.
Method 2: For this method, separate Compound Poisson Processes were made for each of the following datasets:

- Rush hour jams (taking place only on working days)

  - For this process the rate was given by $\frac{1}{\lambda}$, where λ is the mean of time distances between rush hour jams. The data was thinned so that the distance between Friday and Monday would be equal to 0 (as if Monday would follow Friday), because no rush hour jams occur during weekends. This type of thinning was also done for the following processes. Gamma, exponential and normal distributions were simulated for the rush hour jams' durations, from which the normal distribution seemed most appropriate. For the time parameter, the total time 30240 minutes was inputed in the simulations, since it resembles the 21 working days in minutes. For the number of runs see section 4.3, where this is detailed.

- Outside rush hour jams during weekends

  - For this process the rate was given by $\frac{1}{\lambda}$, where λ is the mean of time distances between outside rush hour jams during weekends. Gamma, exponential and normal distributions were simulated for the outside rush hour jams' durations that take place in weekends, from which the gamma distribution seemed most appropriate. For the time parameter, the total time 10080 minutes was inputed in the simulations, since it resembles the 9 weekend days in minutes.

- Outside rush hour jams during working days

  - For this process the rate was given by $\frac{1}{\lambda}$, where λ is the mean of time distances between outside rush hour jams during working days. Gamma, exponential and normal distributions were simulated for the outside rush hour jams' durations that take place during working days, from which the normal distribution seemed most appropriate.

- Jams caused by accidents during weekends

  - For this process the rate was given by $\frac{1}{\lambda}$, where λ is the mean of time distances between jams caused by accidents that take place during weekends. Gamma, exponential and normal distributions were simulated for the accident caused jams' durations that take place during weekend days, from which the gamma distribution seemed most appropriate.

- Jams caused by accidents during working days

  - For this process the rate was given by $\frac{1}{\lambda}$, where λ is the mean of time distances between jams caused by accidents that take place during working days. Gamma, exponential and normal distributions were simulated for the accident caused jams' durations that take place during working days, from which the normal distribution seemed most appropriate.

4.2 Implementation

First are presented results derived from the sample. Later the Python implementation and results from the method 1 and 2 derived processes simulations are shown.
From the sample, the following results were obtained:
S(t) - total duration of jams: 26324.896999999866
Rush hour jams total duration: 20903.361
Outside rush hour jams during working days total duration: 997.499
Outside rush hour jams during weekend days total duration: 144.844
Accidents caused traffic jams during working days total duration: 1457.634
Accidents caused traffic jams during weekend days total duration: 292.816

4.2.1 Method 1

Method 1 is described in the previous chapter. For this method and also for Method 2, the function simulateCompoundPoissonProcess(lam, jumpDist, T) was used. Below the implementation is shown:

```
def simulateCompoundPoissonProcess(lam, jumpDist, T):
    arrivalTimes = deque()  # the most efficient data structure
    levels = deque()  # the levels of the CPP
    currentLevel = 0
    expDist = stats.expon(scale=1 / lam)
    t = expDist.rvs()
    print(t)
    while t < T:
        print(t)
        arrivalTimes.append(t)
        currentLevel += jumpDist.rvs()
        levels.append(currentLevel)
        t += expDist.rvs()

    return arrivalTimes, levels
```

This function simulates a compound poisson process (it adds the durations of each traffic jam together on and on, until time T is reached). The T used in method 1 was the number of minutes in one month. The function was called like this:

```
jumpdist = stats.expon(1 / lam)
data_simulation = simulateCompoundPoissonProcess(rate, jumpdist, 43200)
```

λ is the mean of time distances between all traffic jams (regardless of their cause or time of occurence). The total duration of all jams is computed as follows:

```
print(data_simulation[1][-1])
```

because the last element on the y-axis in the process is the total duration of jams. This method resulted in S(t) = 23163.883738420227, which is a decent result, but still far from reality.

4.2.2 Method 2

Method 2 is described in the previous chapter. First, the data was divided into subcategories. For each category sensitivity tests were performed. Exponential, normal and gamma distributions were tested, and the one that came closest to the real data from the sample was chosen. Below are depicted the results from fitting those distributions:

	Rush hour jam	Outside rush hour jam (weekend)	Outside rush hour jam (working days)	Accident caused jam (weekend)	Accident caused jam (working days
exponential	23386.81	310.57	1352.33	693.95	2294.31
gamma	31388.57	**219.54**	2844.60	**310.70**	749.45
normal	**21810.30**	293.19	**1294.92**	683.12	**2253.59**
sample value	20903.36	144.84	997.49	292.81	1457.63

Figure 4.1: Total duration of jams in minutes, from different jumpDist distributions

In order to try several distributions for the traffic jam durations, the jumpDist took different values:

```
jumpdist = stats.norm(mu, sqrt(sigma2)) # Normal distribution
jumpdist = stats.gamma(alpha, 1/beta) # Gamma distribution
jumpdist = stats.expon(1/lam) # Exponential distribution
```

After choosing the right distribution for each case (the ones in bold), the compound poisson process was simulated a specific number of times, in order to get an accuracy level (see the next chapter). For simulating the compound poisson process multiple times, the following code was used:
Note: the Poisson processes that correspond to those Compound Poisson Processes were also simulated, see the Appendix (A.8 - A.12 and Python code for Poisson process simulation: A.13)

```
nr = 1000
for i in range(n):
    sum_s = 0
    for i in range(nr):
        data_simulation = simulateCompoundPoissonProcess(rate, jumpdist, T)
        sum_s += data_simulation[1][-1]
        data_for_confidence.append(data_simulation[1][-1])
    data_sums.append(sum_s / big_simulations)
```

where n is the number of runs calculated in the next section.
The total duration for the case of traffic jam was computed as follows:

```
print(np.mean(data_sums))
```

where data-sums is a list of n S(t)s for the respective Compound Poisson Process.
After this, each Compound Poisson Process was run n times, and from this the mean value of S(t) of the process was calculated. In the end, in order to obtain the total duration of all traffic jams, all S(T)s of the compound processes were added together, alongside a margin of error which represents the total time duration of traffic jams caused by other things than the ones distributed. Calculated from the sample, this value is equal to 2528 minutes.The final result was:
- rush hour traffic jam durations process: S1(t) = 21807.29
- outside rush hour weekend traffic jam durations process: S2(t) = 1295.65
- outside rush hour working days traffic jam durations process: S3(t) = 170.97
- accident caused weekend traffic jam durations process: S4(t) = 240.82
- accident caused working days traffic jam durations process: S5(t) = 743.82
So, the process of the duration of all traffic jams will be equal to:

$$S(t) = S_1(t) + S_2(t) + S_3(t) + S_4(t) + S_5(t) + margin = 24258.55 + margin = 26768.55min.$$

This approximation is much better than the one obtained with method 1.

4.3 Accuracy and Number of runs

The following piece of code was used for calculating the number of runs:

```
s2 = np.var(data_for_confidence)
m = np.mean(data_for_confidence)
z = 1.96
n = (z**2) * s2 / (E * E)
```

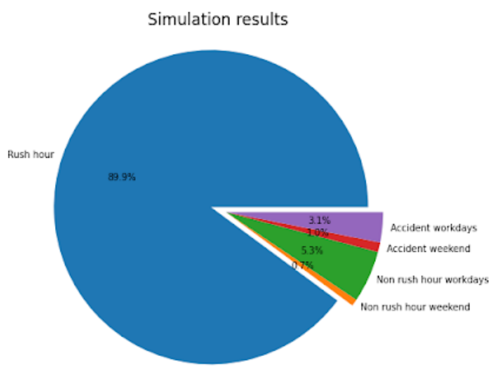
where E is the precision, and n is the minimum number of runs required to get precision E with level of confidence 0.05. Below are depicted the precisions for each compound Poisson process.

	Rush hour jam	Outside rush hour jam (weekend)	Outside rush hour jam (working days)	Accident caused jam (weekend)	Accident caused jam (working days
E in minutes	900	20	100	20	50
minimum number of runs	4	19	8	53	17
confidence interval	(20952.14, 22662.45)	(151.03, 190.91)	(1097.53, 1493.77)	(220.96, 260.68)	(694.89, 792.75)

Figure 4.2: Precisions and number of runs for each Compound Poisson Process

4.4 Results

The simulation created by using method 2 gave the following results compared to the actual data:



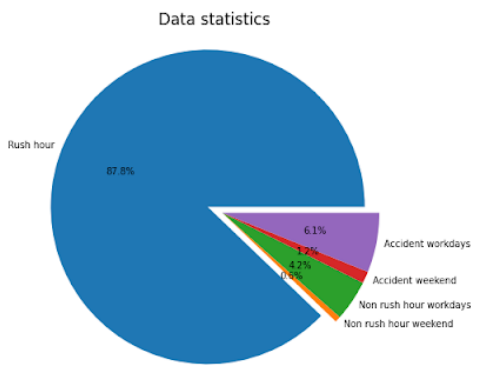


Figure 4.3: Number of traffic jams per each cause

The results depict a similarity between reality and simulated data. However, most of the causes had very little data that is hard to fit into a distribution. This led to an inaccuracy in results. Ignoring the small datasets from other causes was the best approach, since by making the model more complicated, more errors would have been introduced.

5 Conclusions and recommendations

Applied to another sample of data, this simulation won't give precise results, because of the ignored other causes of traffic jams, which occur very rarely and have causes that are beyond this report's purposes. However, the presented data results and the simulation data obtained using the second method (the one where separate Compound Poisson processes were added one to another by superposition), give a clear image on the overall situation.

Although not completely accurate, the simulation shows that most traffic jams occur during rush hours, and during the working days. This problem can be solved if drivers try to avoid the highway during those times as much as possible, or use the extra lanes. With regard to safety, weekends are much less prone to accidents, since there are no rush hours then.

A Appendix

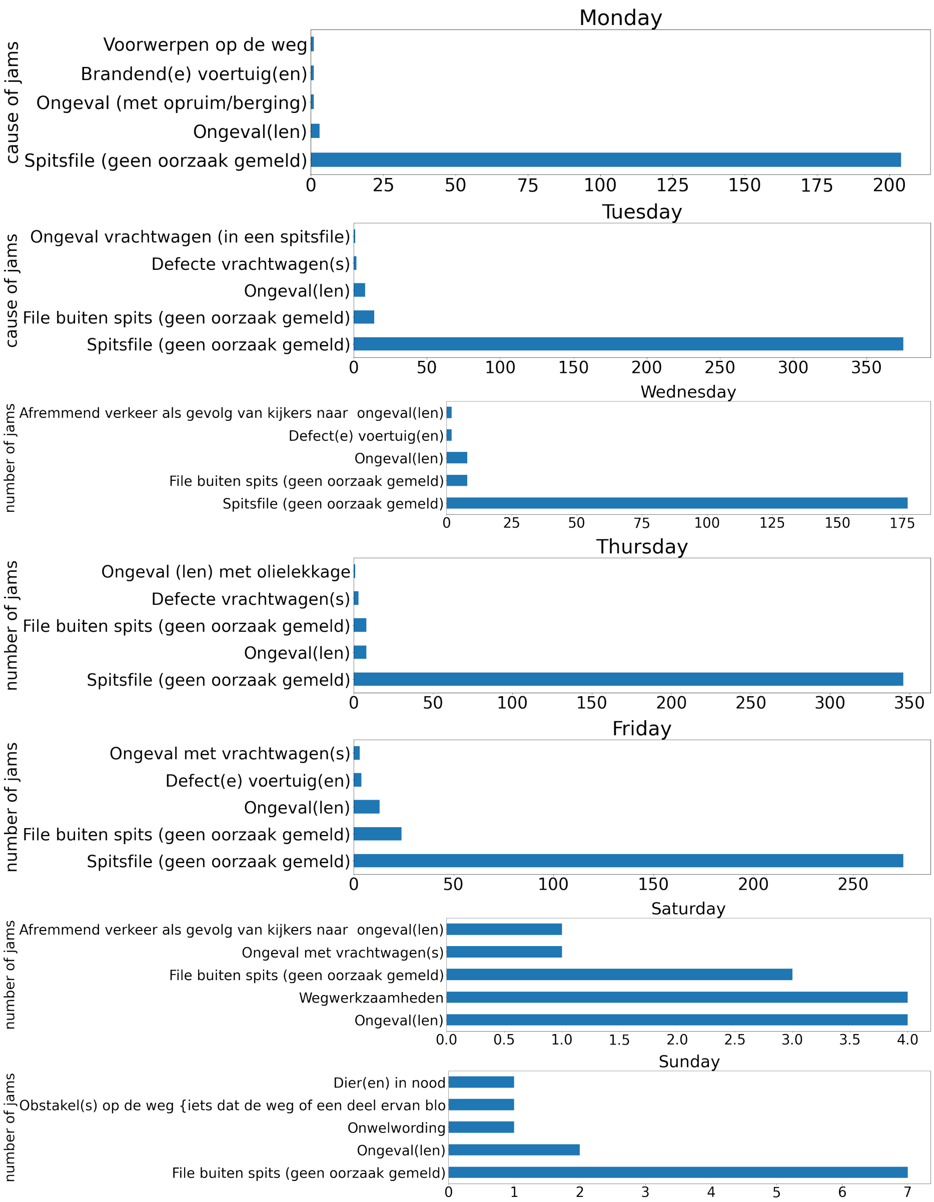


Figure A.1: Bar plots of the number of traffic jam per day

```python
def simPoissonProcess(lam, t):
    expDist = stats.expon(scale=1/lam)
    time = expDist.rvs()
    nT = 0
    while time < t:
        nT += 1
        time += expDist.rvs()
    return nT
```

Listing A.1: Python code for simulating a Poisson process

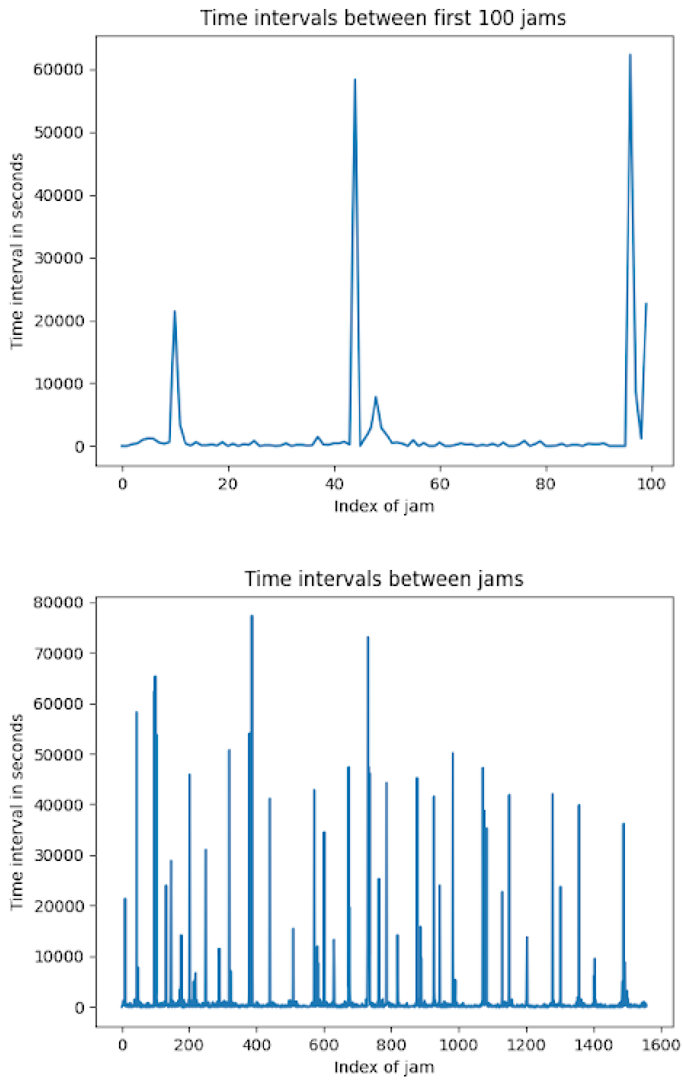


Figure A.2: Time intervals between traffic jams

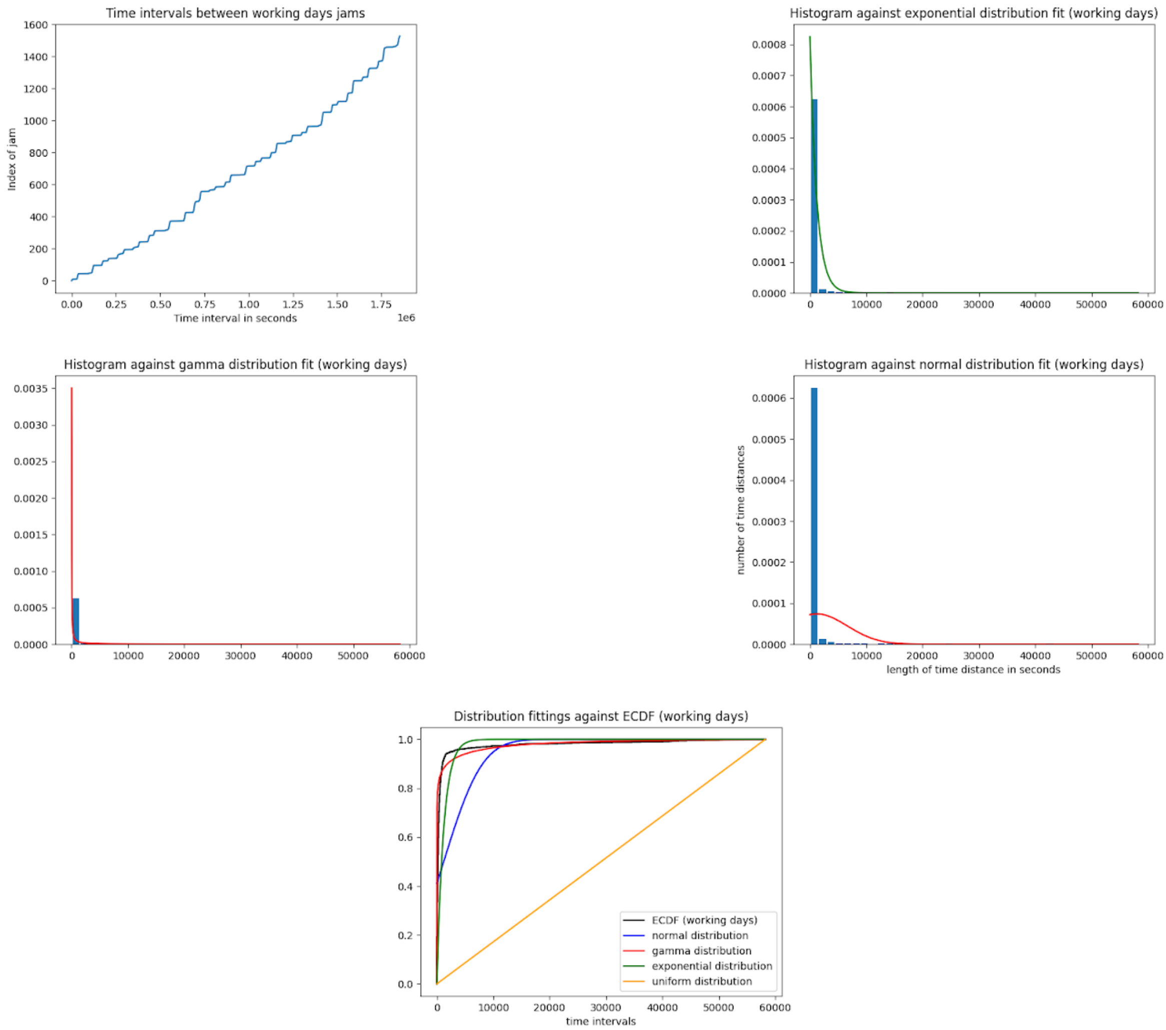


Figure A.3: Fitting different distributions for the time intervals between working days traffic jams

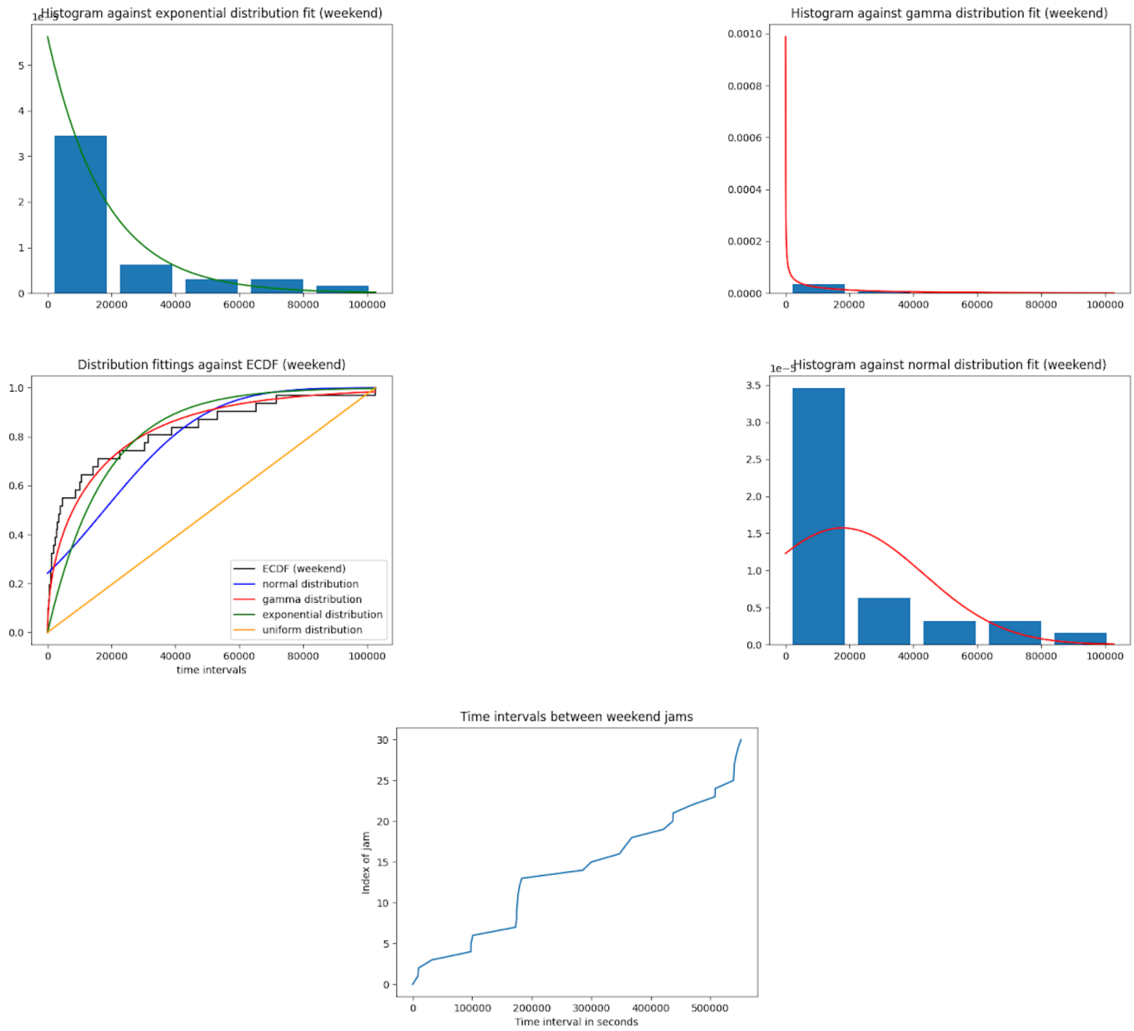


Figure A.4: Fitting different distributions for the time intervals between weekend traffic jams

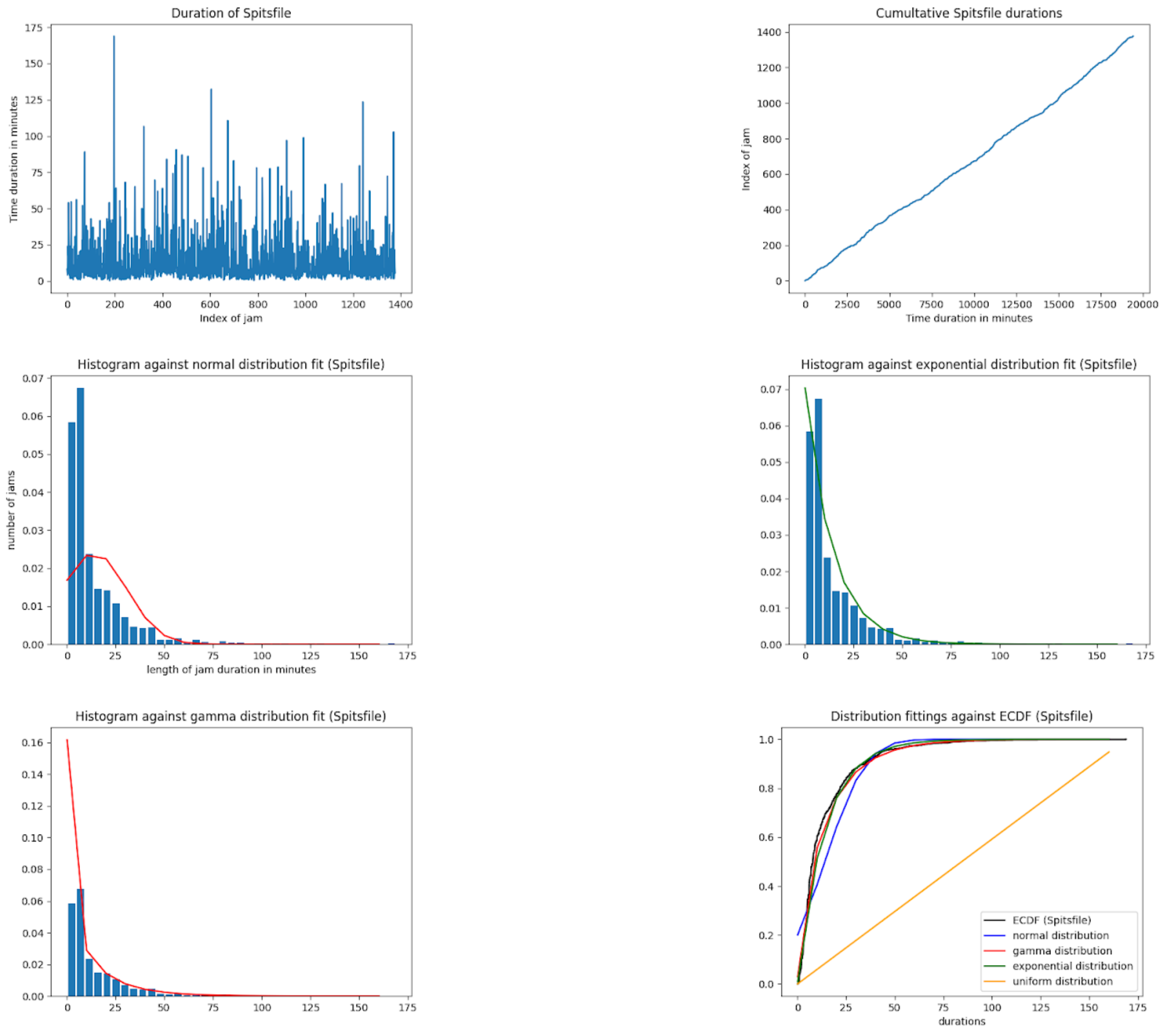


Figure A.5: Fitting distributions for time duration of Spitsfile (Rush hour traffic jams)

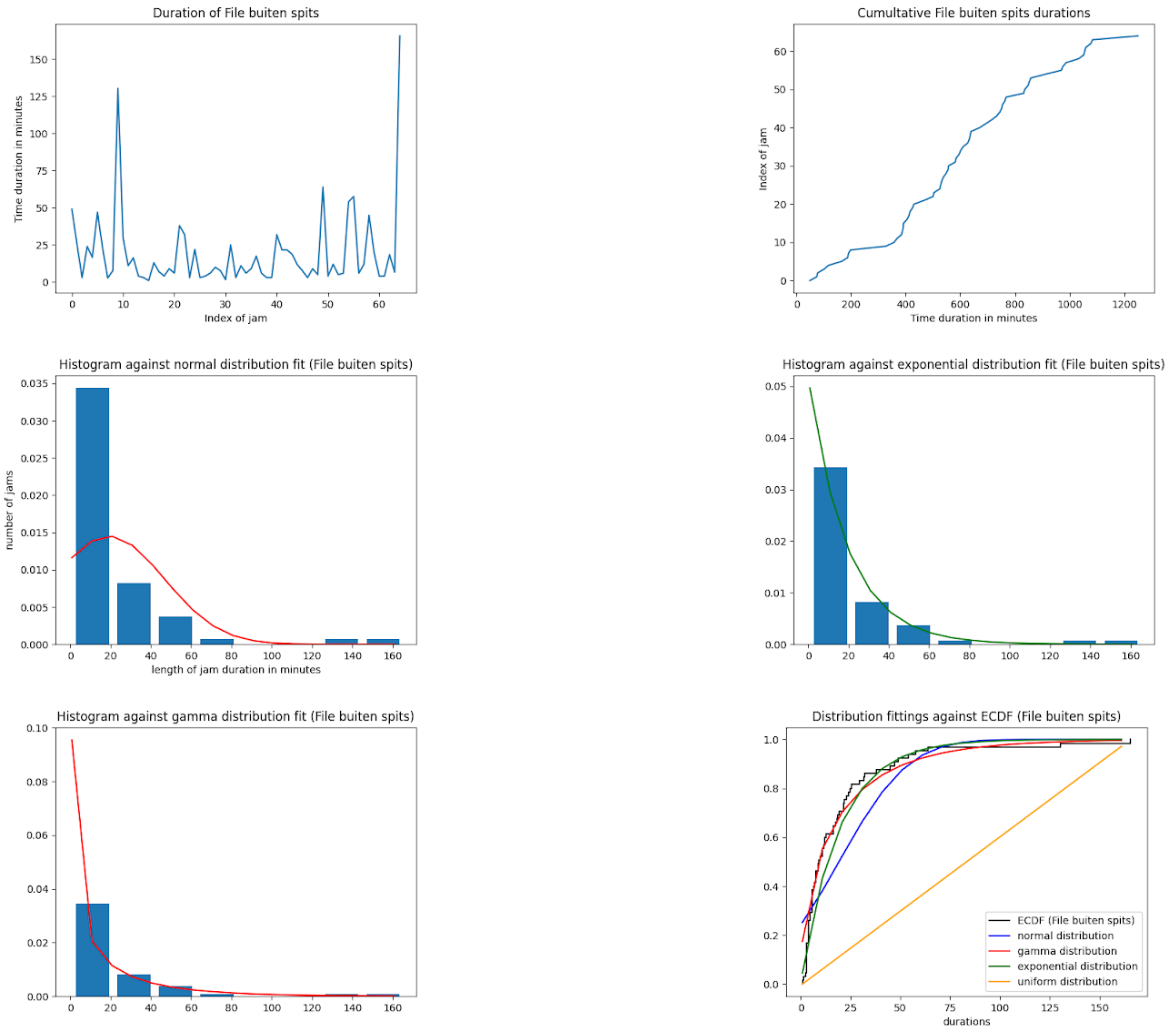


Figure A.6: Fitting distributions for time duration of File buiten spits (Traffic jams outside rush hours)

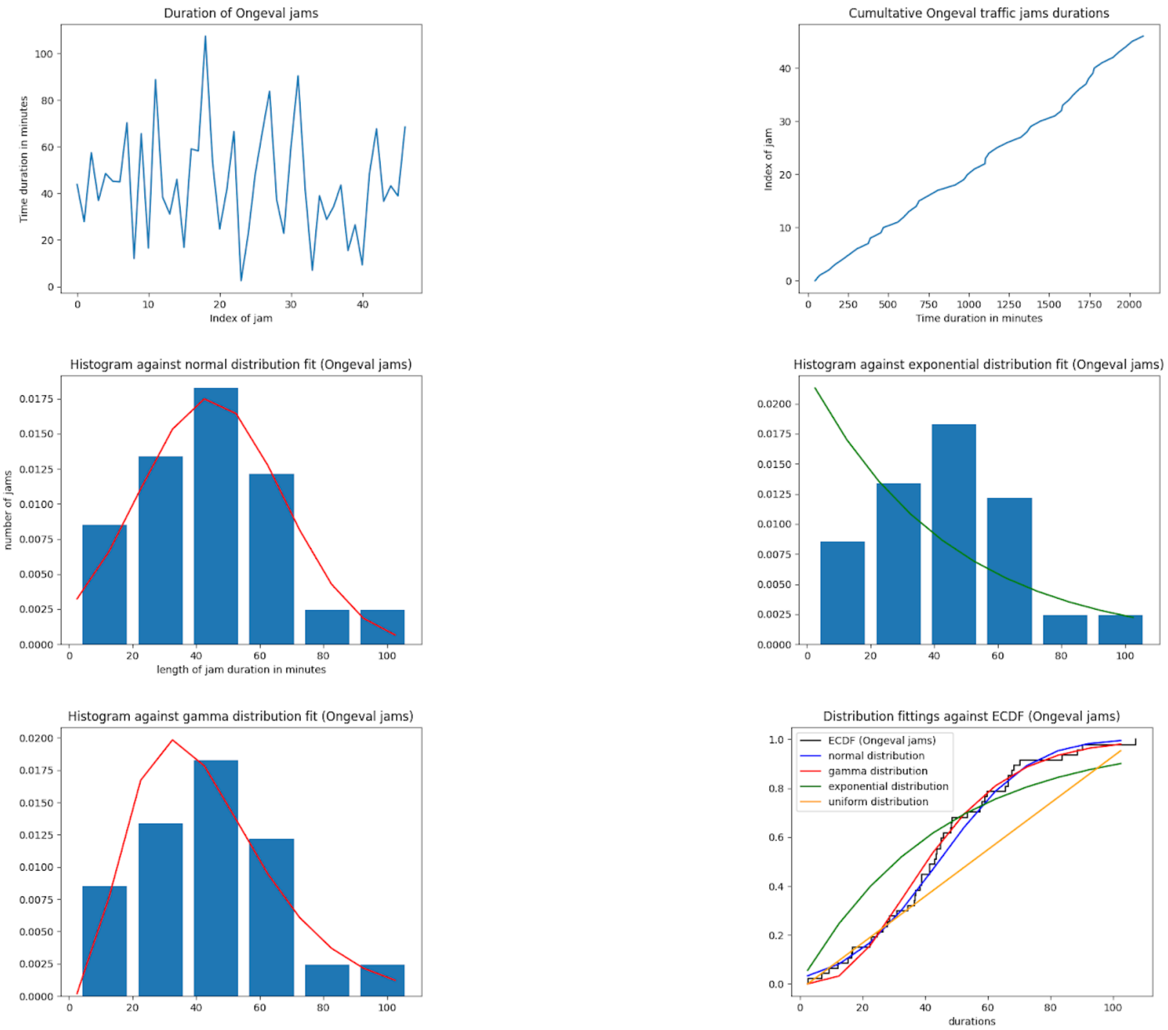


Figure A.7: Fitting distributions for time duration of Ongeval (accidents) caused jams

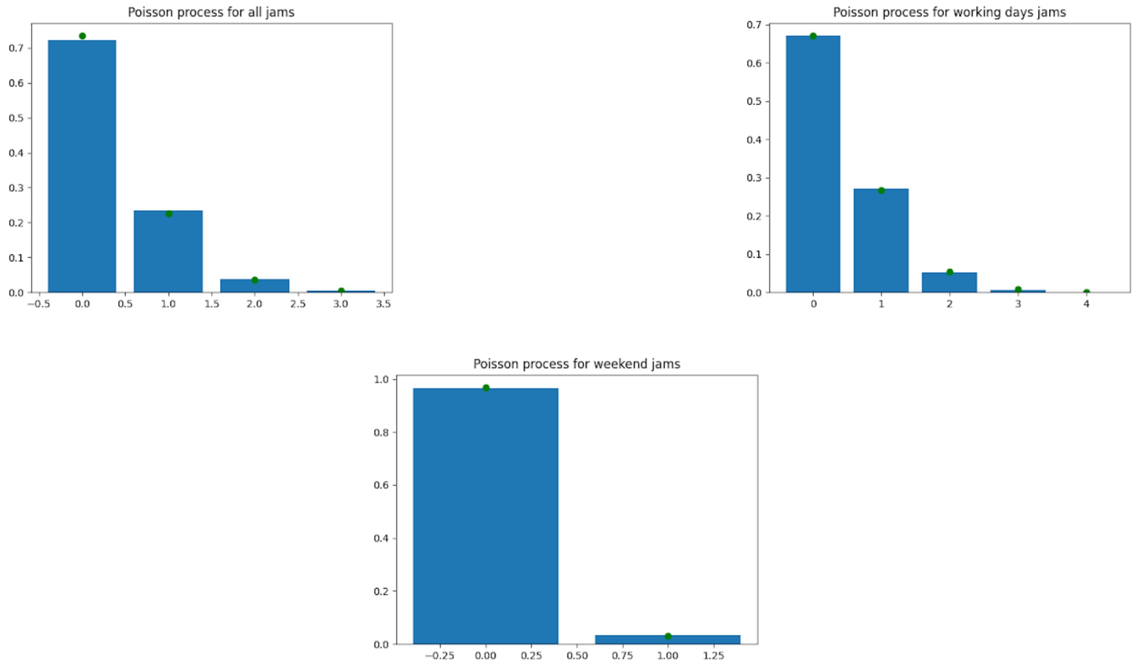


Figure A.8: Poisson processes by time of occurence of traffic jams

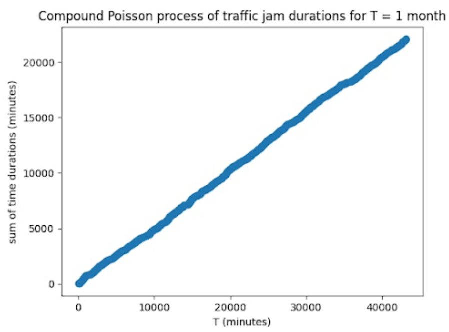


Figure A.9: Compound Poisson process of traffic jams for a month

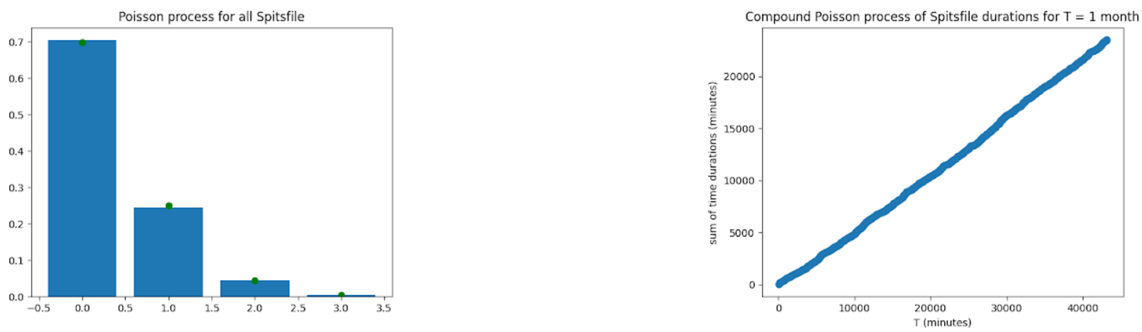
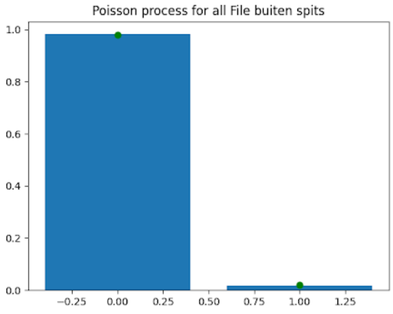


Figure A.10: Compound Poisson process for Spitsfile (Rush hour traffic jam)



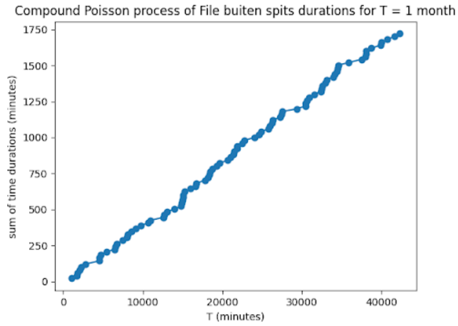
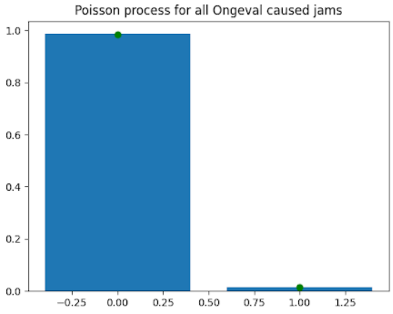


Figure A.11: Compound Poisson process for File buiten spits (Traffic jams outside rush hours)



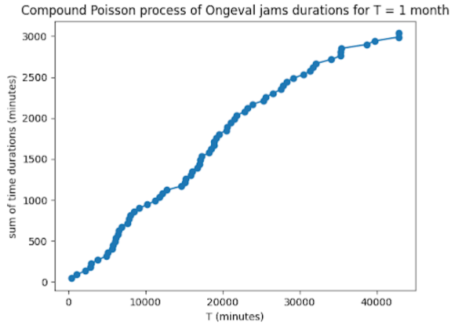


Figure A.12: Compound Poisson process for Ongeval (accidents) caused jams

	Rush hour jam	Outside rush hour jam (weekend)	Outside rush hour jam (working days)	Accident caused jam (weekend)	Accident caused jam (working days
exponential	0	0.05	3.48e-12	0.3	0.17
gamma	0	0.07	4.47e-6	0.31	0.09
normal	6.49	0.25	4.22e-6	0.58	0.06

Figure A.13: Kolmogorov-Smirnov test results

	Monday	Tuesday	Wednes day	Thursda y	Friday	Saturday	Sunday
number of jams	214	407	205	374	328	17	14
accidents	3	8	8	9	13	4	2
rush hour jams	204	375	177	346	275	-	-

Figure A.14: General data from the sample

Bibliography

[1] A2 motorway (Netherlands), January 2021. Page Version ID: 1003217502.