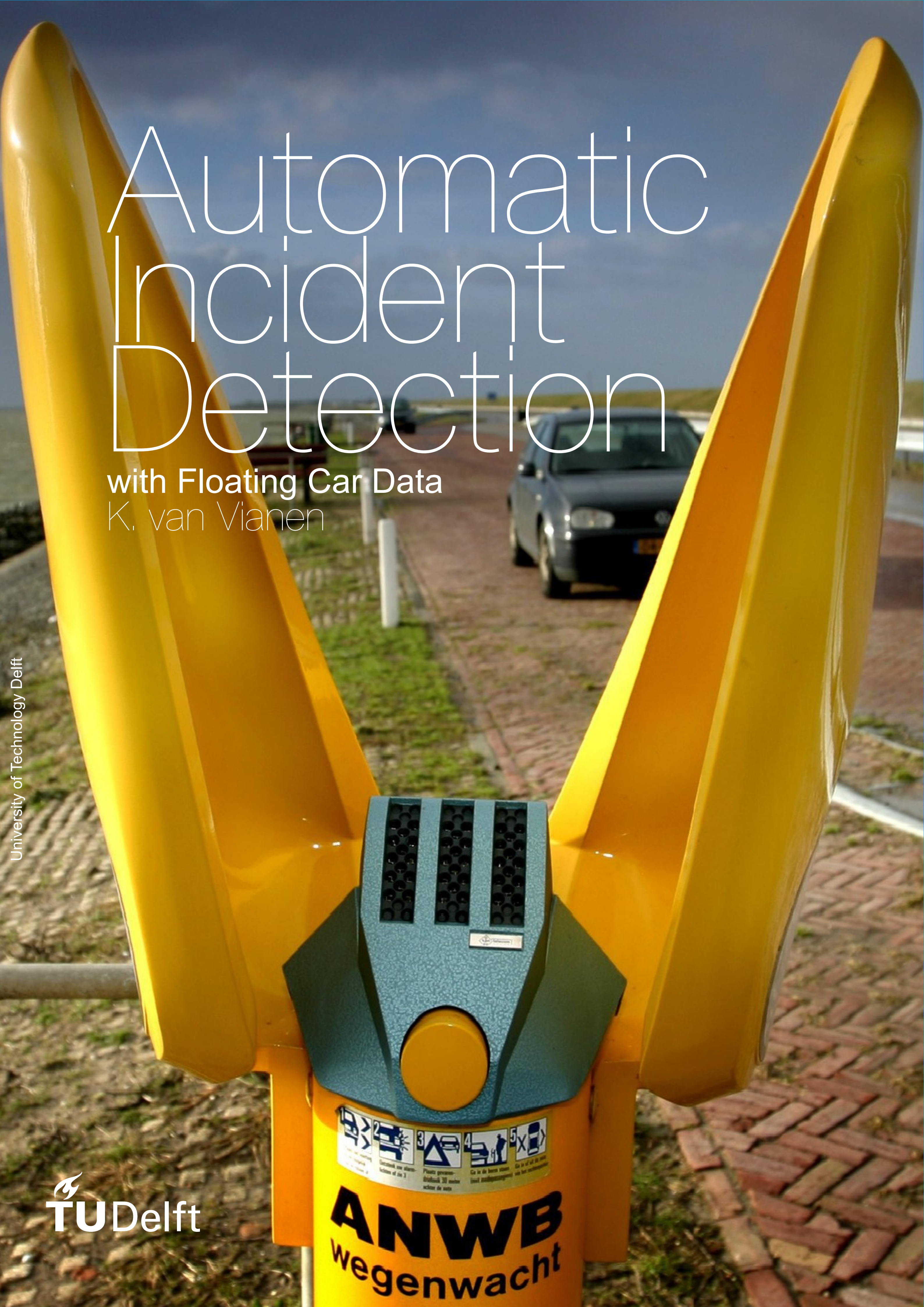
Automatic Incident Detection

with Floating Car Data

K. van Vianen

University of Technology Delft



TU Delft

Automatic Incident Detection
with Floating Car Data

by

K. van Vianen

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on 17 October 2017.

Student number:	4086708	
Thesis committee:	Prof. dr. ir. J.W.C. van Lint,	CiTG, TU Delft
	Ir. W.J. Schakel,	CiTG, TU Delft
	Prof. dr. ir. C. Vuik,	EWI, TU Delft
	Ir. Y. Dierikx - Platschorre,	Rijkswaterstaat

Delft University of Technology

Rijkswaterstaat
Ministerie van Infrastructuur en Milieu

Preface

After a bachelor in Applied Mathematics, I wondered where I really wanted to apply this knowledge. I was interested in traffic jams already, so I found an interesting application in Traffic Management. So I started the master Civil Engineering with the specialisation Transport & Planning, this thesis concludes the programme of this master at Delft University of Technology.

This thesis is made possible by Rijkswaterstaat, since I was part of the department Traffic Management Development during an internship for my thesis. I want to thank Rijkswaterstaat for the opportunities I got and especially my colleagues for helping me out, from chats with a cup of tea to giving feedback on the texts. Also thanks to my supervisors for their guidance and assistance in particular last summer.

An ode to the Dutch characteristic yellow emergency phones is on the cover of this thesis. Formerly the only way to call the emergency services in case of incidents, now they became unnecessary and are shut down.

K. van Vianen
Delft, October 2017

Summary

Congestion is still a problem on the highways, especially in times of economic growth and increasing welfare. Most of the congestion is caused by a high flow. But even 26% of the congestion is caused by unpredictable events like accidents, broken cars, lost cargo etc., also called incidents. Besides time that can be gained by early detection of this incidents, early detection is also beneficial for traffic safety.

Beside the role of incidents in the total congestion, another development is important: floating car data (FCD). It provides traffic information about individual vehicles and is increasingly available for traffic management purposes. To see how this FCD can help in detecting incidents, the following research question will be answered: *'How can floating car data be used to detect incidents?'*

Incidents Incidents is the collective name for accidents, lost cargo, broken vehicles and other issues that will block (part of) the road. In case of an incident, three main scenarios can be divided. The traffic conditions of a bottleneck congestion or other incident-like traffic conditions are quite similar to the traffic conditions of an incident. So to distinguish an incident from another cause is the main challenge for the incident algorithms. The most important characteristics of an incident:

- Head of the queue is standing still;
- Downstream of the incident location is a vacuum in terms of flow and density, along with a high speed;
- Upstream of the incident location there is a queue with high density and low speed (note: not in case of a flow that is lower than the remained capacity during the incident);
- One or more lanes of the road are blocked at the incident location, results:
  - No or less vehicles just upstream, at and just downstream the incident location;
  - More lane changes just upstream and just downstream the incident locations.

Traffic data The detection algorithms need traffic data to detect incidents, the current used data is provided by loop detectors. These loop detector data are giving information about the average flow and speed measured on a detector location. Limitation of the loop detector is that they provide discontinuous data over space, only per 500 meter on average in the Netherlands.

Floating car data (FCD) is data that is generated by or in an individual vehicle, so the car is used as a moving observer. The combined FCD of multiple vehicles represent the overall traffic conditions.This data comprises classic vehicle telemetry like speed, the direction of the vehicle and off course the position of the vehicle. Not all vehicles have the possibilities to produce FCD. The percentage of vehicles that produces FCD is the penetration rate. The consequence of a limited penetration rate FCD is that the density and flow cannot be measured directly and this comes with an uncertainty in flow and density calculations and thus an uncertainty in the measured traffic state.

Incident detection algorithms Based on the traffic characteristics of incidents and the characteristics of the available traffic data, some requirements for incident detection algorithms are determined:

1. Detect all incidents, not missing an incident;
2. Not giving a signal when there is no incident, a false alarm;
3. Quick detection of incidents.

Three existing algorithms are researched. First the Blokkadedetector, which aims to recognise incidents to search for differences in flow. The seconds algorithm is the Presikhaaf algorithm, which is comparing speeds to detect incidents. The McMaster algorithm is the third and last existing algorithm that is researched. This algorithm defines four states in the occupancy-flow graph, to recognise

incidents. In terms of false alarm rate, detection rate and detection time the McMaster algorithm is classified as the best algorithm of these three. So the performance of this algorithm will be compared to the performance of the new designed algorithm.

New algorithm The new algorithm is based on the fact that vehicles have to change lanes in case of an incident, since the road is partly blocked. So the number of lane changes per minute of a road section is compared to a situation without an incident on the same road section. The information about lane changing can be provided by floating car data.
Note that for every road section of 200 meter in length, the number of lane changes is measured separate since the road lay-out will heavily influence the number of lane changes (lane drops, on and off ramps etc.).

For clarification an example is given in figure 1. In this figure the dots are the historical data without an incident and the stars are data with an incident. The density-axis is divided in strips of 10 veh/km and a threshold is determined for each strip.

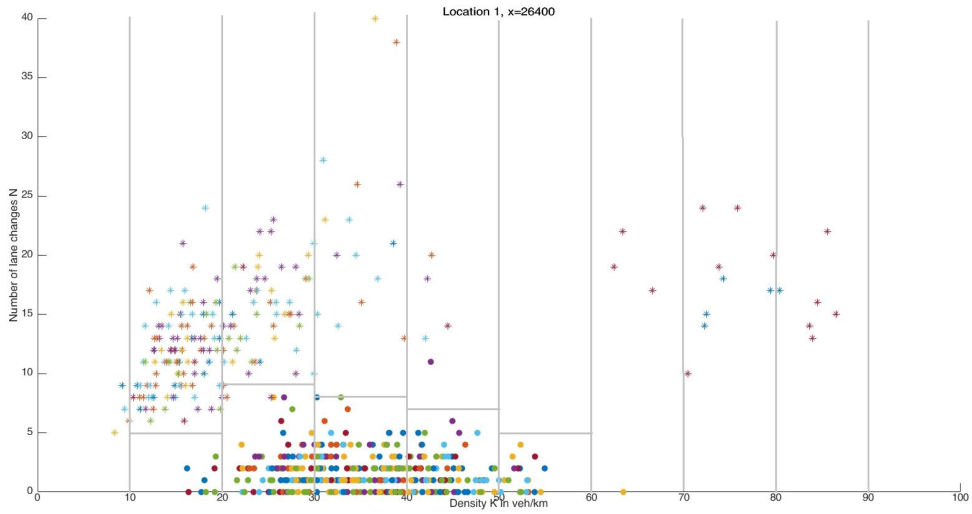


Figure 1: Example of results with an incident; dots are scenario without incident, stars of scenario with incident. Vertical grey lines display the intervals in K for which the thresholds are made, the horizontal grey lines are the thresholds based on historical data without incident.

Same procedure can be followed for a limited penetration rate of floating car data.

Evaluation of the algorithms The McMaster algorithm and the new algorithm are evaluated using a simulation study. In this simulation study four different scenarios are evaluated. These scenarios corresponds with four locations: a simple road stretch and a road section with a lane drop, both contains a scenario with an incident on the left lane and a scenario with an incident on the right lane. In these scenarios different variations are made:

- Incident location. Position relative to the loop detector.
- Traffic conditions. Free flow conditions on the simple road stretch and congested conditions on the road stretch with lane drop.
- Penetration rate of the floating car data: 100%, 50%, 20% and 10%.

The performance of the algorithms is judged by using three key performance indicators:

- The detection rate D is the amount of correct detected incidents I_{det} in proportion to the total incidents I_{tot} that took place. Note that I_{total} is the sum of correct detected incidents I_{det} and not (correct) detected incidents;

$$D = \frac{I_{\text{det}}}{I_{\text{total}}} \tag{1}$$

- The false alarm rate F is the amount of times that the algorithm indicates an incident when there is not in relation to the total data points;
- The detection time t_{det} is the time it take for the algorithm to give the incident message since the real incident.

Incident detection by the new algorithm turns out to be very well possible. Especially if lane change information can be gathered from all vehicles, then the new algorithm is performing better than the McMaster algorithm for all locations. Even a perfect score of 1.0 on the detection rate is achieved.

Also the performance of 50%, 20% and 10% FCD are tested to see the added value of an increased penetration rate.

The new algorithm is performing better than the McMaster algorithm for the simple road lay out, even with a limited penetration rate. The results for 10% FCD and the McMaster algorithm are comparable in terms of false alarm rate, detection rate and detection time. Adding more vehicles that can provide FCD, mainly the false alarm rate and the detection time will decrease.

For the second road layout with the lane drop, the McMaster algorithm is performing badly. Especially the low detection rate is a huge problem. The new algorithm is also having some troubles, but can detect most incidents correctly, although it takes a longer detection time than for the McMaster algorithm when it does detect an incident.

For both locations the new algorithm with 10% FCD can detect incidents pretty well, although a lower false alarm rate is desirable.

In general the new algorithm is better in detecting the tested incidents than the McMaster algorithm. As expected the McMaster algorithm does have more problems when there was already congestion at the incident location and time. Also the lane drop can have an effect on the performance of the new algorithm. The expectation was indeed that the new algorithm would have some difficulties at this location, happily it did detect all incidents.

A major drawback of the research is the simulation of lane changes in the simulation program. This program is literally programming vehicles to change lanes in a certain area before an road blockade and in the new algorithm you are measuring that they are changing lanes. This is highly influencing the performance of the new algorithm. The results of this research can therefor only be seen as an first indication that incidents can be detected by comparing the number of lane changes to data without incidents.

So incident detection is possible on lane level by comparing the number of lane changes for a situation without an incident and a situation with a possible incident. Another conclusion is that floating car data can be used to gather this lane change information and in this way can be used for incident detection.

Contents

1

Introduction

Can incidents be detected? How is that currently done? How can an incident be recognised? What is the influence of quicker detection of incidents? Will the total congestion decrease in case of quicker detection? What is floating car data? What are the weaknesses and chances in incident detection?

A lot of questions, a lot of answers to give. First a short introduction is given with the relevance of this research, the scope, research questions and the reading guide, before really kicking off.

Social and scientific relevance Congestion is a commonly discussed subject in the daily journals, the newspapers and on the internet. Congestion decreases the accessibility of cities and activities, and is also an obstacle for economic growth. Especially during the last three years the congestion severity is increasing, as can be seen in figure 1.1 , which presents the congestion severity in the Netherlands. The expectation is that this growth will continue, due to factors such as economic growth and increasing welfare, among others.

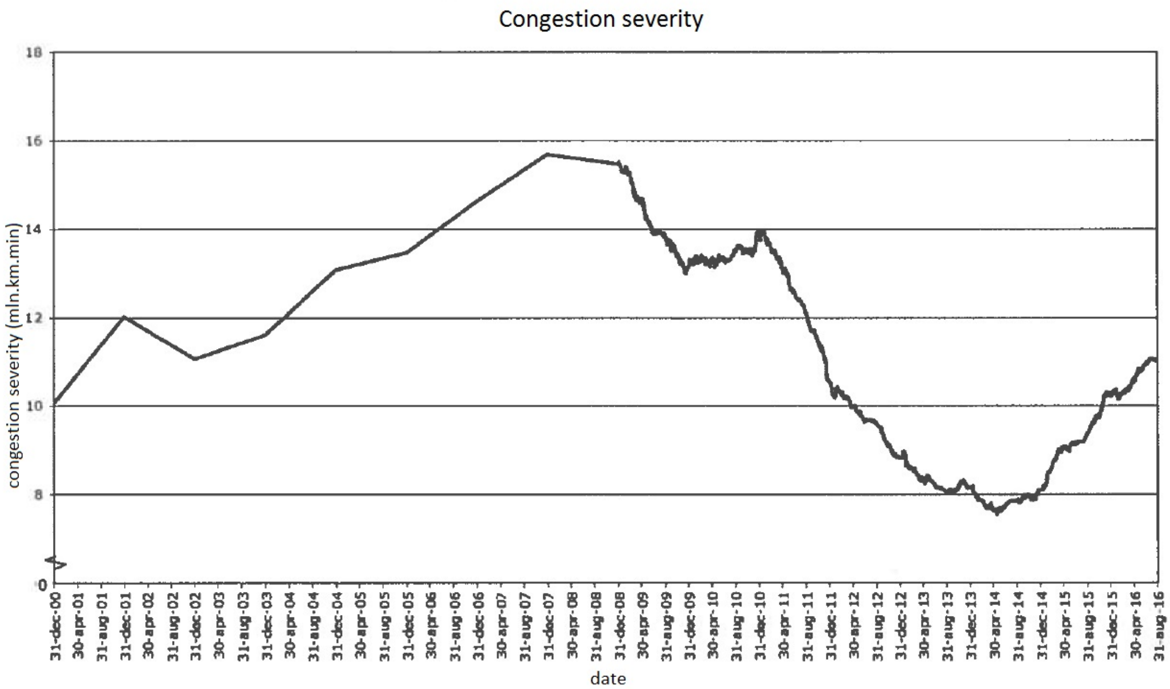


Figure 1.1: Development of the congestion severity from December 2000 till September 2016 [11]

Besides the increasing congestion, another interesting development is a new data source; floating car data (FCD). It provides traffic information about individual vehicles and is increasingly available for traffic management purposes. This is interesting for road authorities, because FCD provides a cheaper alternative to the currently used loop detectors, which are expensive in maintenance and purchase. Another reason for striving to be independent of loop detectors are the regularity of faults in the data they provide. Finally, FCD is also interesting for scientific purposes, like understanding the individual vehicle behaviour.

These two developments are the main reason for this thesis. Furthermore it appears that the current incident detection (AID) algorithms are performing not good according to literature. The detection of incidents is difficult with the current techniques and available data. This results in slow detection, frequent false alarms and/or a low rate of detected incidents.

Scoping Earlier detection of incidents will limit the congestion as results of the incident. A common used rule of thumb is that 3 minutes of total congestion can be saved for each minute that the road is cleared sooner. Besides the time gained, this is also beneficial for traffic safety. Traffic management measures can be taken to prevent (secondary) accidents and thereby more nuisance for road users.

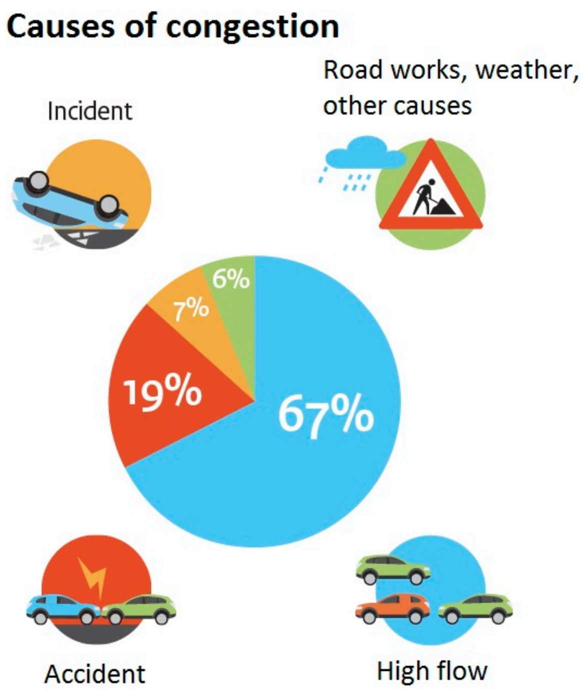


Figure 1.2: Causes of congestion in the Netherlands [11]

A yearly research in the Netherlands [11] has shown that in 26% of the cases the cause of congestion is an accident or an incident, see figure 1.2. In this case an incident is also a blockade of the road by a broken car or lost cargo for example. Therefore by focusing on the detection of incidents an important part of the congestion problem can be addressed.

The scope of this thesis is thereby the detection of incidents with floating car data on the traffic management side of the problem. As such this thesis shall not cover technical specifications of the floating car data, the gathering of this data and the implementation in the current structure of road side systems.

Research questions In this thesis a new detection algorithm will be designed to detect incidents, this algorithm will use floating car data. This new algorithm will be evaluated using a simulation study.

The research question is therefore defined as follows:

How can floating car data be used to detect incidents?

The following subquestions will help to answer the main research question:

1. How can an incident be recognised?

2. What information can loop data and floating car data provide?
3. What are the requirements of the algorithms to detect an incident?
4. Which current detection algorithms are available and what is their performance according to the literature?
5. How can a new algorithm be designed to detect incidents using floating car data?
6. How can incident detection algorithms be evaluated?
7. What is the performance of the new designed algorithm compared to existing algorithms?

Reading guide The first four subquestions will be answered in the following chapter 'Incidents, traffic data and detection in literature'. After more detailed information about incidents, loop detector data and floating car data, a few criteria are discussed for proper incident detection, after which a few existing algorithms will be compared using these criteria.

The new algorithm is designed using the lessons of the previous chapter and the design is elaborated in chapter 3 'New incident detection algorithm development using floating car data' and thus subquestion 5 will be answered.

In the fourth chapter 'Research methodology for evaluation of incident detection algorithms' the sixth subquestion will be answered about the reasons for choosing simulation to evaluate the algorithm and the set-up of the simulation.

The last subquestion about the performance of the new designed algorithm will be discussed in chapter 5 'Results of incident detection algorithms' performance'.

Before the conclusion, a critical look on the results is taken in chapter 6. The thesis is finished with answering the main research question in chapter 7 'Conclusions', together with a critical view on the results in the discussion and some recommendations for further research.

The structure of this thesis is also displayed in figure 1.3.

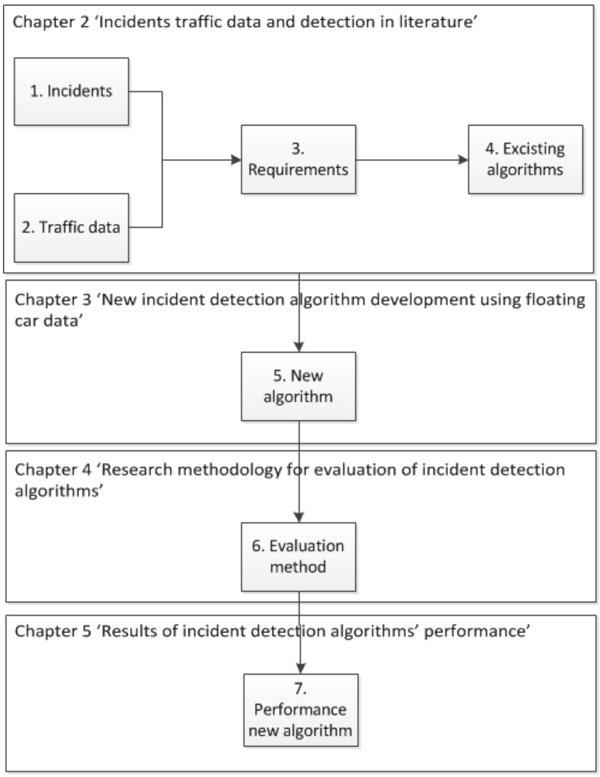


Figure 1.3: Reading guide

2
Incidents, traffic data and incident detection in the literature

To design a new algorithm that will detect incidents using floating car data and thus to answer the main research question, more information is needed about incidents and traffic data. Therefore the characteristics of incidents will be discussed in this chapter, after which the features of floating car data and loop data will be discussed. Knowing the characteristics of incidents and the data, some criteria are constructed for detection algorithms. To conclude this chapter, a few existing algorithms are described and checked by these criteria.

2.1. Incidents

By Gall and Hall, an incident is defined as "a random event that may disrupt the orderly flow of freeway traffic" [3].Incidents are unpredictable by nature, like the breakdown of vehicles, crashes and lost cargo. These incidents can disturb the traffic flow and be a cause of congestion at unpredictable locations and times. A critical note on the definition of Gall and Hall is that also unexpected driver behaviour can disrupt the flow and is not an incident, in this report this definition is expanded with the notion that an incident will (partly) block the road and thus disturb the traffic flow.

Early detection of incidents is very important to limit the consequences like delay for road users, higher speed, less costs, less time commitment for emergency services, but also to prevent (secondary) accidents.

2.1.1. Traffic characteristics incident

To recognise and detect an incident, it is important to know what traffic characteristics can be expected in case of an incident.

In figure 2.1 different scenarios are displayed in case of an incident. In scenario *a*, the road is completely blocked due to an incident. The consequences will be clearly visible. No vehicles will be measured just downstream of the location of the incident and the flow will be zero, also no speed is measured. In scenario *b*, the rest capacity after the incident is lower than the demand. In this case the upstream measured flow is decreased in comparison to the moment before the incident happened. Upstream a standing queue will occur with low speed and high density. Downstream the free-flow conditions hold. In scenario *c*, the incident is more difficult to detect, since the road is only partly blocked and the road offers enough capacity for the demand of vehicles, the traffic flow variables does not show big changes. Although the incident will not cause a queue in this case, it is important to detect it for example to prevent secondary incidents and to clean the incident before the flow is increased till more than $Q_{restcapacity}$.

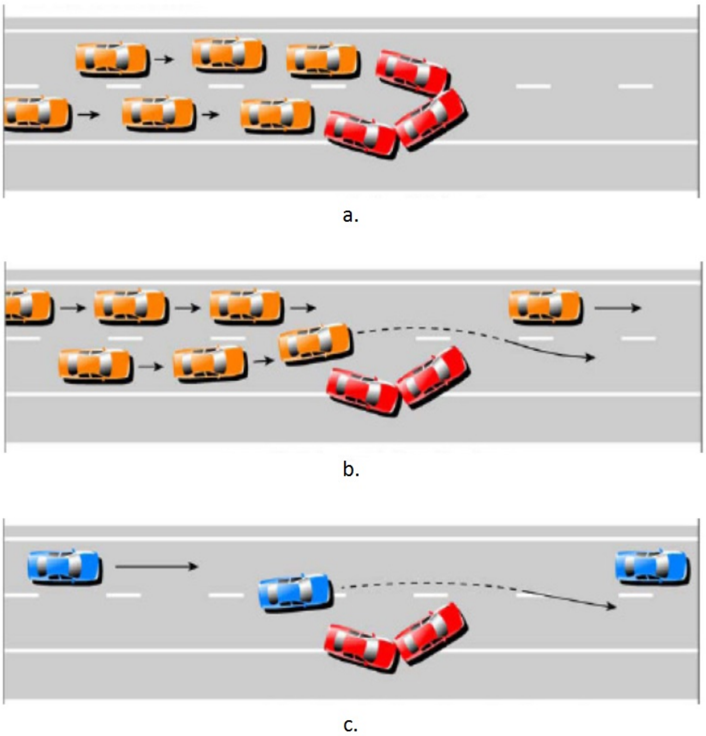


Figure 2.1: Traffic characteristics in case of an incident; a. Road totally blocked; b. Road partly blocked, high flow; c. Road partly blocked, low flow [15]

A traffic jam that is caused by an incident is a standing queue, the head of the queue is standing still at the location of the incident. The tail of the queue moves upstream if the demand is higher than the offered capacity. In case of a partly blocked road, the queue can dissolve if the demand is lower than the offered capacity. Solving the congestion will always be accelerated if the incident is cleaned by emergency services or road inspectors.

2.1.2. Traffic flow analysis

Based on traffic flow theory an analysis of the traffic states is made for the situations that are sketched in figure 2.1. The specific used theory is shock wave theory to capture queueing dynamics [5].

A triangular fundamental diagram is chosen to describe the relationship between flow and density. The different traffic states are indicated with red numbers in figure 2.2.

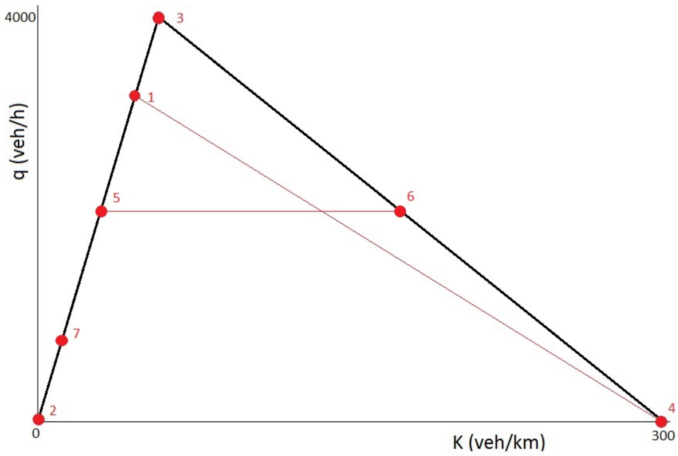


Figure 2.2: Fundamental diagram 2 lanes

The space-time diagrams of a total blocked road and a partly blocked road are displayed in figure 2.3. The traffic flow before the incident is 3200 veh/h, indicated as state no. 1. In case of a total blocked road, the flow downstream the incident location is zero, indicated as traffic state no. 2. This zero flow in combination with a high flow upstream is a clear indication for a total blockage of the road.

When one lane is blocked, the flow downstream will be equal to the flow on one lane, 2000 veh/h, indicated by traffic state no. 5. The duration of the congestion will last shorter, visible in figure 2.3. Still a difference in flow is visible, upstream and downstream of the incident location. But in this situation the difference in flow is less and therefor more difficult to distinguish from congestion with another cause.

The last situation that was sketched in figure 2.1 is a partly blocked road and a low flow. This flow is denoted by no. 7 in figure 2.2. In this case shock wave theory gives no other flows upstream or downstream. So this type of incident and a low flow cannot be detected by the quantities flow and density. Note that this also means that this type cannot be detected by measuring the quantity speed.

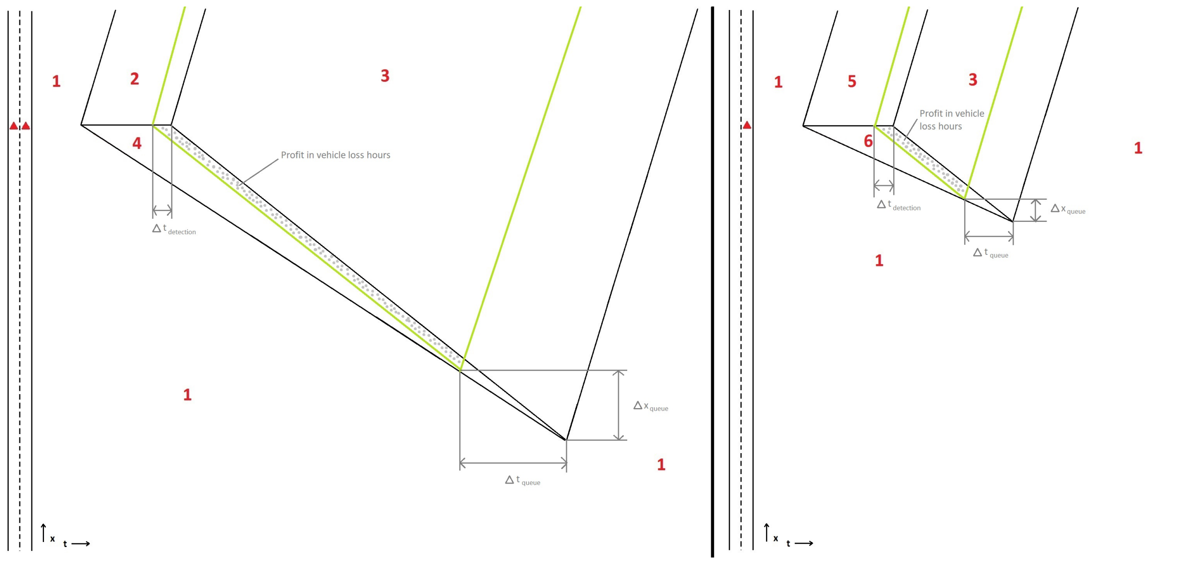


Figure 2.3: Shock waves in space-time diagram; left a blockade of both lanes, right a blockade of one lane. In green the shock waves in case of solving the incident earlier. The red rectangles indicate the incident location.

An earlier detection will lead to a smaller surface of respectively the area of state 4 in the left part of figure 2.3 and area of state 6 in the right part. This means less vehicles that suffer from congestion and also less vehicle loss hours. The green lines in figure 2.3 represents the shock waves in case of solving the incident earlier, made possible by an early detection. Besides a smaller surface of respectively state 4 and state 6, also the surface of the area with state 3 is smaller and thus less vehicle loss hours. Rule of thumb in traffic management is a profit of 3 minutes congestion for each minute that the incident is solved earlier.

The traffic conditions of a bottleneck congestion or other incident-like traffic conditions are quite similar to the traffic conditions of an incident. The main characteristics of an incident:

- Head of the queue is standing still;
- Downstream of the incident location is a vacuum in therms of flow and density, along with a high speed;
- Upstream of the incident location there is a queue with high density and low speed (note: not in case of a flow that is lower than the remained capacity during the incident);
- One or more lanes of the road are blocked at the incident location, results:
  - No or less vehicles just upstream, at and just downstream of the incident location;
  - More lane changes just upstream and just downstream of the incident locations.

Note that especially the last three characteristics are specific for a totally blocked road or a partly blocked road and a flow higher than the left capacity (see scenario a and b in figure 2.1). In case of a partly blocked road and a flow lower than the left capacity, the vacuum and the congestion will not occur and the number of lane changes will change less [16].

2.2. Traffic data

An upcoming source of traffic data is floating car data. Currently loop detector data is commonly used to gain knowledge about traffic states. Both data sources will be discussed and elaborated.

2.2.1. Floating car data

Floating car data (FCD) is data that is generated by or in an individual vehicle, so the car is used as a moving observer. The combined FCD of multiple vehicles represent the overall traffic conditions. This data comprises classic vehicle telemetry like speed, the direction of the vehicle and the position of the vehicle [10]. The vehicle trajectories of the cars that generate FCD can been visualised in a space-time diagram like figure 2.4. In this graph, each curve represents the trajectory of an individual vehicle. Together with the speed-time diagram the data will give an indication of the traffic state on the measured road.

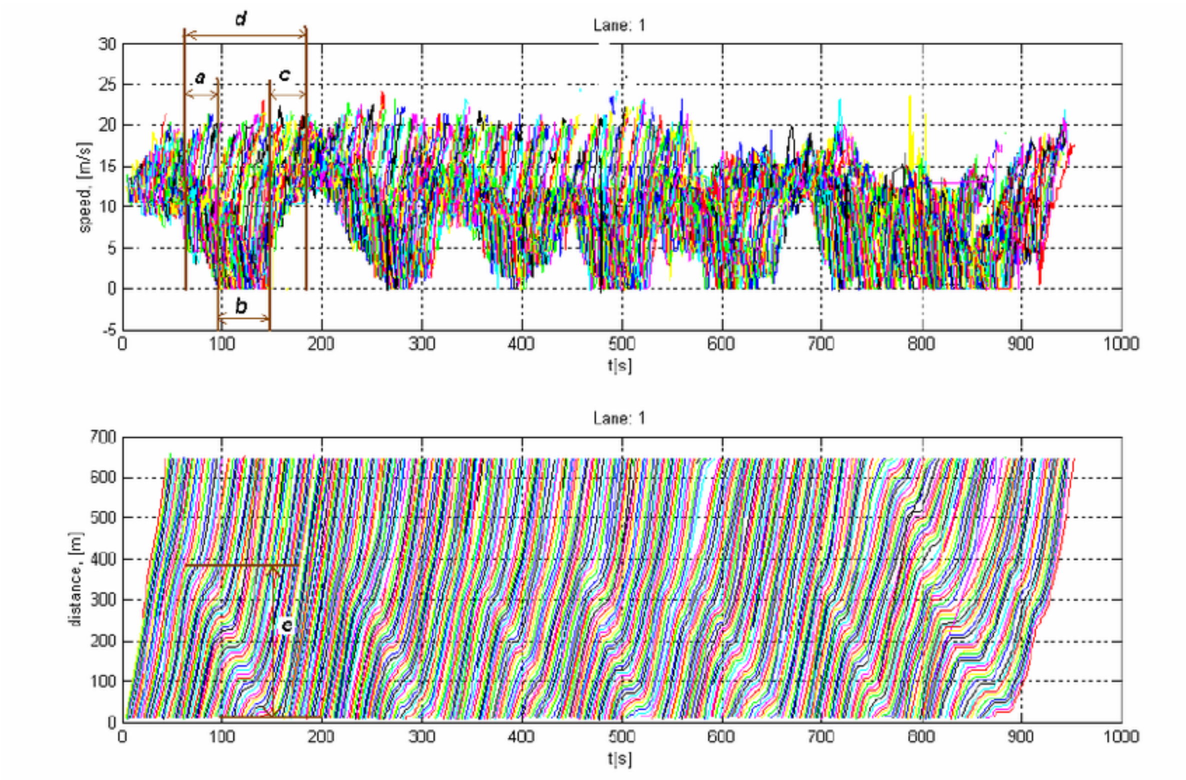


Figure 2.4: Speed-time diagram at the top and vehicle trajectories in space-time diagram at the bottom [6]

Relating the vehicle trajectories and speed information to shock wave theory, an analysis of the traffic state can be made. The time referred to as *a*, is the time when the congestion is building up, time *b* is the congested static state and the recovery of the congestion to free flow state is time *c*. The total duration of the congestion is the sum of these three, also referred to as time *d*. The congestion occurred over a distance *e* in the distance-time diagram [6].

In the distance-time figure the surface can give information about the density, but only when all trajectories are available. Current issue with FCD is that not all vehicles deliver this kind of data. FCD can be gathered by separate devices that are taken into the vehicle, like a mobile phone or by in-car devices such as a navigation device. Not all vehicles have the possibilities to produce FCD. The percentage of vehicles that produces FCD is the penetration rate [4]. An increasing number of vehicles, both private and commercial (trucks for example), are equipped with devices that can produce FCD.

The current penetration rate FCD on a highway in the Netherlands is between 5% and 10%, depending on location and time of the day [12]. The consequence of a limited penetration rate FCD is that the density and flow cannot be measured directly and this comes with an uncertainty in flow and density calculations and thus an uncertainty in the measured traffic state. Since only an unknown percentage of the passing vehicles provides FCD, the exact flow and thus density cannot be measured. A possible solution is to measure the flow with loop detectors and use it in the calculations for penetration rate and density.

Another practical limitation of FCD is the frequency of this data. Depending on the measuring method and the data servers, the FCD is saved in a certain frequency and not immediately sent to the server. So for example the data is saved per 20 seconds and sent in a frequency of one per 60 seconds. Currently some private parties are experimenting with FCD and are developing methods that can give FCD in high frequencies. Some parties have already succeeded to deliver second based FCD [2]. However little is known about the exact quality of this FCD, so there is more research needed by independent institutes.

The accuracy of the FCD in terms of location will improve in the coming years, the accuracy of a few meters will increase first to decimetres and eventually also to centimetres. To recognise behaviour like lane changes, an accuracy of about 1 meter is needed, since the width of the lanes is 3.5 meter for highways in the Netherlands. Although this is still an issue, the assumption is made that the FCD is this accurate and will indicate lane changing of vehicles.

2.2.2. Loop detector data

Currently loop detectors are a common used source to collect data of traffic flows. These loop detectors, also called induction loops, are milled in the pavement of roads, usually on each lane. These loop detectors can give information about the traffic flow on highways, but are also used for queue detection at controlled intersections in urban areas. On highways, usually a lateral spacing between 300 and 1000 meters is used with an average spacing around 500 meters.

A limitation of the loop detector is that it gives discontinuous data over space. Only per 500 meters on average a loop detector gives information about the traffic flow. But on some highways in the Netherlands, there are no variable message signs and detectors. This is for example the case on the A1 between Barneveld and Apeldoorn in both directions, a road stretch of about 30 kilometres long. Another disadvantage is missing or otherwise unreliable data, on average 5-10% with extremes up to 30% unreliable data [14].

Two types of induction loops exist: single and double loops. Single loops measure occupation time and count vehicles. The occupancy of a detector is the fraction of time that the detector is occupied. The occupancy consists of the time that the detector is occupied, τ_{occupied}, and the time that the detector is not occupied, $\tau_{\text{not occupied}}$.

$$o = \frac{\tau_{\text{occupied}}}{\tau_{\text{occupied}} + \tau_{\text{not occupied}}} \tag{2.1}$$

The occupation time can be calculated from the speed v, the length of the vehicle L_i and the length of the detector L_{det}. The distance that the vehicle has to cover from the moment it arrives at the detector up to the time it totally passed it, is equal to the length of the detector plus the length of the vehicle itself. Hence, the occupancy time equals

$$\tau_{\text{occupied}} = \frac{L_i + L_{\text{det}}}{v} \tag{2.2}$$

The distance that the following vehicle has to travel to reach the detector is the gap between the leading and the following vehicle minus the length of the detector. So the time this takes, the time that the detector is not occupied is:

$$\tau_{\text{not occupied}} = \frac{s - L_i - L_{\text{det}}}{v} \tag{2.3}$$

Substitution of the occupation and not occupation time in equation 2.1 gives:

$$o = \frac{L_i + L_{\text{det}}}{s} \tag{2.4}$$

Usually the detector length is known for a certain road configuration and the occupancy is measured. When assuming the vehicle length, the space headway can be calculated from equation 2.4. Since the density can be calculated from the space headway, this can also be done when assuming the vehicle length.

$$k = \frac{1}{\langle s \rangle} = \frac{1}{\left\langle \frac{L_i + L_{\text{det}}}{o} \right\rangle} \tag{2.5}$$

So the density is calculated with a certain fault depending on the fault made in the vehicle length [5].

Double loops consists of a pair of loops per measurement location, the two loops are placed with a known distance of Δx between each other on the same lane. The speed of the passing vehicle can now be calculated since the time of passing loop 1 t_1 and the time of passing loop 2 t_2 is measured.

$$v = \frac{t_2 - t_1}{\Delta x} \tag{2.6}$$

Since the time headway between two vehicles can be measured as well, the space headway can be calculated and the density as well.

$$k = \frac{1}{\langle s \rangle} = \frac{1}{\langle hv \rangle} \tag{2.7}$$

Second advantage of double loops is that they can measure the vehicle length. The vehicle category is determined based on these measured vehicle lengths.

So double loops give a better measurement for the density and therefor monitor the roads better. In the Netherlands the roads are equipped with double loop detectors.

2.2.3. Comparing floating car data and loop detector data

The biggest difference between floating car data (FCD) and loop detector data is that FCD gives information about a limited amount of vehicles and a loop detector measures all vehicles at one specific location. The biggest disadvantage of loop detector data is that it is discontinuous in space, the disadvantage of FCD is the limited penetration rate.

The drawback of a limited penetration rate is that the flow and density cannot be measured, because a certain amount of vehicles with FCD is tracked in a time and space interval, but the penetration rate is unknown. A solution for this problem is to combine loop detectors with FCD. When measuring the total flow Q_i at the beginning of a (longer) road section and count the number of cars with FCD n_i, the penetration rate of the FCD can be calculated. Now the density and flow can be emanated from the FCD. Note that a new loop detector is only needed after each on- and off-ramp, since the flow will change much at these locations.

As a practical example of FCD versus loop detector data, we will compare FCD and loop detector data for a road stretch of the A1 in the Netherlands from Apeldoorn to Amersfoort. The data is plotted in an x,t-diagram in figure 2.5, the loop detector data in the graph at the top and the FCD in the lower graph. A congestion is visible in this graph, starting around 9:00 at Amersfoort and building up towards Apeldoorn in time, till the congestion is resolved around 12:30.

The first thing that stands out is the lack of loop detector data, the grey parts of the graph represents that there is no information available. And of course the loop detectors are only measuring at specific locations, visible in vertical lines in the x,t-diagram. In the graph with FCD, there is also a grey area. This is downstream of the congestion, in this case the shortage of data is explicable by the low flow downstream of the congestion. Because of the low flow, a limited number of vehicles is driving there at the time of the congestion. Since only a percentage of 5-10% of the vehicles is providing FCD, the possibility of no FCD is larger in case of a limited number of vehicles [1].

The graphs are showing that FCD can give more information over the spacial axis. Nowadays they are researching the actual quality of this FCD, especially in relation to the low penetration rate.

So the most important differences:

- FCD contains traffic information of individual vehicles, loop detector data contains averages about the traffic.

- FCD is continuous in the driving direction, it gives trajectories. A loop detector is measuring at a specific location, on average each 500 meter. But on some roads there are no loop detectors available at all.

- FCD gives trajectories with location and speed (and in the future maybe more advanced data is possible), loop detector data gives average speed and average flow per minute.
- FCD gives information about a limited amount of vehicles, the penetration rate. Loop detector data is measuring all vehicles.

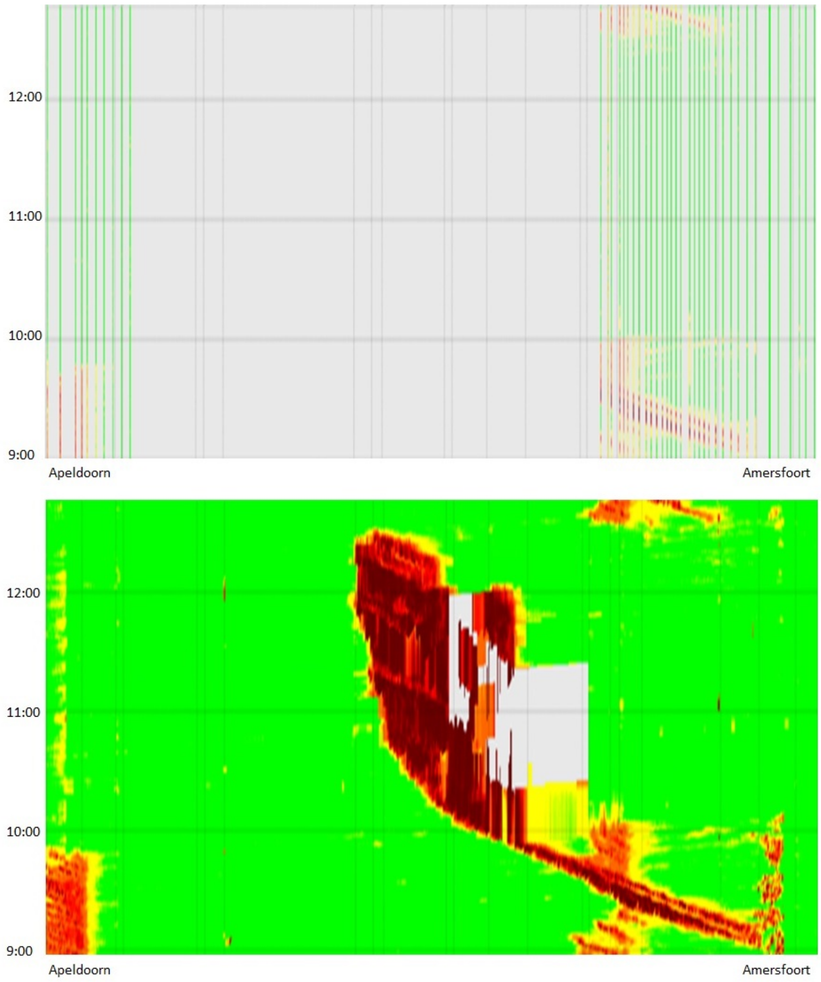


Figure 2.5: Space-time-diagram of A1 in a morning peak, upper graph of loop detector data, lower graph of FCD [1]. The colours represents the speed, from green to red respectively free flow speed to a speed of zero

2.3. Automatic incident detection algorithms

When traffic data is available, from FCD or loop detectors, and traffic characteristics of incidents can be recognised, incidents can be detected by an algorithm. First the requirements of the automatic incident detection (AID) algorithms will be discussed based on the traffic characteristics of incidents. In the second part of this section, some existing AID algorithms will be reviewed with their performance in the last part of the section.

2.3.1. Requirements AID algorithm

As discussed in previous section, incidents can block the road totally or partially. The easiest incident to detect is an incident that blocks the road totally. In this case a clear vacuum in terms of flow and density can be observed downstream the incident location. More difficult is the partly blocked road, the difference in flow and density downstream and upstream of the incident location will be less. Especially when the flow upstream the incident location is very low, there is a possibility of not observing any difference comparing the flow upstream and downstream the incident location. Biggest problem in this detection method is the detection of incident that block the road partly and especially in case of low flows.

To detect incidents, most existing AID algorithms are detecting the difference in flow, density and/or speed upstream and downstream of the incident location. This can be done directly or first detecting congestion and then checking if the characteristics of an incident are present. The biggest challenge is distinguish an incident from another cause of congestion, when measuring at least two of the traffic quantities (speed u, flow Q, density K) the specific characteristics for incident can be better checked and in this way a more accurate detection is possible.

Besides the traditional quantities for traffic measuring (flow, density and speed), FCD can bring new quantity in the future: lane changing. In case of a partly blocked road a lane change is needed from the vehicles using the blocked lane upstream of the incident location. But also downstream of the incident location are lane changes expected to the blocked lane. This can be an opportunity for the detection of incidents that are blocking the road partly.

Another important criteria for the algorithm have to do with the statistical type I and type II errors. The criteria is that the type I and type II errors have to be minimal. A type I error is a false positive, also called a false alarm, this is the case when the algorithm indicated an incident, but there is not. The type II error is a false negative, thus an occurred incident that is not detected by the algorithm. Ideally both errors are zero, but at least we want to find all incidents correctly, so the type II error as close to zero as possible.

So the requirements for AID algorithms are:

1. Detect all incidents, not missing an incident;

2. Not giving a signal when there is no incident, a false alarm;

3. Quick detection of incidents.

2.3.2. Existing AID algorithms

Since 1970 several AID algorithms are developed which facilitate this early detection and to make quicker detection possible. Most of the current AID algorithms are based on loop detector data. A lot of the developed algorithms from 1970 are still being used. Different algorithms use different criteria for the detection, different features of an incident in traffic engineering terms are used. In general the algorithms can detect the congestion as a result of an incident. After the congestion detection, the algorithm checks the reason of the congestion. So these algorithms need to distinguish a traffic jam caused by an incident from jams caused by a bottleneck and stop-and-go waves.

The theory of three incident detecting algorithms will be elaborated in this subsection: Blokkadedetector, Preshikhaaf and McMaster algorithm. The Blokkadedetector is now developed in the Netherlands and is comparable with the well-known California algorithm, it recognises patterns in density. The second algorithm, the Preshikhaaf algorithm, recognises patterns in speed to detect incidents. The last algorithm is the McMaster algorithm, this one is used outside of the Netherlands. Three different approaches to detect incidents, these three will be discussed in the following paragraphs, concluding with a review on the performance of the algorithms with the deployed requirements for AID algorithms.

Blokkadedetector The Blokkadedetector is developed by Rijkswaterstaat and as data source loop detectors are used, which give data averaged per minute. This algorithm is comparable to the California algorithm, a pattern-based algorithm. A significant difference with the California algorithm is that the Blokkadedetector uses flow and speed as input, while the California algorithm uses speed and occupancy, quantities that are more common outside the Netherlands.

The performance of this algorithm is not optimal, evaluation studies give a relative low detection rate of approximately 30% for this algorithm with a quite high detection time. Further research show that at least 25% of the traffic has to be blocked for good detection with this algorithm.

The algorithm The blokkadedetector checks in different time steps the traffic characteristic of an incident. The basics of this algorithm are visible in figure 2.6. The three checks are executed in three subsequent time steps, since loop detectors provide data every minute, the checks are executed in three subsequent minutes. An incident message can be provided after every check of the algorithm. The different steps will be discussed below.

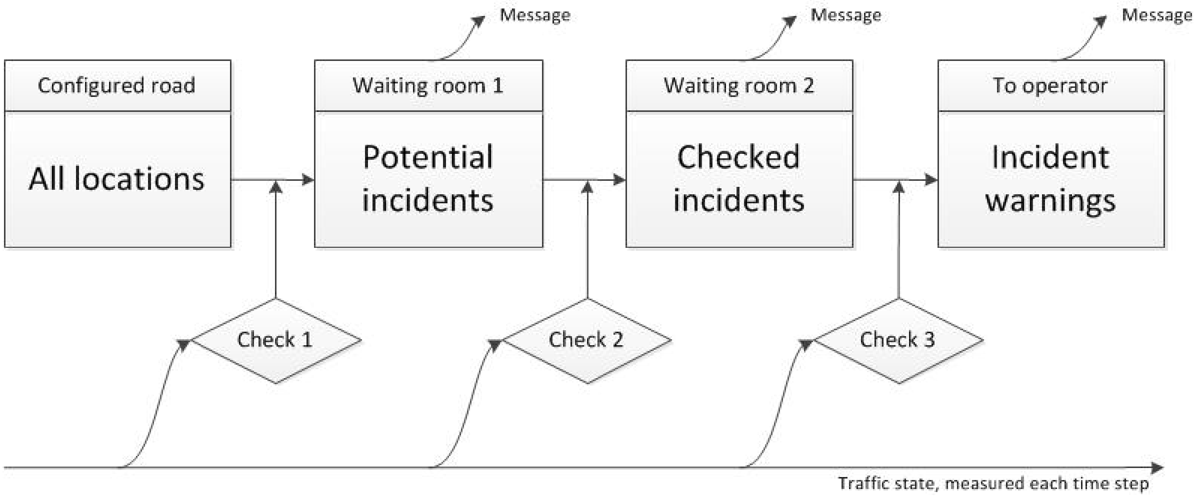


Figure 2.6: Schematic description of the Blokkadedetector [15]

Step 1: time t
The first check will be executed at all measurement locations and at all time steps. In this first check is searched for a flow vacuum, since this is a specific characteristic of an incident. Thereby the expansion speed of the flow vacuum downstream of the incident is higher and thus the vacuum can be measured easier at an early stadium. See also figure 2.7.

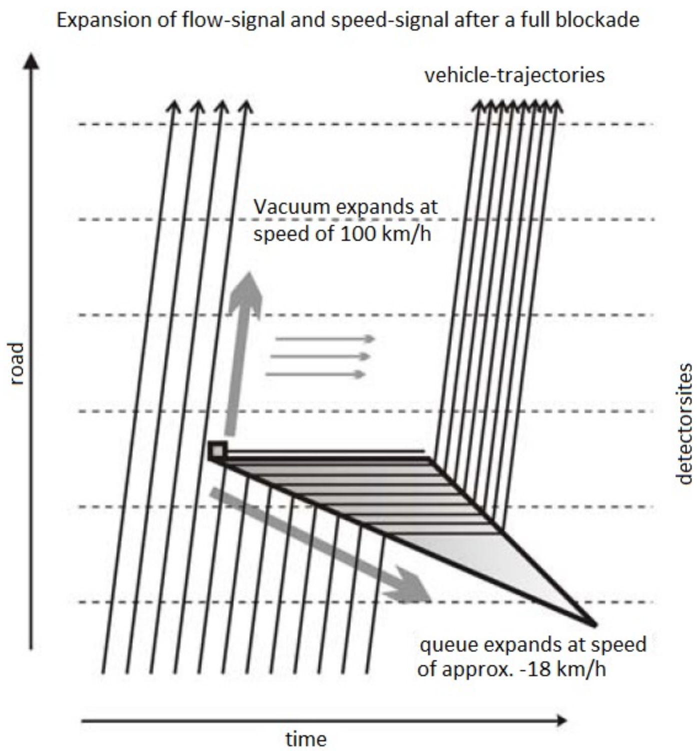


Figure 2.7: x,t-diagram for an totally blocked road [15]

To recognise a flow vacuum for every time step at every measurement location the measured flow is compared with an expected flow. The expected flow is determined by looking back in time and location and checks if these characteristics match with the measured characteristics.

$$Q_e(t,x) = (1 - c(t,x)) \cdot Q_r(t,x-1) + c(t,x) \cdot Q_r(t-1,x-1) \tag{2.8}$$

x	= measurement location
$x-1$	= first upstream measurement location
Q_e	= expected flow
Q_r	= relative flow
c	= weight factor dependent on distance and speed

The weight factor $c(t,x)$ is the travel time from measurement location $x-1$ to x.

$$c(t,x) = \max\left(60 \cdot \frac{d}{v(t,x)}, 1\right) \tag{2.9}$$

v	= speed in km/h
d	= distance from measurement location $x-1$ to x

The influence of this factor represents the share of data that the platoon of traffic was at $x-1$ in this minute and the share that was there one minute ago. This means that in case of a higher speed, the traffic of previous minute has a higher share and the estimated flow $Q_e(t,x)$ is highly based on the data of previous minute. In case of a low speed, the factor $c(t,x)$ is low and in the formula of estimated flow the part on time step t has a high share.

To prevent systematic errors due to failure of loop detectors and to limit influences from on and off ramps, the relative flow Q_r is calculated with measured flows Q_m and the moving average Q_s of two locations.

$$Q_r(t,x) = \frac{Q_s(t,x+1)}{Q_s(t,x)} \cdot Q_m(t,x) \tag{2.10}$$

$x+1$	= first downstream measurement location

The moving average is calculated with the following formula.

$$Q_s(t,x) = A \cdot Q_m(t,x) + (1-A) \cdot Q_s(t-1,x) \tag{2.11}$$

with

$$A = \begin{cases} 0.3 & \text{if } Q_s(t) > Q_s(t-1) \\ 0.4 & \text{otherwise} \end{cases}$$

Now the measured flow will be compared with the expected flow. In case the measured flow is a lot less than the expected flow, an incident is recognised. A threshold value $H_{q-little}$ is used. If equation 2.12 holds a message of an incident is given by the algorithm.

$$\frac{Q_m(t,x)}{Q_e(t,x)} < H_{q-little} \tag{2.12}$$

If the measured flow is less than the expected flow, but does not directly trigger the algorithm to send an incident message, the situation can be sent to waiting room 1 as potential incident and check 2 is executed in the next time step. This is the case if equation 2.13 holds with a bigger threshold value H_{q-big}.

$$\frac{Q_m(t,x)}{Q_e(t,x)} < H_{q-big} \tag{2.13}$$

Step 2: time $t + 1$
This second step is only executed for all potential incidents in waiting room 1 after check 1. For these potential incidents is checked if the vacuum still exists and if the speed downstream is high enough.

When at the location downstream a flow drop rises, an incident is detected. This is the case if the condition of equation 2.14 hold.

$$\frac{Q_m(t, x+1)}{Q_e(t-1, x)} < G_{q-little} \tag{2.14}$$

If not, a few more checks are done to identify the vacuum in this next time step.

$$\frac{Q_m(t+1, x)}{Q_e(t+1, x)} < G_{q-big} \tag{2.15}$$

$$\frac{Q_m(t, x-1)}{Q_e(t, x-1)} < G_{q-big} \tag{2.16}$$

$$\frac{Q_m(t+1, x-1)}{Q_e(t+1, x-1)} < G_{q-big} \tag{2.17}$$

In case one of above three conditions holds, an extra check is executed to check the downstream speed.

$$v(t+1, x+1) > V_{check} \tag{2.18}$$

If at least one of the conditions of equations 2.15, 2.16 and 2.17 is true and also condition 2.18 is true, then a possible incident is detected and this situation is send to waiting room 2 for check 3.

Step 3: time $t + 2$
This third and last step is only executed for all potential incidents in waiting room 2 after check 2. A control step is done in the time step $t + 2$ if the speed and flow differences are significant high to detect an incident.

$$\frac{v(t+2, x-1)}{v(t+2, x+1)} < F_{v-alarm} \tag{2.19}$$

$$\frac{Q_m(t+2, x+1)}{Q_s(t-1, x+1)} < F_{q-alarm} \tag{2.20}$$

Only if both above mentioned conditions hold an incident message is given by the algorithm.

Presikhaaf algorithm This algorithm is also pattern-based and developed as an alternative for the above Blokkadedetector. Instead of detecting differences in flow like the Blokkadedetector, this algorithm detects differences in speed. Like is seen in figure 2.7 a low speed is measured upstream of the incident and downstream the incident a critical high speed.

The algorithm Just like the Blokkadedetector, the Presikhaaf algorithm consists of three steps in sequent minutes to be sure of an incident on the road. Instead of following platoons of vehicles like the Blokkadedetector, the Presikhaaf algorithm uses the moving average. To prevent errors from heterogeneous traffic and errors in the data the current minute is compared with the moving average of the previous five minutes.

Step 1: time t
For every minute and every measurement location the measured speed is compared with the moving average of previous five minutes. If the difference between them is above threshold, a possible incident is detected. In this case equation 2.21 must hold.

$$\frac{v(t, x)}{v_{ma}(t, x)} < A_1 \tag{2.21}$$

With $v_{ma}(t, x)$ the moving average of the five previous minutes of t on location x.

But a lower speed can also be a result of congestion. To check this the downstream speed is checked, this speed needs to be high.

$$v(t, x+1) > A_2 \tag{2.22}$$

A lower speed can be caused by an stop-and-go wave. This option has to be excluded with the following condition.

$$v_{ma}(t, x+1) < A_3 \tag{2.23}$$

If all three above discussed conditions holds, a possible incident is detected and step 2 is done.

Step 2: time $t+1$

One minute later the second step is executed for all possible incidents that are detected in step 1. In this time step is checked if the speed is still decreasing compared with the moving average and if the downstream speed is higher than the threshold of the end of an traffic jam. Both undermentioned conditions must hold.

$$\frac{v(t+1, x)}{v_{ma}(t+1, x)} < B_1 \tag{2.24}$$

$$v_{ma}(t+1, x+1) < B_2 \tag{2.25}$$

But also the downstream flow has to be lower than the moving average of the flow, since this is one of the characteristic specific for a congestion due to an incident.

$$\frac{Q(t+1, x+1)}{Q_{ma}(t+1, x+1)} < B_3 \tag{2.26}$$

If all three above discussed conditions holds, the possible incident still exists in this time step and step 3 is done.

Step 3: time $t+2$

This step is only executed for the situations that in step 2 are still recognised as possible incidents. A last check is done on the current speed and a comparison with the moving average of the speed.

$$v(t+2), x) < C_1 \tag{2.27}$$

$$\frac{v(t+2, x)}{v_{ma}(t+2, x)} < C_2 \tag{2.28}$$

Only if also these two conditions are true, an incident message is given by the algorithm.

McMaster algorithm The McMaster algorithm is an example of an algorithm from the catastrophe theory, where incidents are detected on the fact that one of the flow variables, flow, density or speed, will have a sudden change while in other chases all variables change gradually. For this algorithm, the occupancy of loop detectors is used [9].

The algorithm The incident is detected in two phases by this algorithm, first the traffic state is determined and in the second phase this information is used to distinguish congestion due to incidents from other causes. First the two phases are explained, thereafter the used parameters are discussed.

Phase 1

To determine the traffic state the flow-occupancy diagram is made and divided in four parts like is done in figure 2.8. Horizontally the graph is divided by the line OC_{max}, the critical occupancy before congestion sets in. Left of the line OC_{max} the curve $g(occ)$ divides this part of the diagram in state 1 and state 2. The curve $q(o)$ is an approximation of the measured occupancy and flow, the free flow part of the fundamental diagram. The curve $g(occ)$ is a fraction of $q(o)$ so that $g(occ)$ is below capacity, see also the extensive explanation further down. At the right side the line q_{crit} is dividing the graph in state 3 and state 4, where q_{crit} represents the critical flow.

The decision process of this first phase is presented in figure 2.9. So four traffic states can be determined:

- State 1 is a non congested state, with high flow and low occupancy
- State 2 represents traffic states downstream of congestion, the flow is lower than $g(occ)$, but the occupancy is lower than the critical value OC_{max}
- State 3 is a second congested form. In this case the flow is lower than the critical value q_{crit} and the occupancy is higher than the critical occupancy OC_{max}

- State 4 describes the accelerating vehicles downstream the blockade, the occupancy is higher than OC_{max}, but the flow is higher than his critical value.

When a traffic state 2 or 3 is determined at a certain measurement location, phase 2 is executed to check if this congestion is a result of an incident.

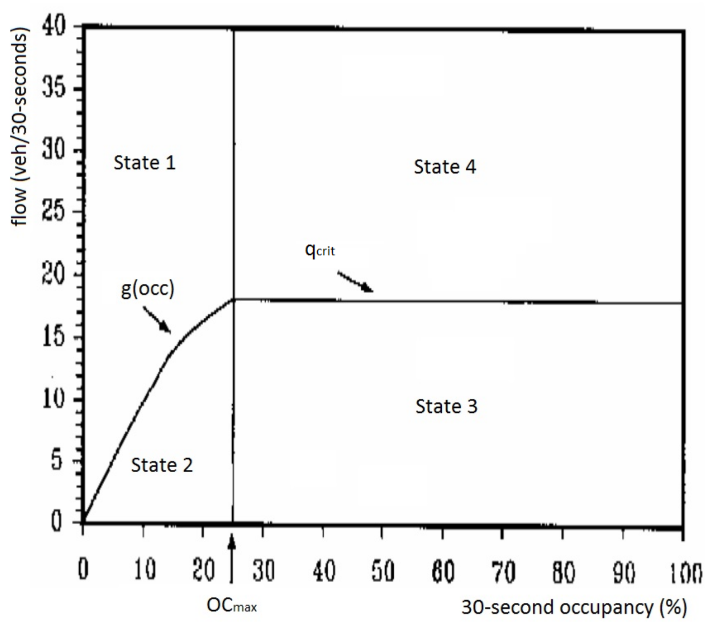


Figure 2.8: Traffic states in an occupancy-flow diagram; state 1 = no congestion; state 2 = congestion; state 3 = congestion; state 4 = accelerating traffic downstream of congestion

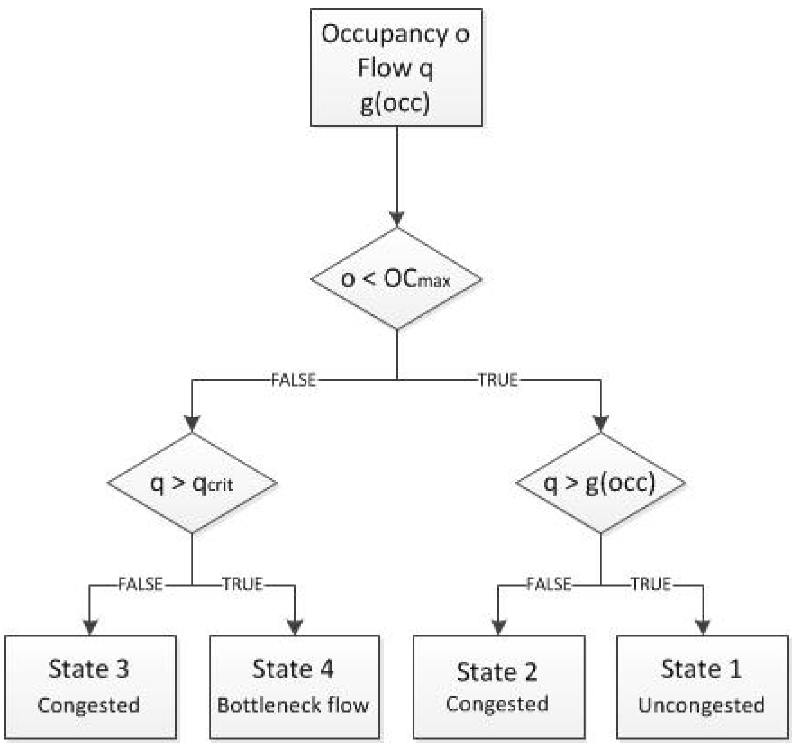


Figure 2.9: Graphical representation of the decision making process in phase 1 of the McMaster algorithm

Phase 2
As said above, only for situations with traffic state 2 or 3 this phase is executed, the detected congestion on a certain measurement location x is analysed further to determine if the cause is an incident. To determine this, the traffic state for the first downstream measurement location $x + 1$ is defined. If the traffic state in location $x + 1$ is state 1 or 2, than an incident is the cause of this congestion.

In figure 2.10, the decision making process of the second phase is presented.

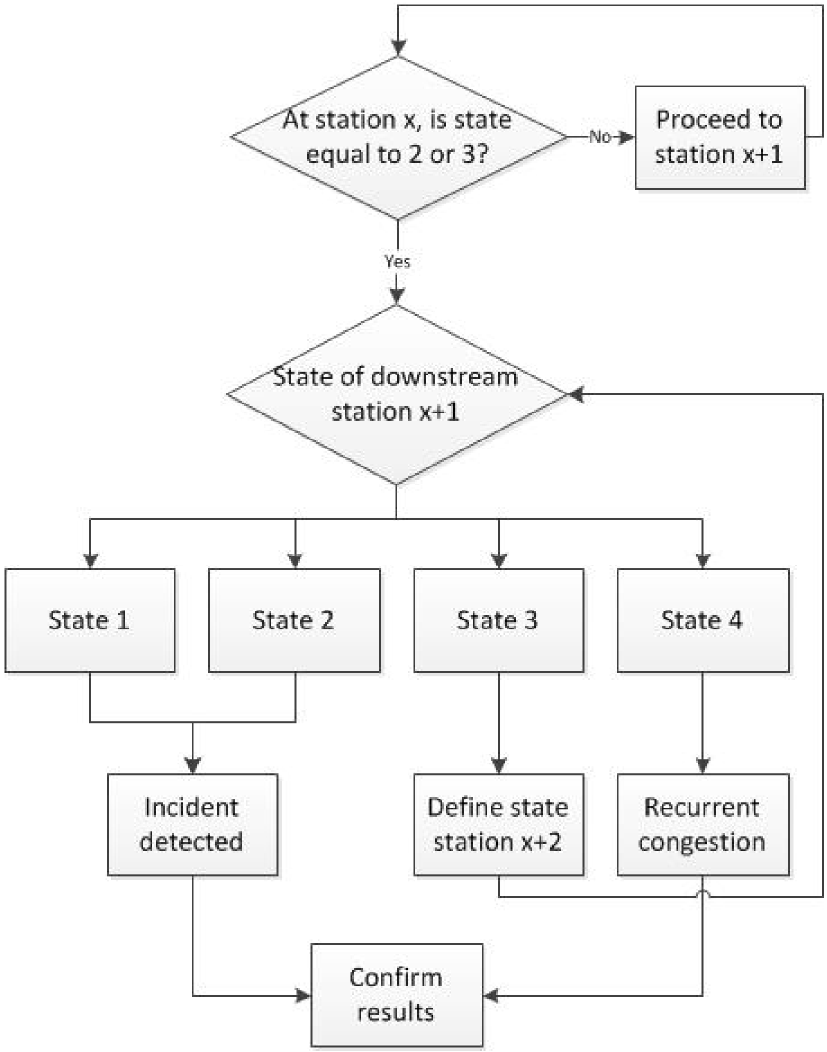


Figure 2.10: Graphical representation of the decision making process in phase 2 of the McMaster algorithm

Traffic behaviour and traffic states are dependent on the local conditions of the location, like on and off ramps and the number of lanes. This means that the thresholds to determine the traffic states are also dependent on local conditions. Below the parameters to determine the four traffic states will be discussed.

$g(occ)$
If all measurement point of occupancy and flow are displayed in a graph, the points represents the free flow part of the fundamental diagram, which can be approximated by the following formula.

$$q(o) = b \cdot o^a \tag{2.29}$$

$q(o)$	= flow
o	= occupancy
a,b	= location dependent parameters

The values for a and b has to be determined for each location separately to fit the free flow part of the fundamental diagram as well as possible. Traffic state 1 and 2 are distinguished by:

$$g(occ) = m \cdot q(o) \tag{2.30}$$

In this formula m is a value between 0 and 1 that moves line $q(o)$ down. An heterogeneous traffic flow is assumed, so deviations will be relatively big, so the value of m has to be between 0.8 and 0.9.

q_{crit}
The critical value of the flow that divide state 3 and 4 can be determined by analysing historical data of incidents and speeds. From other literature studies is learned that $q_{crit} = 1250$ per lane. So the value of q_{crit} have to be calculated by multiplying the number of lanes at a detector location with 1250.

OC_{max}
The critical occupancy can be calculated by calculating the intersection of q_{crit} and $g(occ)$. Since q_{crit} and $g(occ)$ differ for each location, the OC_{max} will also differ per detector location.

2.3.3. Performance of existing algorithms

The main limitations of existing AID algorithms according to literature are the unsatisfactory quality of the traffic data and the limited filtering before using the data by the algorithms. In addition the current algorithms are not effective in distinguishing an incident from other congestion causes [7].

To assess the performance of the above described algorithms, three scenarios and the requirements of the beginning of this section are used, to repeat:

1. Detect all incidents, not missing an incident;
2. Not giving a signal when there is no incident, a false alarm;
3. Quick detection of incidents.

The described requirements will be reviewed for all three previous discussed algorithms. But first three scenarios are sketched to check the theoretical performance of the three algorithms. Based on these scenarios, the requirements will be checked as well.

Testing the algorithms for three scenarios The three scenarios are visible in figure 2.11. The first scenario is a two lanes road with an incident. In this case the flow Q is lower than the remained capacity of one lane at the incident location, so no congestion will occur after the incident. In the second scenario, the flow is bigger than the capacity of one lane, so an queue will build up after the incident. The last scenario is one without an incident, there is congestion building up from a busy off-ramp.

Blokkadedetector
The results of the first algorithm, the Blokkadedetector, for the three scenarios:

Scenario a There is no congestion occurred in this case, so no differences in flow will be found, the Blokkadedetector will not detect the incident. **A false negative**.

Scenario b Difference in flow will be found in this case, so the incident is detected. **A correct detection**.

Scenario c In this case there is a chance to find a significant difference in the flow to characterise this situation as an incident. **A false positive**.

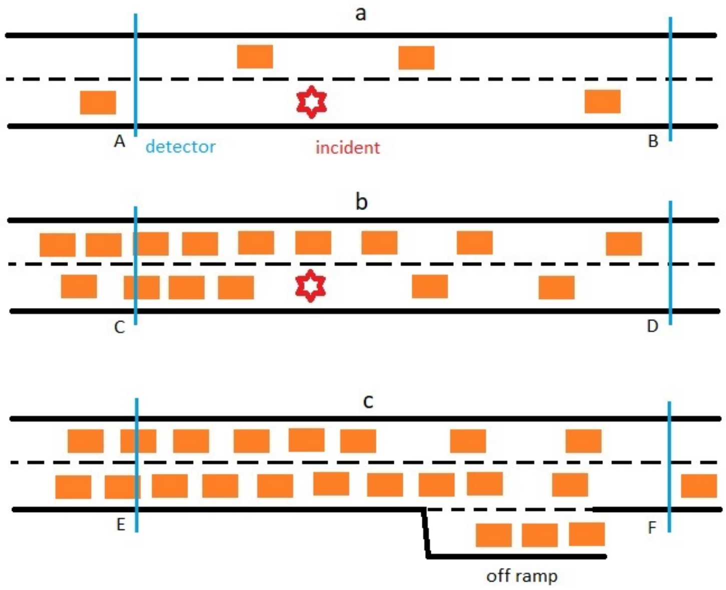


Figure 2.11: Three scenarios to test the algorithms; a. $Q < Q_{capacityonelane}$ with incident; b. $Q > Q_{capacityonelane}$; c. off-ramp that is causing congestion on the main road, no incident.

Presikhaaf algorithm
The results of the Presikhaaf algorithm are discussed below:

Scenario a Like the Blokkadedetector, the Presikhaaf algorithm will not detect this incident, since there is no difference in speed and flow when comparing before and after the incident. **A false negative**.

Scenario b In case of congestion after an incident, a difference in flow and speed will be found and this incident will be detected. **A correct detection**.

Scenario c Also the Presikhaaf algorithm will sometimes detect this scenario as an incident, partly due to the stochastic nature of traffic. So this is **a false positive.**

McMaster algorithm
The scenarios are again presented in a figure, with a o, Q-diagram at the side. At the right side of the figure the graph is sketched that is used by the McMaster algorithm to detect incidents, with the fundamental diagram in grey. See figure 2.12. The results of the McMaster algorithm for these scenarios:

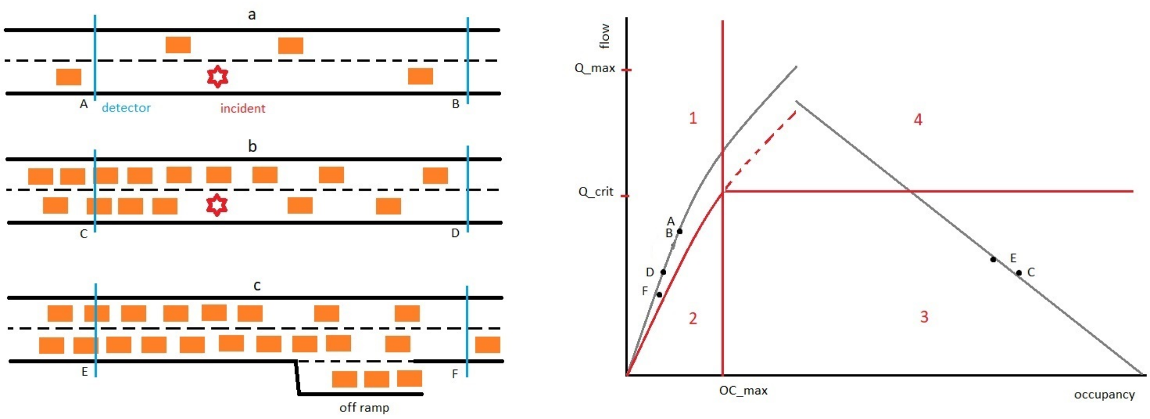


Figure 2.12: Left are three scenarios visible and at the right hand the graph for the McMaster algorithm, in grey the fundamental diagram.

Scenario a In the case of a flow q_i that is less than the capacity of one lane q_{max1}, so $q_i < q_{max1}$, than no congestion will occur in case of an incident. Both detectors A and B will give data on the free flow branch of the fundamental diagram, so in state 1 of the McMaster algorithm. Therefor the McMaster incident will not detect this incident. A type I error, also called a **false negative**.

Scenario b In case of a flow $q_i > q_{max1}$, a queue will occur. Now the detector C will give data on the congested branch of the fundamental diagram with a low flow and detector D gives data with the same flow, but on the free flow branch. So state 3 for detector C and state 1 for the next downstream detector D and the McMaster algorithm does recognise this incident. A **correct detection**.

Scenario c Now an off-ramp that is causing congestion on the main road. The traffic at detector E is congested with a low flow and high occupancy, the downstream detector F gives data of free flow conditions. So state 3 for detector E and state 1 for detector F. In this way, the McMaster algorithm will indicate this situation as an incident. A type II error, also called **false positive**.

So all three algorithms potentially give many false alarms (type II errors).

1. Detect all incidents, not missing an incident As seen when testing the scenarios earlier in this section, all algorithms are having troubles with distinguish the off-ramp that is causing congestion from an incident. Mainly a congestion in combination with an infrastructure change is causing problems for the algorithms, exactly where often queues will occur.

Like is presented above, the algorithms will not detect an incident in case of $Q < Q_{restcapacity}$. Since all three algorithms need congestion to detect the incident. Although the incident will not cause a queue in this case, it is important to detect it for example to prevent secondary incidents and to clean the incident before the flow is increased till more than $Q_{restcapacity}$.

Another reason for bad detection is that the Blokkadedetector and the Presikhaaf algorithm are only checking for differences in one traffic quantity to identify an incident, respectively flow for the Blokkadedetector and speed for the Presikhaaf algorithm.

In literature, there are some detection rates found for incident detection algorithms. Although the values for Blokkadedetector and the Presikhaaf algorithm cannot be found. The California algorithm is also detecting incidents based on patterns in flow and speed, so the results will probably be comparable. The found detection rate is visible in table 2.1. The concepts detection rate, false alarm rate and detection time will be discussed more extensively in chapter 4.

Table 2.1: Detection rate of existing incident detection algorithms according to [8] and [9].

Algorithm	**Detection rate**
McMaster	0.63-1.0
California algorithm	0.67

The detection rate is ideally close to 1.0, which means that all incidents are detected. The literature gives different detection rates for the McMaster algorithm from 0.63 till 1.0. Since a case if found when the McMaster algorithm will not detect the incident (with low flow), the 0.63 seems more realistic than the 1.0 given in literature.

2. Not giving a signal when there is no incident, a false alarm In the tested scenarios is concluded that all algorithms will possibly indicate the third scenario as an incident. This will depend on the stochastic nature of the traffic and thus the variety in the traffic characteristics. The false alarm rate of the algorithm is presented in table 2.2.

Table 2.2: False alarm rate of existing incident detection algorithms according to [8] and [9].

Algorithm	**False alarm rate**
McMaster	0.000012-0.0004
California algorithm	0.0013

Again different false alarm rates for the McMaster algorithm in different literature articles, but in general very low false alarm rates. Also the California algorithm has a low false alarm rate, as well the Blokkad-edetector and the Presikhaaf algorithm. Side-note is that the exact scenarios on which the algorithms are tested are not clear. Since the notice that these algorithms will have troubles with infrastructural changes, the false alarm rates will possibly be higher when adding this scenario to the research.

3. Quick detection of incidents A queue have to build up before all three algorithm can detect the incident, so that will take some time. Also the distance from the incident location to the next upstream detector location will influence the detection time. The detection times from literature give the shortest time for the McMaster algorithm, displayed in table 2.3.

Table 2.3: Detection time of existing incident detection algorithms according to [8] and [9].

Algorithm	Detection time[s]
McMaster	90.0-132.0
California algorithm	174.6

So in general all algorithms have problems with infrastructural changes like an off-ramp that is causing a queue on the main road. This will often be recognised as an incident instead of just congestion. Another problem is detecting an incident when the flow is lower than the remained capacity at the incident location.

The McMaster algorithm is classified as the best of these three according to the literature, so it will be interesting to compare the new algorithm to the McMaster algorithm.

The new designed algorithm will be explained in the next chapter, after which the research set-up is stated in chapter 4.

3

New incident detection algorithm development using floating car data

With the lessons learned from the existing algorithms and the criteria for incident detection algorithms in chapter 2, in this chapter a new approach for detecting algorithms is presented. By designing the new algorithm the new possibilities of floating car data (FCD) are taken into account, as are the difficulties of detecting incidents bij previous algorithms, especially in case of a low flow.

The general concept of the new algorithm is discussed first, after which some important details will be explained.

3.1. General concept

As seen in chapter 2, an incident can be detected by a combination of a flow vacuum and high speed downstream of the incident location and high flow and low speed upstream of the incident detection. The new algorithm is not based on detecting these characteristics, but based on the observed lane changes upstream and downstream of the incident.

In case of an incident, probably there will be an increase in the number of lane changes compared to a situation without an incident. This increase is due to a blocked lane, so road users will change to another lane upstream of the incident location. Downstream the road users will divide again over all lanes so some vehicles will change to the previous blocked lane.

The general idea of the new algorithm is counting the number of lane changes on a road section per minute. A threshold on the number of lane changes will be made based on historical data without an incident. If a new data point is collected, these are compared to the threshold. A higher number of lane changes will indicate an incident.

The new algorithm step-by-step:

Step 1 Determine thresholds $T(X, K_i)$ for road sections of 200 meter, with

X	= road section number
K_i	= density interval with $i = [0, 10), [10, 20), [20, 30), ...$

1. Count number of lane changes $N(X, K_i)$ per minute;
2. Measure flow $Q(X)$ and speed $u(X)$ in a minute;
3. Calculate the average density $K(X)$ of the road section in a minute;

$$K(X) = \frac{Q(X)}{u(X)} \tag{3.1}$$

4. Take the upper boundary $t_u(X, K_i)$ of a 95%-confidence interval that is made of the number of lane changes $N(X, K_i)$

$t_u(X, K_i)$ = upper boundary of 95%-confidence interval of $N(X, K_i)$

5. Calculate threshold $T(X, K_i)$

$$T(X,K_i) = \begin{cases} roundup(t_u(X,K_i)) + 2 & \text{if } t_u(X,K_i) < 10 \\ roundup(t_u(X,K_i)) + 4 & \text{if } 10 \geq t_u(X,K_i) < 20 \\ roundup(t_u(X,K_i)) + 8 & \text{if } t_u(X,K_i) \geq 20 \end{cases} \tag{3.2}$$

Step 2 Repeat step 1 for all road sections X

Step 3 Detecting an incident

1. Count number of lane changes $N(X)$ per minute on road section X;
2. Measure flow $Q(X)$ and speed $u(X)$ in a minute on road section X;
3. Calculate the average density $K(X)$ in a minute and determine i (with $i = [0, 10), [10, 20), [20, 30), ...$
4. Take the threshold $T(X, K_i)$ which responds to the measured density K_i. So if $K_i = 54.3$ on road section X, take $T(X, [50, 60))$;
5. If the measured number of lane changes $N(X)$ is smaller than the threshold $T(X, K_i)$, an incident is detected.

$$N(X) < T(X, K_i) \qquad \Rightarrow \qquad \text{incident detected} \tag{3.3}$$

Note that for every road section of 200 meter in length, the number of lane changes is measured separate since the road lay-out will heavily influence the number of lane changes (lane drops, on and off ramps etc.).

The most important design details are discussed more extensively below.

3.2. Design details

A few choices are made in the design of the new algorithm as presented above. The most important design details will be discussed below: the dependency of the number of lane changes on the density, the determination of the thresholds and the length of the road sections. Also some alterations to the algorithm will be discussed when a lower (than 100%) penetration rate of FCD is available.

3.2.1. Dependency on density

When counting the number of lane changes on a road section, a higher number of lane changes is found when the road is crowded than on a quiet moment. The simulation that is set-up in chapter 4, is showing this relation in figure 3.1. The crowdedness of the road is expressed in the density K on the x-axis of the graph. The number of lane changes N is presented on the y-axis. Each point in the graph gives information about the average number of lane changes and average density of one minute on a road section of 200 meter. The colours in the graph represent the ten runs in the simulation program FOSIM, which can be seen as 10 different days to determine the threshold. Knowing from which run a point is, is not needed for the algorithm, but there are different runs needed to determine reliable thresholds and results.

The relation between the density and the number of lane changes is visible in figure 3.1. The maximum of lane changes is equal to 3 for $10 \leq K < 20$ and equal to 11 for $40 \leq K < 50$. This is taken into account when determining the thresholds for the algorithm. Especially the outliers are more extreme for higher densities. These outliers are measured, since the lane changing process is a stochastic process.

Note that there are no observations of the traffic for really low density ($0 \leq K < 10$) and high density ($K > 70$). Another important notice is that the density will be calculated from speed u and flow q, based on loop detector data, since no flow can be measured with FCD. A disadvantage of this method is that the density of congested traffic conditions is underestimated.

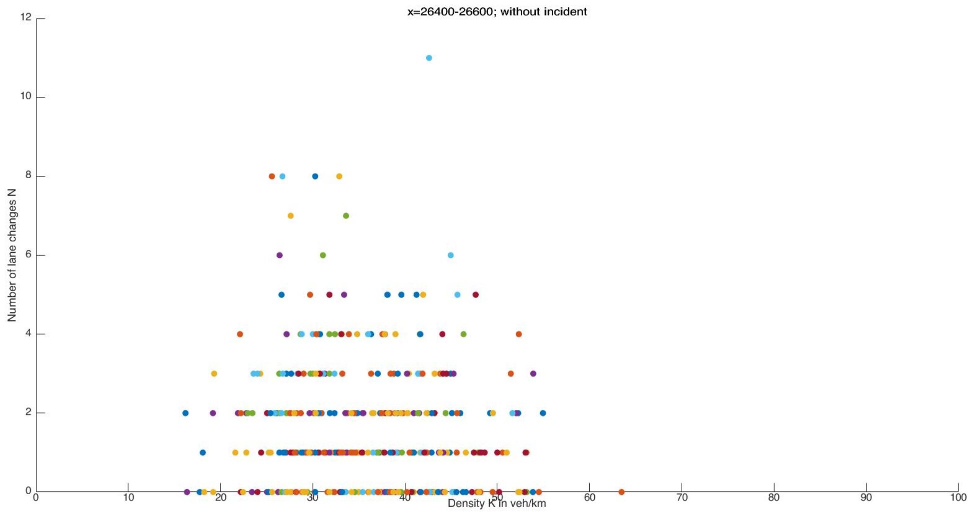


Figure 3.1: Relation between density K and number of lane changes N.

3.2.2. Threshold determination

As said before, the thresholds are determined for each road section of 200 meters long, since the road lay-out heavily influences the number of lane changes N. Even a series of thresholds is determined for each road section, because the number of lane changes is dependent on the density, as seen in the previous subsection.

The threshold determination is explained based on the graph in figure 3.2.

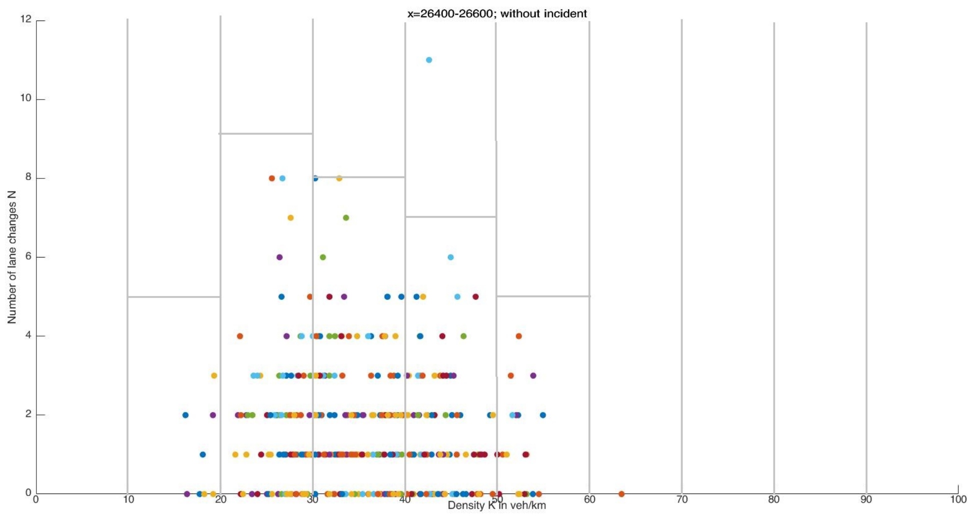


Figure 3.2: Example threshold determination. Vertical grey lines display the intervals in K_i for which the thresholds are made, the horizontal grey lines are the thresholds based on historical data without incident.

Again the results of road section 26400-26600 are visible in this figure, but now the threshold determination is added. The graph is divided in columns of 10 veh/km width, indicated with vertical grey lines. A threshold is determined for each column. So in column $10 \leq K < 20$ the threshold $T(X, [10-20)$ is equal to 5.

Since the lane changing is a stochastic process, there are outliers in the number of lane changes. Therefore the threshold is determined as follows:

- For each road section of 200 meter, called X;
- For a density interval of 10 veh/km width

$$K_i \quad \text{with} \quad i = [0,10), [10,20), [20,30), ... \tag{3.4}$$

- Take the upper boundary $t_u(X, K_i)$ of a 95% confidence interval that is made of the number of lane changes in the specific density interval K_i, $N(X, K_i)$;

$$T(X, K_i) = ceil(t_u(X, K_i)) + \alpha \tag{3.5}$$

  The safety limit α represents the variance of the number of lane changes, since this variance is higher in case of a high number of lane changes, the value of α will be higher for high number of lane changes.

  – If this upper boundary $t_u < 10$, round up to the next integer and add two as a safety limit.
  – If this upper boundary $10 < t_u \leq 20$, than round up to the next integer and add four as a safety limit.
  – And if this upper boundary $t_u \geq 20$, than round up to the next integer and add eight as a safety limit.

$$T(X, K_i) = \begin{cases} ceil(t_u(X, K_i)) + 2 & \text{if } t_u(X, K_i) < 10 \\ ceil(t_u(X, K_i)) + 4 & \text{if } 10 \geq t_u(X, K_i) < 20 \\ ceil(t_u(X, K_i)) + 8 & \text{if } t_u(X, K_i) \geq 20 \end{cases} \tag{3.6}$$

In this manner the effect of extreme outliers on the determination of the threshold is decreased.

Note that multiple observations per interval on the density K are needed to determine a good threshold. In this research it is chosen the require at least five observations. If less than five observations are available on an interval on K, then no threshold is determined and thus no incident can be detected for this range in the density.

In figure 3.3 the results of scenario with an incident are added. The filled dots in this figure are the historical data points on which the thresholds are made, and the stars are the data points in a scenario with an incident.

3.2.3. Road section length

The number of lane changes is counted on a road section of 200 meter in the new algorithm as discussed in the section ‘General concept’. In this subsection is also a longer road section discussed. The results of a 200 meter long road stretch and a 400 meter long road stretch (same location, 200 meter longer in the downstream direction) are presented in figure 3.4. Again the filled dots are the historical data to determine the thresholds and the stars are the data points in case of an incident.

The differentiation between the stars and the dots is more clear in the left graph. When looking at the columns of 10 veh/km width, there are no stars lower than the highest dots. In case of the 400 meter road section, there are multiple stars lower than the highest dot in the columns of 10 veh/km width. These stars are indicated with a red circle, especially in $20 \leq K < 30$.

Based on the results for this location and other locations that are checked, a road section length of 200 meter is chosen for the new algorithm.

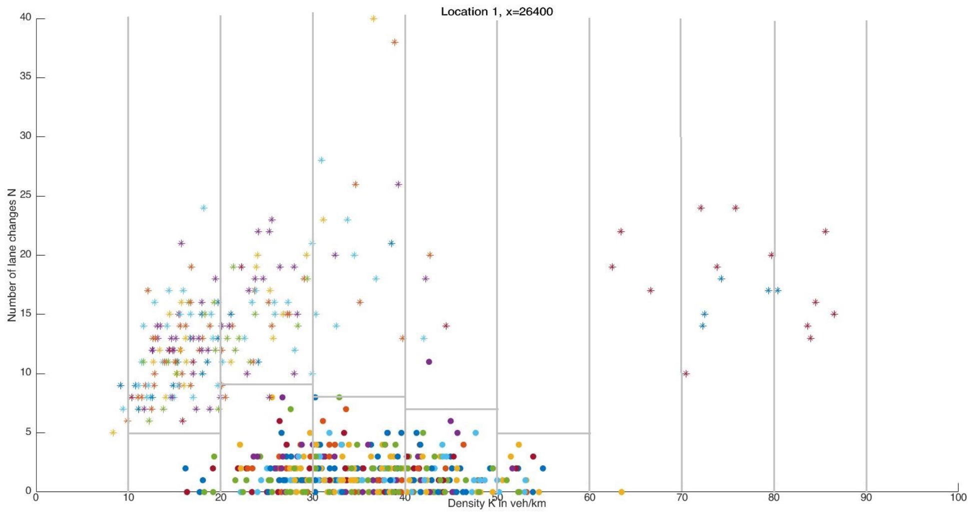


Figure 3.3: Results of an incident measured on location x=26400-26600; dots are scenario without incident, stars of scenario with incident. Vertical grey lines display the intervals in K for which the thresholds are made, the horizontal grey lines are the thresholds based on historical data without incident.

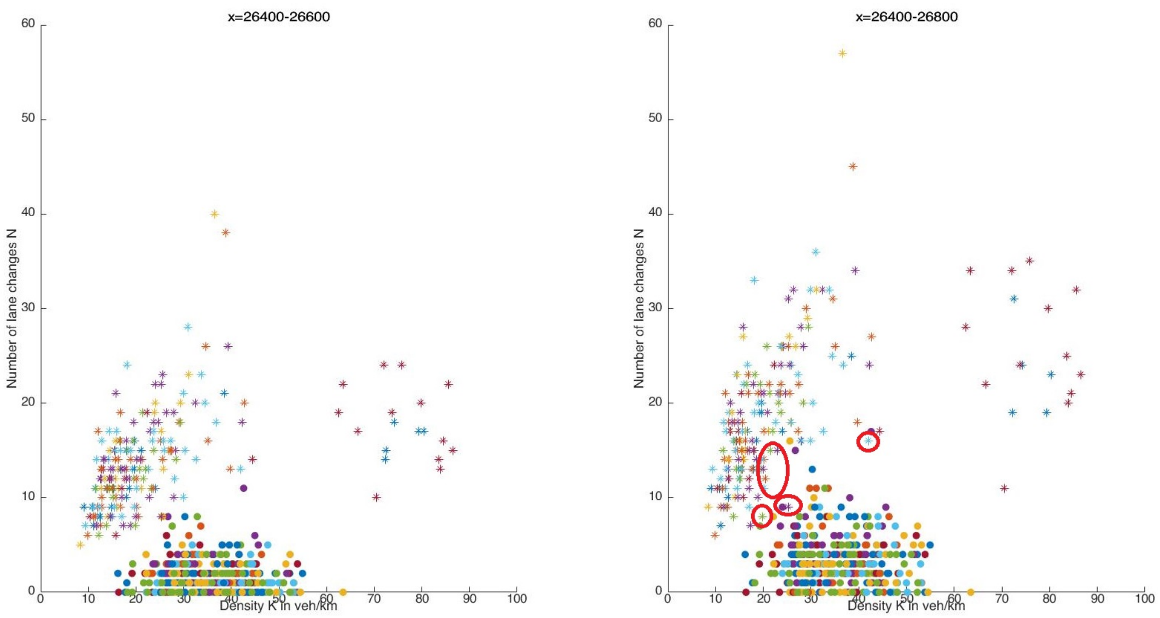


Figure 3.4: Results for a road section of 200 meter (left) and 400 meter (right).

3.2.4. Algorithm for lower penetration rate of FCD

There is no full coverage of floating car data on the roads at this moment and probably will not be available within the coming ten (or more) years. So the algorithm is adjusted to detect incidents with lower penetration rates. Especially the thresholds for the number of lane changes have to be adjusted, since the number of measured cars will highly influence the measured number of lane changes. So a special threshold for each penetration rate has to be made. The method of threshold determination is the same, for each road section and interval of density K a confidence interval plus a safety limit. But now also for a specific penetration rate P, see equation 3.7.

$$T_P(X,K_i) = \begin{cases} ceil(t_{u,P}(X,K_i)) + 2 & \text{if } t_{u,P}(X,K_i) < 10 \\ ceil(t_{u,P}(X,K_i)) + 4 & \text{if } 10 \geq t_{u,P}(X,K_i) < 20 \\ ceil(t_{u,P}(X,K_i)) + 8 & \text{if } t_{u,P}(X,K_i) \geq 20 \end{cases} \tag{3.7}$$

In the case you measure n_i cars (thus only cars with FCD) on a road section in one minute, the penetration rate P is not given. But the loop detector is measuring the flow Q_i. So from n_i and Q_i the penetration rate can be calculated and the right threshold values can be used to detect the possible incident.

$$P = \frac{n_i}{Q_i} \tag{3.8}$$

So this new algorithm aims to detect incidents based on an increase in the number of lane changes in case of an incident. This concept will be tested in a simulation and compared with the McMaster algorithm. In the next chapter the set-up of the simulation and the different scenarios in this simulation will be explained.

4 Research methodology for the evaluation of Incident Detection Algorithms

To examine the performance of the new Automatic Incident Detection algorithm and the already existing algorithm 'McMaster', an investigation is set up. The research will consist of a simulation study. Main reasons for choosing a simulation are the feasibility and adaptability. The floating car data (FCD) and loop detector data can be generated by the simulation, different scenarios for the penetration rate of FCD can be tested easily and incidents can be researched safely in a simulation.

The results of the new algorithm will be compared to the results of the McMaster algorithm in the next chapter 'Results of Incident Detection algorithms' performance'. In this chapter the set-up of the simulation is discussed, with the input and output of the simulation and the scenario's that will be researched.

4.1. Simulation scope

Most important for the simulation is to create a realistic traffic situation to test the AID algorithms. In section 2 is seen that the detection of incidents is easier in case of a high flow and more difficult in case of a low flow. Both cases will be tested. In dialogue with Rijkswaterstaat the criteria of a highway in the Netherlands is added. Below the used criteria for selecting a suitable research area:

- A motorway stretch in the Netherlands, because loop data and infrastructure design information is available for these roads;
- Congestion in the morning or evening peak on the chosen road section, not necessarily on all days of the week;
- Free flow conditions in a certain period on the chosen road section;
- No complicated infrastructure design like weaving sections and parallel carriageways.

To simulate a realistic traffic pattern on a realistic road, there is chosen to use the road lay out from a Dutch motorway. The Dutch motorway A27 from interchange Lunetten at Utrecht to interchange Gorinchem complies with above criteria. The chosen research area is visible in figure 4.1.

To get an realistic traffic pattern, the input of the simulation will be based on real traffic data on the chosen motorway A27. The days and times of the day to research are selected with following criteria:

- A regular working day, so no holiday, vacation or weekend;
- A day of the week with congestion.

Figure 4.1: Reseach area in blue: A27 from interchange Lunetten to interchange Gorinchem

For above reasons the traffic data of Monday 3 October till Friday 7 October is researched to choose two suitable time/place slots, one time/place slot with congested conditions and one slot with free flow conditions. Different locations are checked since the input for the simulation is collected at a different measurement location than the location of the traffic jam will be. For all locations the average flow per minute in veh/h and the speed per minute in km/h is researched.

In subsection 4.2.3 the simulation with this input is calibrated.

4.2. Simulation set-up

First step is to chose a suitable simulation program, knowing the limitations of the different options. After a short explanation about the selection of FOSIM, the input and output of this program is stated. Concluding this section with the calibration of the simulation model.

For the traffic simulation there were different possible programs to use, with their own advantages and disadvantages. A few that were considered to use:

- FOSIM: easy to use for inexperienced programmers, input can be given in a (no so friendly) user interface. Not all characteristics of the simulation can be adjusted manually, for example the driver characteristics are programmed in the back end of the program. However there is chosen for mean Dutch driver characteristics, so that will hopefully fit the characteristics of the drivers at the A27 at the chosen days.

- VISSIM: easy to use for inexperienced programmers, mostly used for micro simulations on smaller scale that current research area. Not all characteristics can be adjusted manually. Availability of the program can be a problem since its license is expensive and limited in the university.

Since my experience with FOSIM and that the options of FOSIM are extensive enough, the simulation will be done in FOSIM. The limitation of this program has to be taken into account in the interpretation and comparison of the results.

In the user manual of FOSIM, some important considerations are mentioned which has to be kept in mind when using this program and to interpret the results wisely. Some relevant considerations are stated below [13].

- Calibration parameters for vehicle types are defined, the five types (representing the different vehicle types from passenger cars to lorries and big trucks) needs to be in every simulation.

- One simulation is not a basis for realistic results of the traffic stream since the simulation in FOSIM is partly dependent on drawing of lots. So different runs per simulation are needed with a different start value.

- There are some restrictions for the simulated road sections:
  - The effect of vertical alignment is not simulated for the traffic settlement.
  - The effect of arcs in the road lay-out can not be simulated.
  - The effect of narrow lanes can not be simulated.
  - FOSIM is using ideal circumstances (in terms of for example weather).

The considerations are used in the set-up of the simulation and in the interpretation of the results. The five vehicle types are used in the simulation and ten runs will be executed for each simulation scenario. The first three restrictions for the simulated road sections will not have a significant influence on the results of the simulation. The last mentioned that FOSIM is using ideal circumstances has to be kept in mind when researching the results of the simulation and the results of the algorithms.

4.2.1. Input FOSIM

This program needs different input parameters, they are discussed concisely below.

- Road layout: lanes, on and off ramps. Number of lanes at location and the location of on and off ramps gathered via the application Mapviewer version 2.4.14 from Rijkswaterstaat. Between Houten and Hagestein the left lane is a peak hour lane and only opened during peak hours.

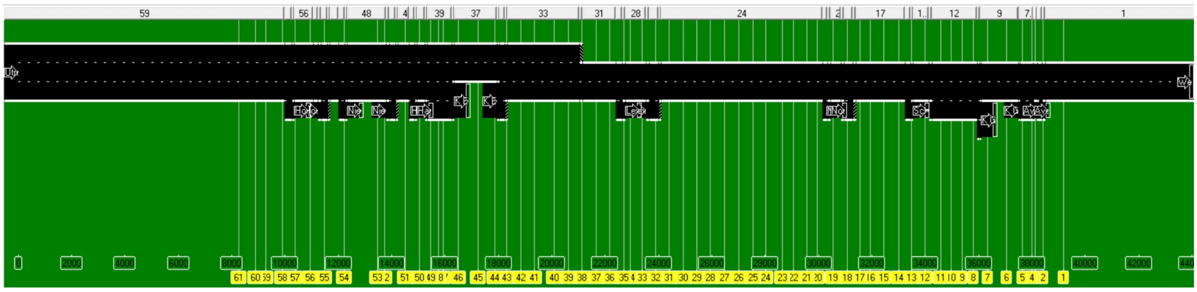


Figure 4.2: Road layout in FOSIM with detectors

- Lane change areas: at the end of an on ramp and when the number of lanes decreases. If possible there is 600 metres mandatory lane change area and 600 metres wished lane change area at the end of a lane.

- Speed limit: for each road section a maximum speed. From Lunetten to Everdingen 100 km/h and from Everdingen to Gorinchem 120 km/h (the actual speed limit is 130 km/h, but this is not a valid input according to FOSIM). When the peak hour lane between Houten and Hagestein is opened, the speed limit is lowered to 80 km/h.

- Detector locations. Gathered from the application Mapviewer version 2.4.14 from Rijkswaterstaat. The detector locations can be found in table 4.1. The given detector numbers are corresponding with the detector labels in FOSIM. The given Δx is the difference in distance from the detector to the downstream detector. For this road stretch the average distance between the detector locations is 507.25 meter.

- Traffic composition: percentage of freight traffic and passenger traffic. In the simulation 10% freight traffic and 90% passenger traffic.

- Flow: given per origin per minute displayed in veh/h in graph 4.3.

- Origin-Destination matrix: the destination is given per origin in percentages, see table 4.2.

- Vehicle parameters: for each vehicle type (five in total) the vehicle type specific parameters can be adjusted. The standard parameter values are used.

- Temporary blockade: a temporary blockade can be used to simulate for example an accident.

Table 4.1: Detector locations as input for FOSIM

Detector nr	Hmp	Δx	Detector nr	Hmp	Δx
41	65.739	745	20	50.070	500
40	64.994	288	19	49.570	519
39	64.706	754	18	49.051	528
38	58.952	550	17	48.523	516
37	58.402	413	16	48.007	507
36	57.989	283	15	47.500	529
35	57.706	201	14	46.971	524
34	57.505	556	13	46.447	546
33	56.949	749	12	45.901	483
32	56.200	617	11	45.418	618
31	55.583	430	10	44.800	415
30	55.153	553	9	44.385	487
29	54.600	502	8	43.898	400
28	54.098	746	7	43.498	607
27	53.352	520	6	42.891	512
26	52.832	522	5	42.379	471
25	52.310	517	4	41.908	410
24	51.793	523	3	41.498	500
23	51.270	524	2	40.998	549
22	50.746	226	1	40.449	
21	50.520	450			

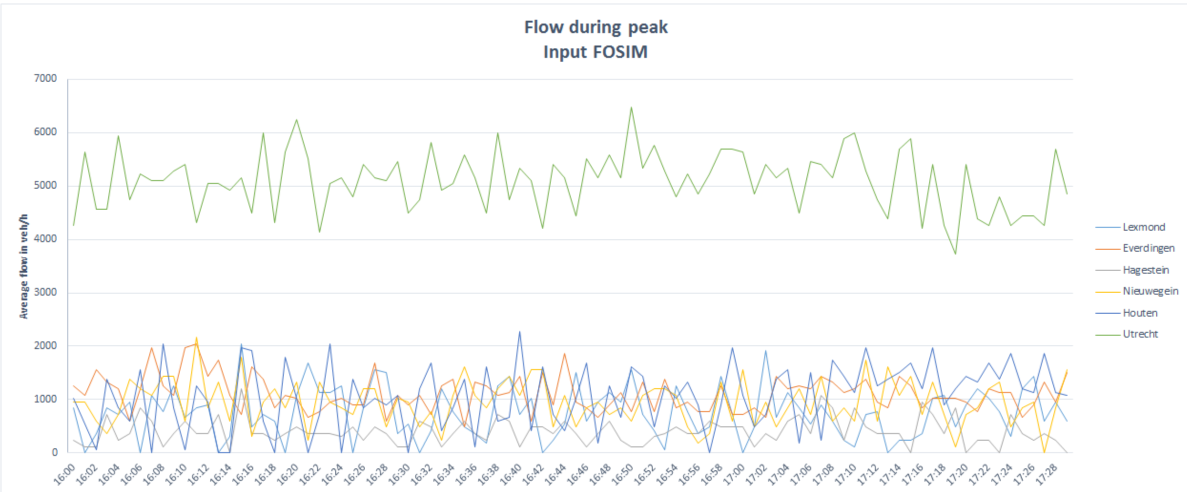


Figure 4.3: Input for FOSIM, flow during peak 16:30-17:30

Table 4.2: ODmatrix as input for FOSIM

O\D	UT	HT	NG	HS	ED	LM	ND	G
Utrecht (UT)	x	20.0%	14.0%	8.0%	20.0%	8.0%	1.0%	29.0%
Houten (HT)	-	x	18.0%	10.0%	24.0%	9.0%	1.0%	38.0%
Nieuwegein (NG)	-	-	x	12.0%	28.0%	10.0%	1.0%	49.0%
Hagestein (HS)	-	-	-	x	32.0%	11.0%	1.0%	56.0%
Everdingen (ED)	-	-	-	-	x	12.0%	1.0%	87.0%
Lexmond (LM)	-	-	-	-	-	x	1.0%	99.0%
Noordeloos (ND)	-	-	-	-	-	-	x	100.0%
Gorinchem (G)	-	-	-	-	-	-	-	x

4.2.2. Output FOSIM

The simulation program FOSIM gives the following output with following information which will be used for this research.

- Micro detection output; every line in the data is a new data point of a passing vehicle at a certain detector. Given information per data point:
  - Position of detector in m;
  - Lane;
  - Time in s;
  - Speed in m/s;
  - Vehicle type;
  - Vehicle ID;
  - Destination.
- Lane change output; every line in the data is a lane change. This data will be used for the new algorithm. Given information per data point:
  - Time in s;
  - Lane change from lane;
  - Lane change to lane;
  - Position of lane change in m;
  - Vehicle type;
  - Vehicle ID.

4.2.3. Calibration of the simulation model

To validate the simulation model, the results are compared with data from the NDW. Especially for the incident locations, which are elaborated in subsection 4.4.

Note that a realistic traffic pattern is needed for realistic results. The simulation does not need to be the same as the data form the NDW, but this data will give an example of a realistic traffic pattern.

In figure 4.4 is visible that the FOSIM results at the chosen detector locations are indeed congested at detector 27 and free flow at detector 15. The traffic data from the NDW (National Data Warehouse) is compared for a chosen time interval when the road is congested and for another chosen time interval in free flow traffic state. Especially the flow and speed on the two research locations is investigated: hmp 53.4 (detector 27) and hmp 47.5 (detector 15).

In figures 4.5 and 4.6 both the results of FOSIM and the NDW data is plotted.

For the speed is clear that the simulation in FOSIM needs time to set-up. The simulation represents the real data from around 17:10. The flow graph of FOSIM is whimsical, but fits the NDW data quite well.

The same research is done for detector 15 in figures 4.7 and 4.8. FOSIM seems to have a lower set-up time for this location, the graphs fit already after 20 minutes. Notable is the lower peak in the graph of NDW data, which is not shown in the graph of the FOSIM data. Presumably this is a failure of the detector or an irregularity in traffic behaviour.

So the results of the simulation of FOSIM is a realistic traffic pattern and can be used for this research.

The result of this investigation is the simulation time period of 16:30-17:50. Which include a simulation set-up time of 40 minutes and the research time period of 17:10-17:50. In the scenarios with an incident, see below, the incident will happen at 17:30.

Data assimilation Since FOSIM only provides simulation data and cannot apply the algorithms and thus lead to our final results, a second program, Matlab, is used for implementing the algorithms and calculate the results.

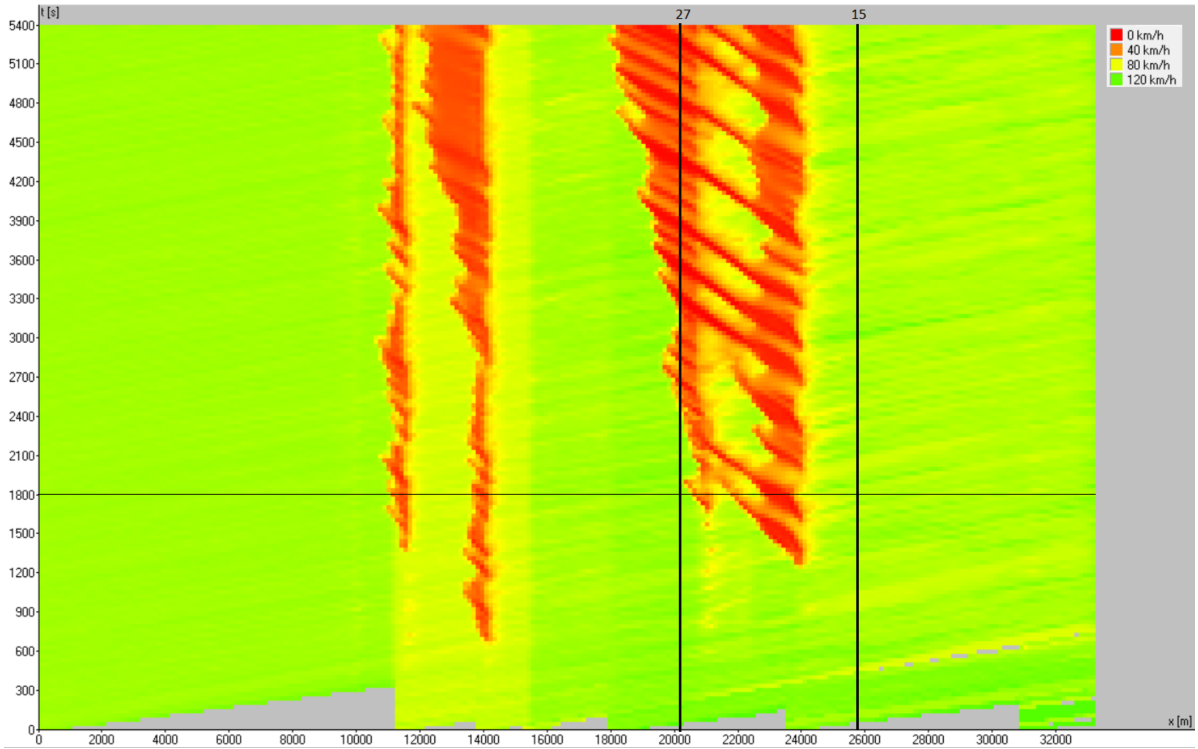


Figure 4.4: Flowdiagram as output of FOSIM for simulation of 4 October 16:00-17:30. Detector locations 27 and 15 are indicated as well as the set-up time needed by FOSIM till 1800 s

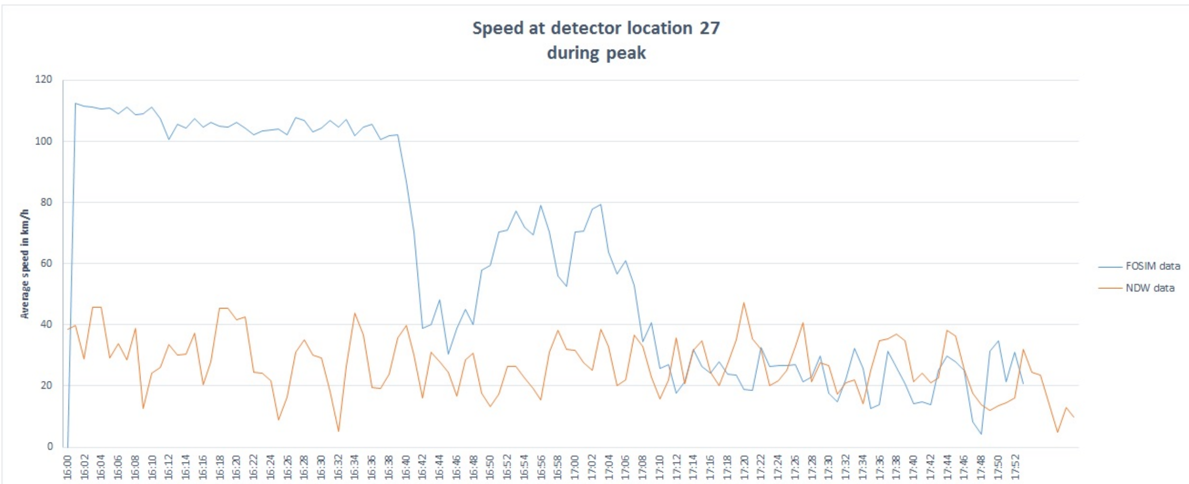


Figure 4.5: Speed FOSIM between 16:00 and 17:30 at hmp 53.35, detector 27 in FOSIM

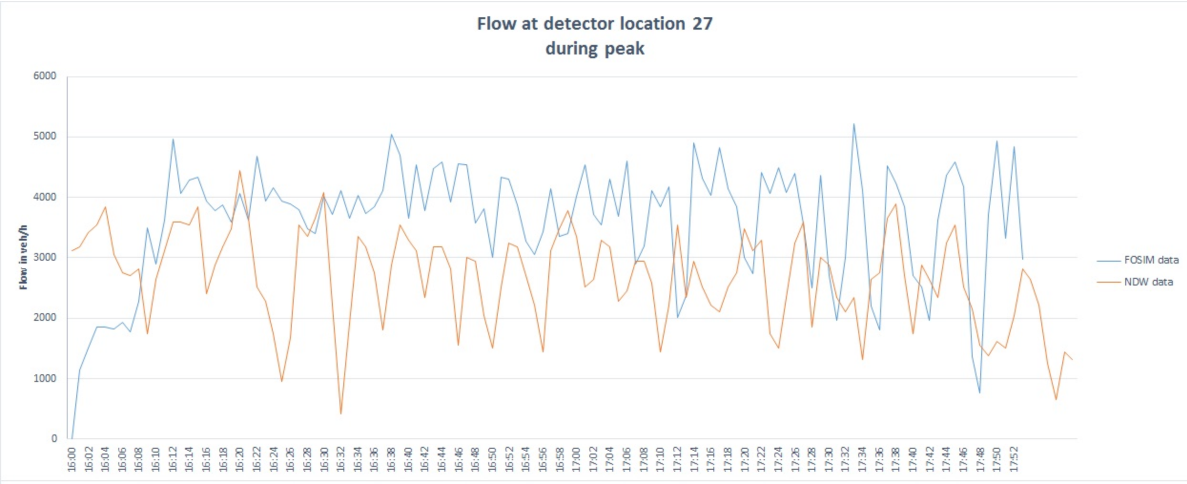


Figure 4.6: Flow FOSIM between 16:00 and 17:30 at hmp 53.35, detector 27 in FOSIM

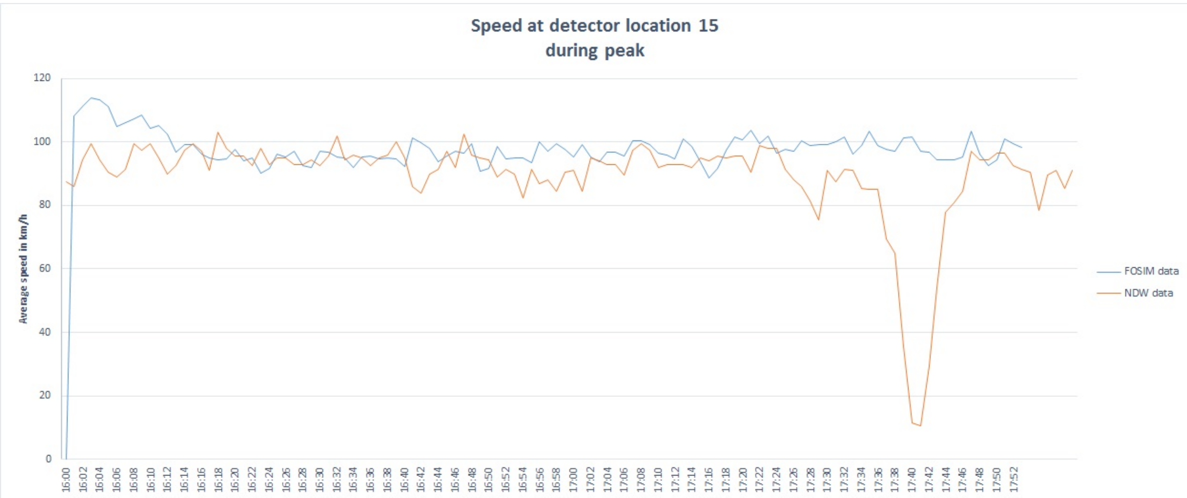


Figure 4.7: Speed FOSIM between 16:00 and 17:30 at hmp 53.35, detector 15 in

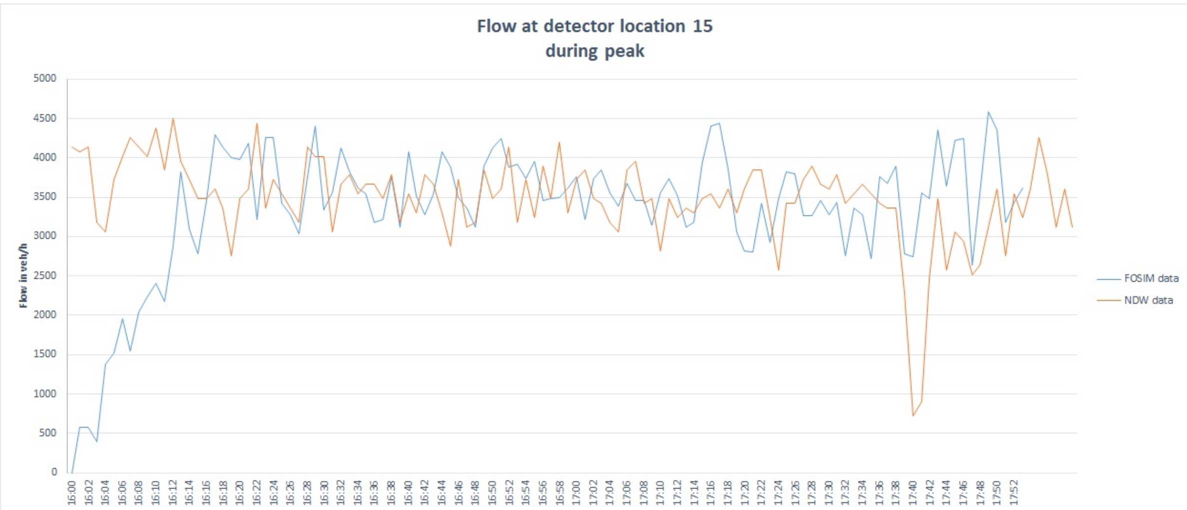


Figure 4.8: Flow FOSIM between 16:00 and 17:30 at hmp 53.35, detector 15 in FOSIM

4.3. Data

Both loop data and floating car data will be used by the researched algorithms, below a short explanation of the data that will be processed by the algorithms.

Loop data The simulation in FOSIM needs flow as input. The flow data is obtained by loops. This loop data is gathered via the National Data Warehouse (NDW). For precision the constraints for this data are:

- Loop data for all real measurement locations on the chosen road section;
- Loop data given per carriageway;
- Data contains mean flow per minute.

Floating Car Data The Floating Car Data is generated by the FOSIM itself. Especially lane changes are needed for the new algorithm. FOSIM can generate lane changing data for all vehicles and does not descretise this data in time neither in space.

4.4. Scenarios

In this simulation study 4 different scenarios are evaluated. These scenarios corresponds with four locations: a simple road stretch (road section 1) and a road section with a lane drop (road section 2), both contains a scenario with an incident on the left lane (scenario 1 and 3) and a scenario with an incident on the right lane (scenario 2 and 4). These four scenarios are drawn up to examine the algorithms in different circumstances. The variables in the scenarios are explained below. In these scenarios different variations are made:

- Incident location. Position relative to the loop detector: part *a* and b of each scenario. These distinction is only made for the McMaster algorithm, since the distance from the incident location to the first upstream detector will influence the performance of the algorithm. For the new algorithm only location *a* is taken.
- Traffic conditions. Free flow conditions on road section 1 (and thus in scenario 1 and 2) and congested conditions on road section 2 (and thus in scenario 3 and 4).
- Penetration rate of the floating car data: 100%, 50%, 20% and 10%. The penetration rates 50%, 20% and 10% are only tested for scenario 1a and 4a to get an indication of the performance of the new algorithm for lower penetration rates.

See figure 4.9 for a graphic view of the scenarios

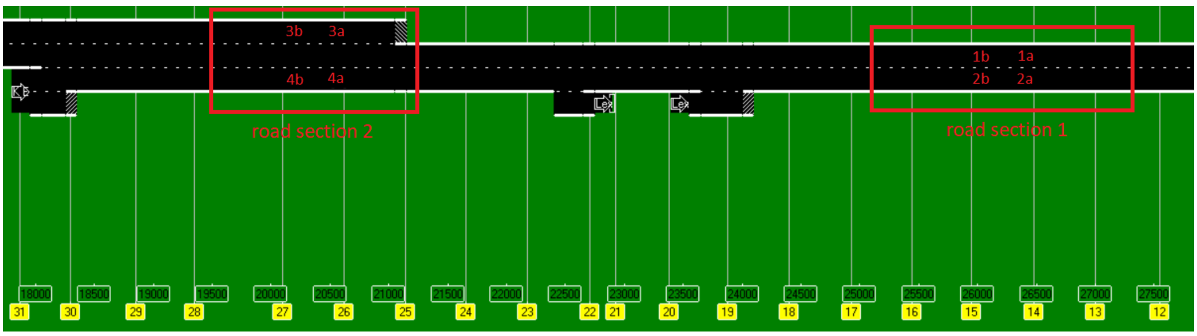


Figure 4.9: Road section downstream between interchange Everdingen and off ramp Noordeloos. With the two road sections and four scenarios.

In case of the new algorithm, the scenarios are first ran without an incident for setting a threshold in the number of lane changes to distinguish the incident. Above scenarios are ran in FOSIM, thereby FOSIM will make 10 runs per simulation scenario to capture differences in traffic behaviour.

The penetration rate will be executed in Matlab before the calculation of the algorithms. A random draw at the size of the penetration rate is taken. Since different random draws can influence the results of the algorithm, 10 different random draws are taken for each penetration rate.

4.5. Key performance indicators

The performance of the algorithms are judged by using three key performance indicators:

- The false alarm rate: rate of data point that are detected as an incident but are not. Calculated as the false alarms relative to the total alarms;

$$\text{false alarm rate} = \frac{\text{number of false alarms}}{\text{number of total alarms}}$$

- The detection rate: rate of incidents that are detected. Calculated as the detected incident relative to the total incidents;

$$\text{detection rate} = \frac{\text{number of detected incidents}}{\text{number of total incidents}}$$

- Detection time: time it takes to detect the incident from the moment it occurs in seconds.

These three parameters are mutually dependent. When the detection rate increases, often the false alarm rate increases as well. Vice-versa the detection rate will decrease if the false alarm rate decreases, since the sensitivity of the algorithm is decreased. When more time is taken to analyse the data, and thus a longer detection time, will regularly lead to improved values for the detection and false alarm rate [15].

Preferably the algorithms detect all incidents in a short time with a low false alarm rate. Since these parameters are mutually dependent, the perfect combination presumably does not exist. The detection of the incidents and the detection time are determined to be more important than false alarms in this case, since an incident needs to be detected for safety reasons. A quick detection is important because in this way the total travel time delay can be reduced.

Now the simulation is set up, the scenarios are clear and the key performance indicators are discussed, the results will be presented in the next chapter. This will be done based on the scenarios for the new designed algorithm discussed in chapter 3 and the for the McMaster algorithm.

5

Results of incident detection algorithms' performance

The results of the Master algorithm and the new algorithm will be discussed in this chapter according to the scenarios presented in the previous chapter. The incident locations are visible in figure 4.9. The results of the McMaster algorithm and the new algorithm will be discussed per incident location. Note that the results for the new algorithm for incident location *a* and *b* will be not significantly different, since the distance till the detector location is not an issue, so these are not tested separately in all scenarios.

As discussed in chapter 4, the new algorithm will be tested for a 100% penetration rate on all incident locations. An few extra tests with 50%, 20% and 10% penetration rate are done for respectively scenario 1 and 4. These results will give an indication of the usability of the new algorithm in the near future, since the current penetration rate is more realistic to be 10% and will grow in the coming years. These results are also compared to the results of the McMaster algorithm.

5.1. Results scenario 1: road section 1, left lane

First two scenarios on road with a simple lay out (road section 1): two lanes without any infrastructure changes. The road lay-out is given in figure 5.1.

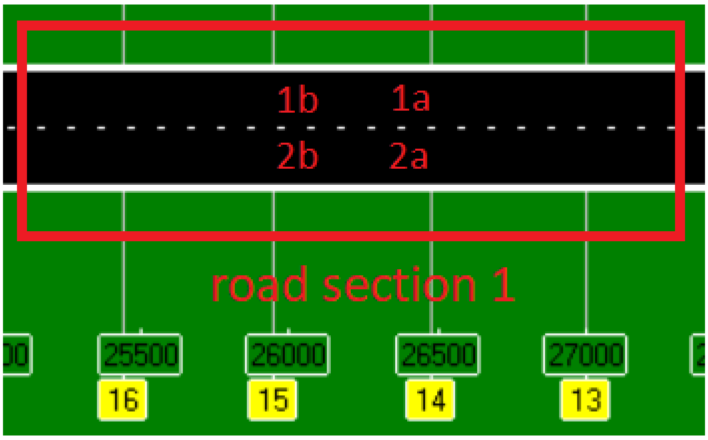


Figure 5.1: Lay-out of road section 1, with part *a* and *b* for the McMaster algorithm

The space-time diagram of an incident on location number 3 is displayed in figure 5.2. Clearly an queue is building up after the incident, so this will probably be detected by the algorithms.

Also lower penetration rates (50%, 20% and 10%) are tested on road section 1, the section is concluded with these results. The results of this first scenario are presented in table 5.1.

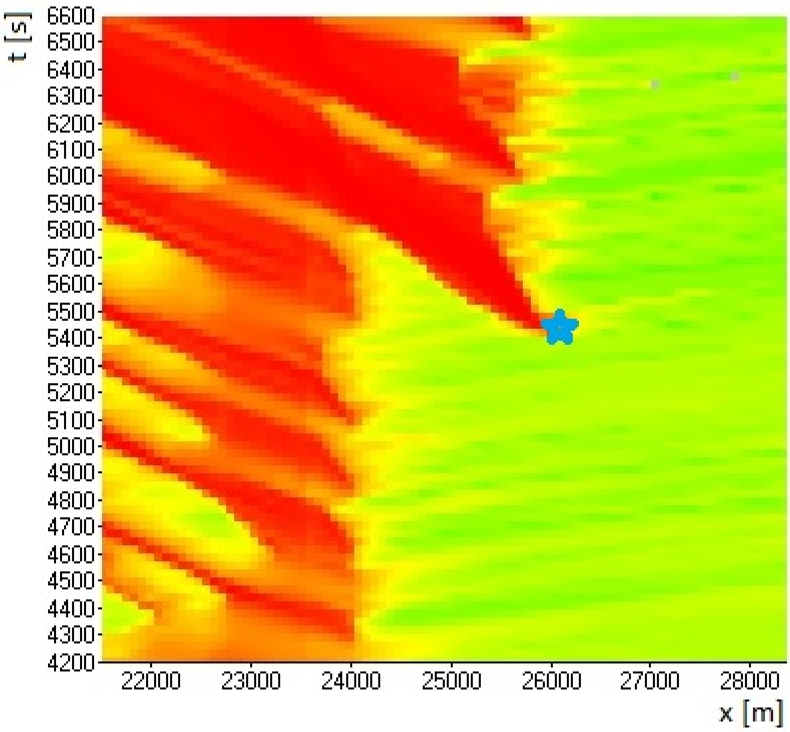


Figure 5.2: Space-time diagram of the traffic conditions with an incident on location 3, the time and location of the incident is indicated by the blue star.

Table 5.1: Results scenario 1: road section 1, left lane

Algorithm	Information	False alarm rate	Detection rate	Detection time [s]
McMaster	location *a*	0.1764	1.0	288.0
McMaster	location *b*	0.1853	1.0	180.0
New algorithm	100% FCD	0.0167	1.0	60.0
New algorithm	50% FCD	0.0367	0.9	126.6
New algorithm	20% FCD	0.0700	1.0	276.0
New algorithm	10% FCD	0.1967	0.9	284.0

McMaster algorithm, location a The first to notice for the performance of the McMaster algorithm on location *a* is the false alarm rate of more than 0.17, far more than the found false alarm rate in literature of less than 0.001. A closer look to the false alarms give that they are mostly located around the intersection 'Everdingen'. Possibly the algorithm have problems with the change in infrastructure at this point. Beside these concentrated false alarms around the intersection, there are some false alarms divided over all detector locations, somewhat higher around infrastructure changes than on a straight stretch.

All incidents at this location are detected by the McMaster algorithm, although it takes relatively long: more than four minutes. Important factor of this detection time is the location of the incident related to the detector locations. This incident is located just downstream of a detector, so it is detected by the next upstream detector which is around 500 meter away. The queue due to the incident has to build up, which takes time, but also has to grow until the upstream detector can measure this queue and detect the incident.

The occupancy-flow diagrams of four detectors are visible in figure 5.3. Remember that the incident is located just downstream of detector 28. As expected, the flow downstream of the incident is decreasing while the occupancy is still low: the yellow dots are located on the free-flow branch of the fundamental diagram. At the downstream of detectors the yellow dots are visibly located under the free-flow branch of the fundamental diagram. Note that the algorithm is using only data from the same minute at different detector locations and in this graph the data of multiple minutes is displayed. Clearly the incident is detected at the detectors downstream of the incident.

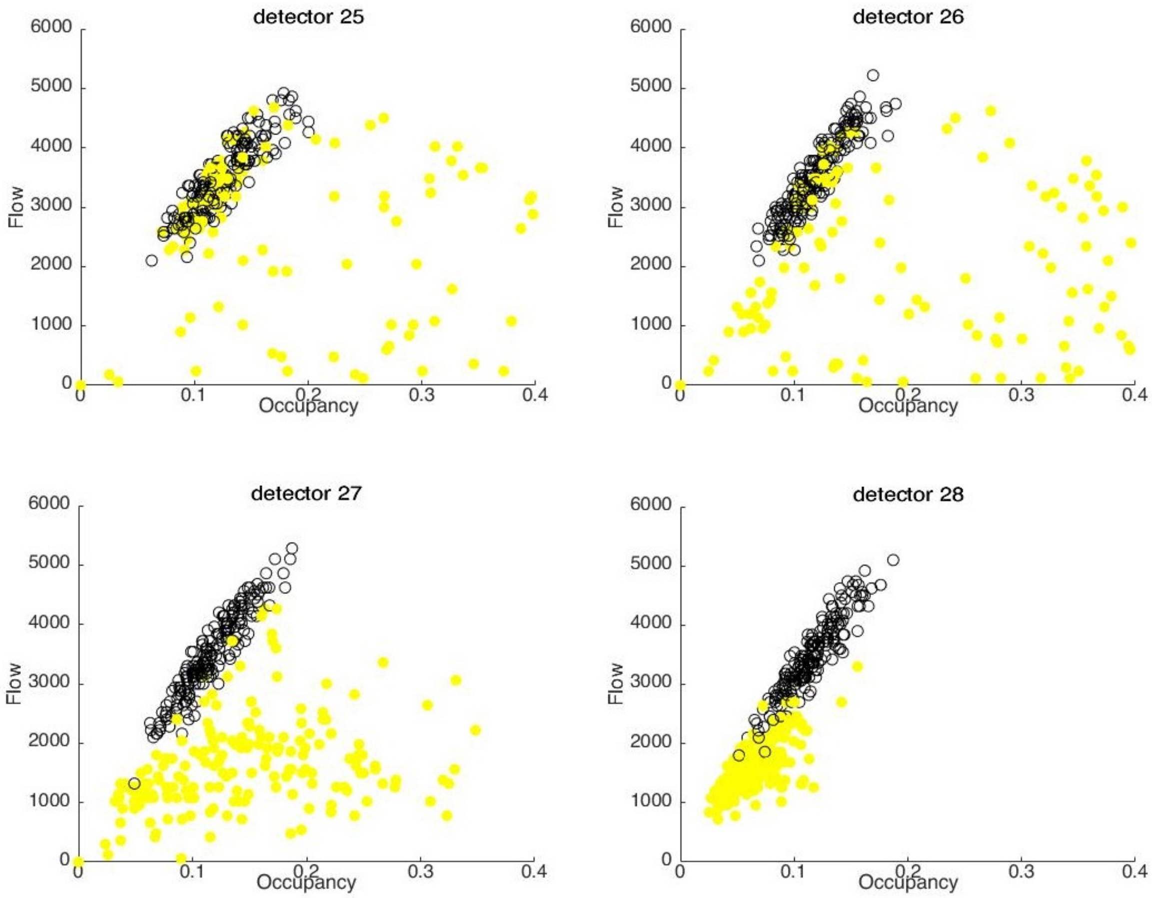


Figure 5.3: Occupancy-flow diagrams around incident location 1*a*; incident located between detector 27 and 28; black dots before the incident, yellow dots after the incident.

McMaster algorithm, location b The distance from the incident location till the first upstream detector is an important factor for the detection time. Therefor two scenarios have been tested with the incident location just downstream of a detector. This is not an issue in case of the new algorithm, so only the results of the McMaster algorithm are presented for these locations.

The detection time for the McMaster algorithm on location *b* immediately stands out, it is about 1.5 minutes quicker than the detection of the incident on location *a*. The false alarm rate and the detection rate are not significantly different from the other location with the McMaster algorithm. As expected the incident location relative to the detector location mainly influences the detection time.

New algorithm, 100% FCD Now the new algorithm is used to detect the incident at the same location as above, first for a penetration rate of 100% FCD. The false alarm rate of the new algorithm is much lower than the McMaster algorithm at the same location, only 0.0167 versus 0.1764. That indicates already that the set thresholds are high enough to prevent a lot of false alarms. Again all incidents are detected, but a lot quicker than the McMaster algorithm: 60.0 seconds. The detection time is the minimum possible value, since the incident is programmed to happen exactly at the beginning of a minute and the algorithm uses data per minute.

As seen in a sneak preview in chapter 3, but also in figure 5.4, the graph with density and the number of lane changes for the road section just downstream of the incident location. This figure also indicated that the incident can be detected by the number of lane changes, since the data cloud of an incident can be clearly distinguished from the data cloud without an incident.

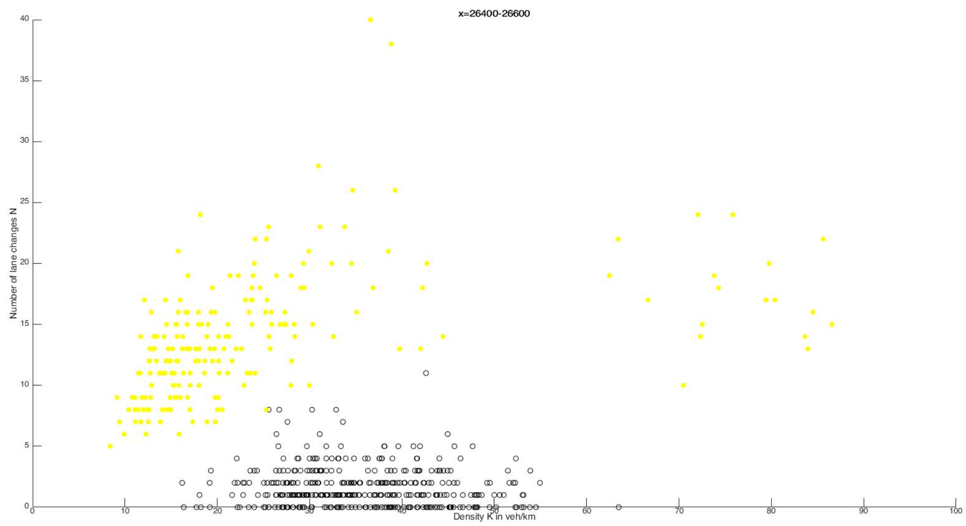


Figure 5.4: Graph results of incident in scenario 1 on location x=26400-26600; black dots before the incident, yellow dots after the incident.

Further research gives that the detection of the incident is mostly done in the road section one downstream of the incident. So the location of the incident can thereby be indicated within a range of 200 meters.

New algorithm, lower penetation rates The results of this first scenario with a limited penetration rate, now the penetration rate of FCD is halved to 50%. The false alarm rate seems to be doubled: from 0.0167 to 0.0367. Also the detection time is doubled now the penetration rate FCD is lower. The detection rate is 0.9 for this case.

These results can be explained by the greater variance in the number of lane changes when only measuring half of the vehicles (thus a penetration rate of 50%). Since only measuring half of the vehicles, there is a change of measuring no vehicles in a minute and thus no lane changes. Another possibility is measuring mainly vehicles that do not change lanes. Contrariwise there is a chance of measuring mainly vehicles that do change lanes. These outliers and its number will grow when the penetration rate is lower.

As expected the performance of the new algorithm is again decreased for a penetration rate of 20% FCD. Again the false alarm rate and the detection time increased compared to the 50% penetration rate. For this scenario there is a detection rate measured of 1.0. The difference with the 0.9 detection rate of 50% FCD can be declared by the stochastic nature of traffic. The expectation is that not always all incidents will be detected by this algorithm with a 20% penetration rate, but it is indicating that most incidents will be detected.

The last step in penetration rate is 10%, closer the current penetration rate FCD that is available nowadays. Remarkable is that the detection time is not decreased significantly compared to a 20% FCD, only 12.0 seconds. But again the false alarm rate increased, to almost 20%, which is very high for a detection algorithm.

For this first scenario the results for 10% FCD are comparable with the results of the McMaster algorithm. Therefor the new algorithm is recommended, even if only 10% FCD available. Although 20% penetration rate adds a added value in terms of the false alarm rate which is significant. A 50% penetration rate mainly add a much lower detection time. So measuring more vehicles do add an added value to the performance of the new algorithm.

5.2. Results scenario 2: road section 1, right lane

Now the results the right lane of road section 1 are presented, see figure 5.1 for the road lay out. The results can be found in table 5.2.

Table 5.2: Results scenario 2: road section 1, right lane

Algorithm	Information	False alarm rate	Detection rate	Detection time [s]
McMaster	location *a*	0.1825	1.0	234.0
McMaster	location *b*	0.1949	1.0	156.0
New algorithm	100% FCD	0.0200	1.0	60.0

McMaster algorithm, location a Incident location 2*a* is located at the same lateral distance as location 1*a*, but on the other lane of the two lane road. So the expectation is that these results are comparable to these of scenario 1, location *a*.

Indeed, the results are comparable with scenario 1. A false alarm rate just under 20% is found. All incidents are detected in on average four minutes.

McMaster algorithm, location b Same story for this scenario of the McMaster algorithm. The detection time is decreased to less than three minutes, again much shorter than the detection time of an incident on location *a*. Also in this case all incidents are detected and the false alarm rate is not significantly different of that on location *a*.

The performance of the McMaster algorithm on the left and right lane of road section 1 are comparable. Apparently the McMaster algorithm does not give significantly different results under these circumstances for an incident on the other lane.

New algorithm, 100% FCD For the new algorithm is only a 100% penetration rate tested for scenario 2. The expectation was that the new algorithm does not give very different results than is seen in scenario 1. The performance in terms of false alarm rate, detection rate and detection time turned out to be almost the same as scenario 1, with a detection rate of 1.0 for the tested incidents in the shortest possible detection time for this algorithm.

Compared to the McMaster algorithm, the new algorithm is performing much better for the false alarm rate and the detection time. Since the performance of the new algorithm for 100% FCD is comparable for the left and right lane, the results of the lower penetration rates FCD will also be comparable. So in this scenario, also the new algorithm is recommended for detecting incidents.

5.3. Results scenario 3: road section 3, left lane

The next two scenarios are executed on a part of the road with three lanes (road section 2), but with a lane drop downstream of the incident locations. The road lay out is given in figure 5.5. In figure 5.6 the space-time diagram for scenario 3*b* is displayed. On this location, there is already congestion at the time the incident occurs. The traffic conditions does change after the incident. The main issue for the algorithms is to detect the incidents while there is already congestion and to deal with the infrastructure change.

The results of this third scenario are presented in table 5.3.

Table 5.3: Results scenario 3: road section 2, left lane

Algorithm	Information	False alarm rate	Detection rate	Detection time [s]
McMaster	location *a*	0.1609	0.4	330.0
McMaster	location *b*	0.1557	0.0	-
New algorithm	100% FCD	0.0100	1.0	168.0

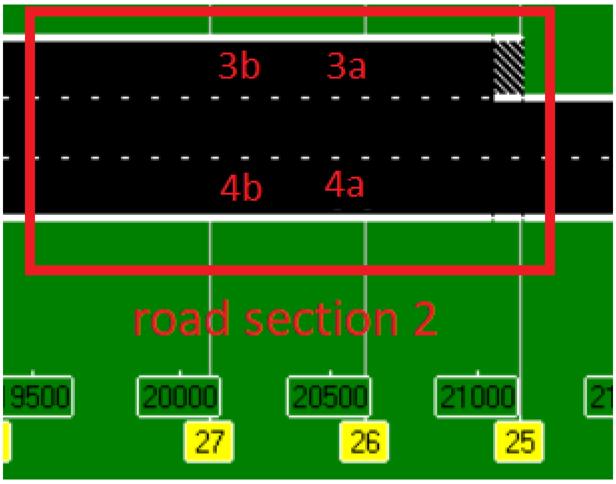


Figure 5.5: Lay-out of road section 2, with part *a* and *b* for the McMaster algorithm

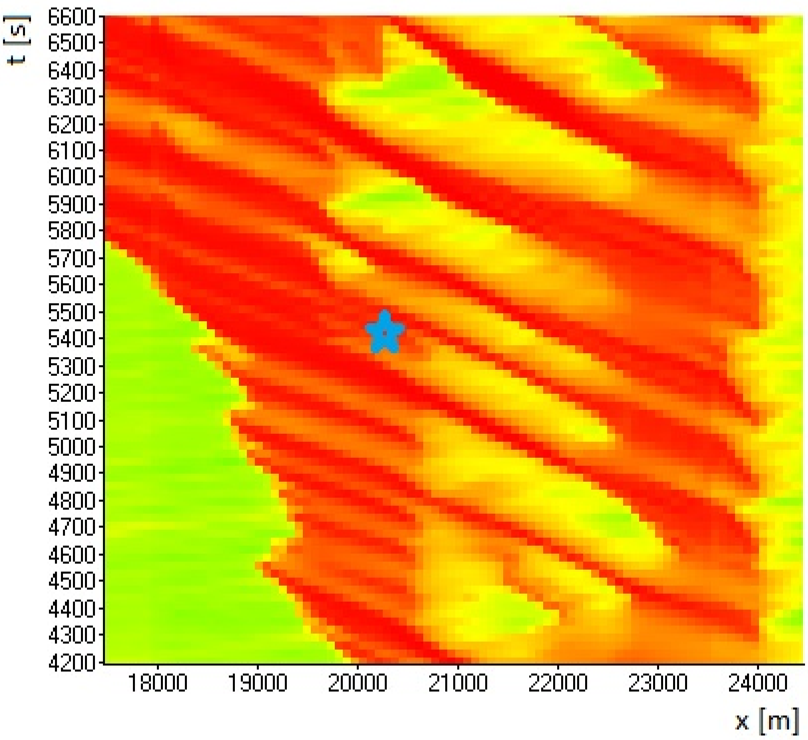


Figure 5.6: Space-time diagram of the traffic conditions with an incident on location *b*, the time and location of the incident is indicated by the blue star.

McMaster algorithm, location a The McMaster algorithm struggles more with this location in detecting incidents. The first results were promising, but a closer look gave more accidental detected incidents than by purpose detected incidents. Based on the theory of the McMaster algorithm from chapter 2, the expectation is that the incident will be detected in multiple consecutive minutes at multiple downstream detectors. This was only the case in 4 of 10 cases. The McMaster algorithm is not really detecting the incident in the other 6 cases. To distinguish an accidental detection with a certain detection, the condition of two consecutive detection signals at a detector is added. Although the queue do not have to build up to the upstream detector, the detection time is still quite long with 5.5 minutes.

A closer look to the occupancy-flow diagrams can help to explain the difficulties of the McMaster algorithm. These diagrams are shown in figure 5.7. The incident takes place between detector 15 and 16, where after the lane drop will occur. The traffic conditions are already congested for detector

15 till 18 before the incident, as visible in the graphs. But the main issue why the incidents cannot be detected is possibly the fact that $Q_i > q_{max}$ for quite a lot of yellow data points (data of minutes after the incident), so these points are in state 4, where just recurrent congestion is detected by this algorithm.

Based on the graphs, detection is most likely when a data point of detector 15 is in state 2 or 3 (meets at least condition $Q_i < q_{max}$, for detector 15 $q_{max} = 3750$) and the data point of the same minute at detector 16 is in state 1 or 2 (meets condition $OC_i < OC_{max}$; with in case of detector 16 $OC_{max} = 0.1095$). Note that detection is only the case if the data is from the same run and measured in the same minute, while in the graph the data of all ten runs and all minutes is displayed.

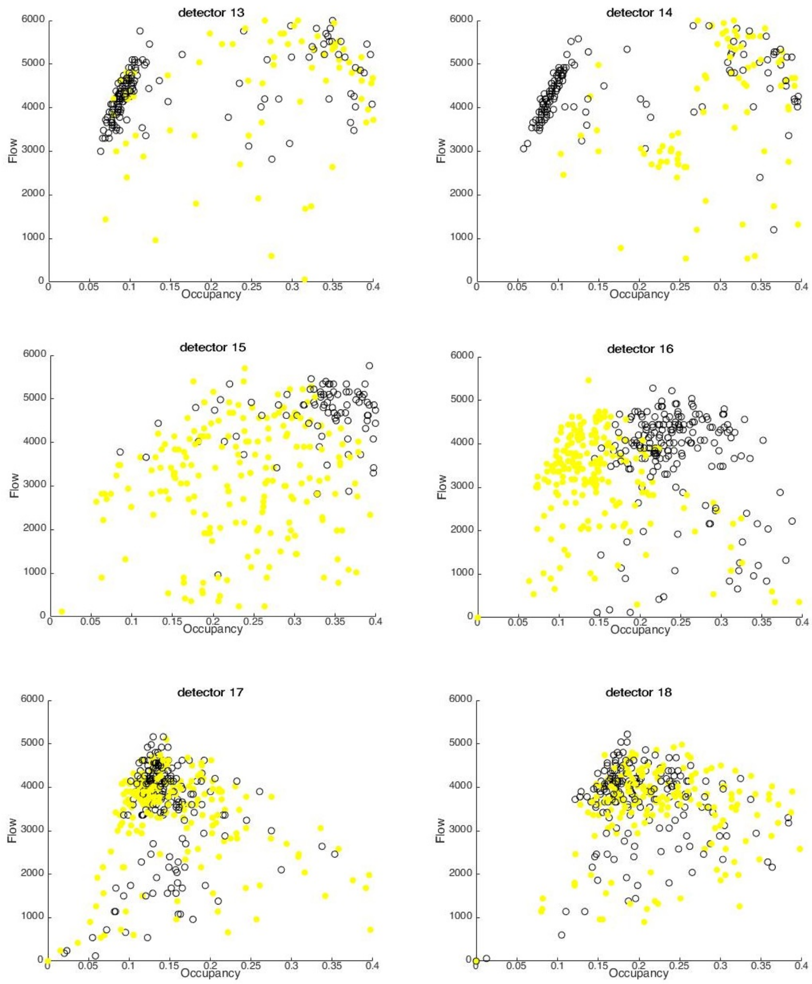


Figure 5.7: Occupancy-flow diagrams around incident location 5; incident located between detector 15 and 16; black dots before the incident, yellow dots after the incident.

McMaster algorithm, location b Again the locations more close to the nearest upstream detector is researched, named location *b*. But in this case the expectation is that this will less influence the detection time, since there is already congestion when the incident occurs. For this location, no incident has been detected by the McMaster algorithm. This is remarkable, since some incidents on location *a* were detected.

New algorithm, 100% FCD For the new algorithm the main challenge is the fact that the lane drop already causes many lane changes. It turns out to be possible to detect a difference in the number of lane changes in case of an incident according to the results presented in table 5.3.

The K,N-graphs are visible in figure 5.8, the incident is located in $x = 20200 - 20400$. Mainly in the road section $x = 20000 - 20200$ there is a difference between the yellow dots from the incident and the black dots without an incident. This road section turns out to be the one in which the incident is detected by a difference in the number of lane changes. In the first scenarios with a simple road lay out, it was the road section downstream of the incident, now it is the road section upstream of the incident. This is explained by the lane drop downstream of the incident, in this case the vehicles will not change back to the left lane. As seen in the graphs, there are even less lane changes in the downstream road section ($x = 20400 - 206000$).

So also in case of this particular road lay out, the location of the incident can be indicated by 200 meters, but in this case the incident has happened downstream of the detector which gives a signal.

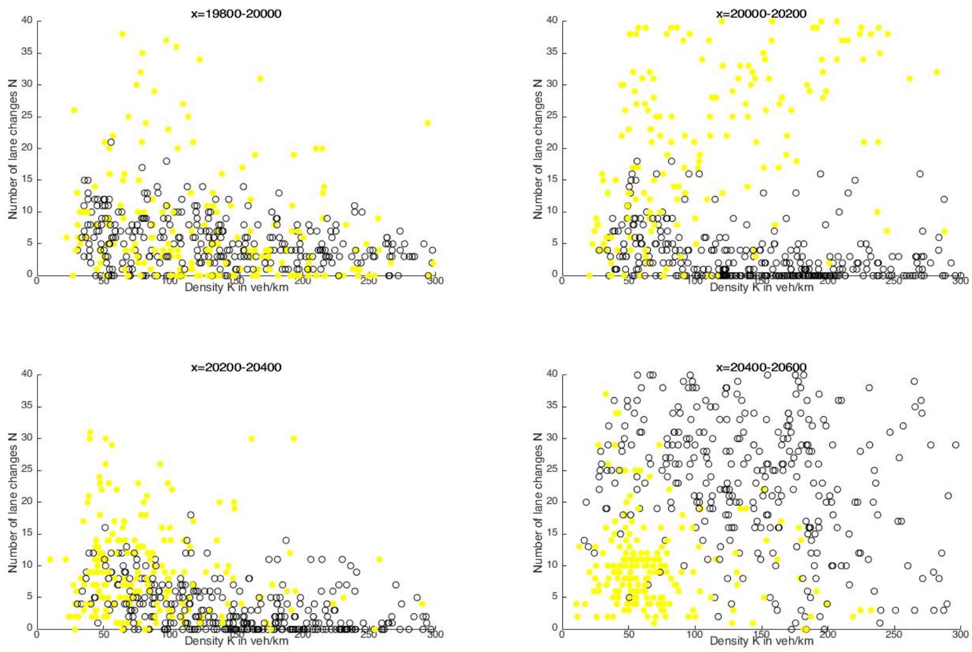


Figure 5.8: Results of incident in scenario 3 for the new algorithm; black dots before the incident, yellow dots after the incident.

For this road stretch with the lane drop and an incident on the left lane, the false alarm rate is still low and all programmed incidents are detected by the new algorithm. Although, the detection time is increased to 168.0 seconds. The increased detection time is mostly the consequence of a higher variance in the detection time of the separate runs.

For this scenario the new algorithm is detecting the incident more accurate than the McMaster algorithm on all three key performance indicators. Note that the new algorithm is using a 100% penetration rate FCD in the tested case, the lower penetration rates are tested for this second road section in scenario 4.

5.4. Results scenario 4: road section 2, right lane

Now the results the right lane of road section 2 are presented, see figure 5.5 for the road lay out. The results can be found in table 5.4, also for lower penetration rates.

McMaster algorithm, location a First that stands out is that an incident on the right is better detected by the McMaster than an incident on the left lane of road section 2 by the McMaster algorithm. Both the detection rate and the detection time have been improved compared to an incident on the left lane. Most likely the traffic is more disturbed in this case, so detecting differences in the occupancy-flow diagram is easier.

Not all incidents are detected by the McMaster algorithm, only 80% in an average time just over two minutes.

Table 5.4: Results scenario 4: road section 2, right lane

Algorithm	Information	False alarm rate	Detection rate	Detection time [s]
McMaster	location *a*	0.1747	0.8	127.5
McMaster	location *b*	0.1608	0.4	345.0
New algorithm	100% FCD	0.0267	1.0	264.0
New algorithm	50% FCD	0.0533	1.0	404.0
New algorithm	20% FCD	0.1033	1.0	420.0
New algorithm	10% FCD	0.1667	0.8	405.0

McMaster algorithm, location b In this case only 4 of 10 incidents are detected, but again the incident on the right lane is detected better than the incident on the left lane. Although the detection time increased with almost two minutes compared to an incident on location *a*.

New algorithm, 100% FCD Nevertheless the results of the new algorithm on the right lane are worse than on the left lane, like was the case for the McMaster algorithm. Especially the detection time did increase, but all incidents are still detected. Possibly less vehicles are changing from the middle to the right lane and vehicles change later from the left to the middle lane, so a less clear difference is present.

Although the performance is worse than the left lane, still all incidents are detected by the new algorithm. Again the detection time is much higher than on road section 1.

New algorithm, lower penetration rates Up to now a 100% penetration rate is used for the new algorithm, but currently there is no 100% penetration rate floating car data available. More realistic is a penetration rate of 10%. The penetration rate of FCD will grow in the coming years, therefor also a 20% and 50% penetration rate is researched. In this way the added value can be seen of the increased penetration rate in terms of performance of the new algorithm.

Since the new algorithm had some more trouble in detecting the incidents with a 100% penetration rate, more difficulties are expected in case of lower penetration rates.

Like the other location the false alarm rate is doubles now the penetration rate is halved to 50%. The detection time did increase from 264.0 to 404.0, respectively 100% and 50% penetration rate. Besides an almost double detection time, the false alarm rate did also become twice as big.

For 20% FCD, again the false alarm rate increased, to 0.1033. Remarkable is that the detection time is not increased significantly while it did raise in the previous step.

For only 10% penetration rate the false alarm rate is increased to 0.1667. This time the detection time is even decreased compared to 20% FCD. This can possibly be declared by the limited amount of test runs that are executed in the simulation, more runs are needed to validate these results.

For the scenarios with the lane drop in the road lay out, both algorithms have more difficulties in detecting the incidents. The McMaster algorithm struggles with detecting the incidents at all and the new algorithm has an higher detection time compared to the earlier road lay out ánd the McMaster algorithm.

The McMaster algorithm did detect 8 of the 10 incidents in on average 127.5 seconds. The new algorithm is detecting 8 of 10 incidents in on average 405.0 seconds for an penetration rate of 10% FCD. So the McMaster algorithm is better for this specific location while the penetration rate of FCD is still low.

Note that the new algorithm was slower for all scenarios on this road stretch. But for other locations on the road stretch, especially location 7, the detection rate was really bad for the McMaster algorithm. So it is wise to use the new algorithm instead, especially when the detection time is less important than the detection rate, which is in many cases.

An overview of all scenarios and their results are displayed in table 5.5, with the scenarios and corresponding road lay out in figure 5.9.

Now the results of all scenarios are presented, the main research question can be answered. But first the results are discussed critically in the next chapter.

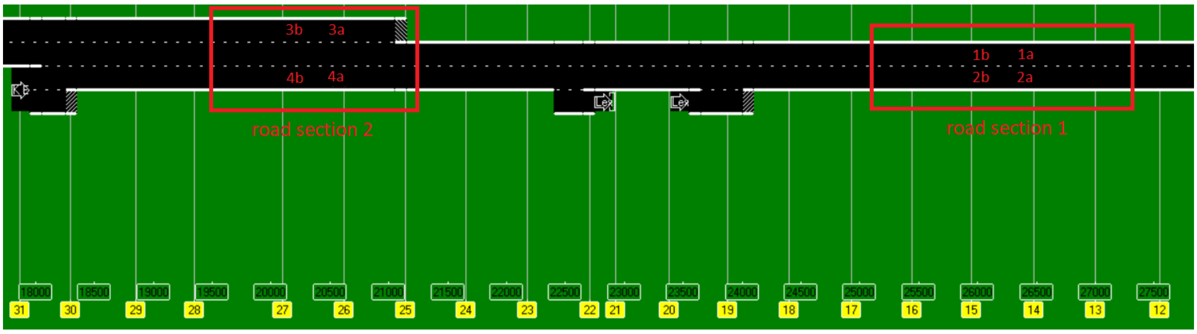


Figure 5.9: Road lay out with the two road sections and four scenarios.

Table 5.5: Overview scenarios and their results

Scenario	Algorithm	Information	False alarm rate	Detection rate	Detection time [s]
1	McMaster	location *a*	0.1764	1.0	288.0
1	McMaster	location *b*	0.1853	1.0	180.0
1	New algorithm	100% FCD	0.0167	1.0	60.0
1	New algorithm	50% FCD	0.0367	0.9	126.6
1	New algorithm	20% FCD	0.0700	1.0	276.0
1	New algorithm	10% FCD	0.1967	0.9	284.0
2	McMaster	location *a*	0.1825	1.0	234.0
2	McMaster	location *b*	0.1949	1.0	156.0
2	New algorithm	100% FCD	0.0200	1.0	60.0
3	McMaster	location *a*	0.1609	0.4	330.0
3	McMaster	location *b*	0.1557	0.0	-
3	New algorithm	100% FCD	0.0100	1.0	168.0
4	McMaster	location *a*	0.1747	0.8	127.5
4	McMaster	location *b*	0.1608	0.4	345.0
4	New algorithm	100% FCD	0.0267	1.0	264.0
4	New algorithm	50% FCD	0.0533	1.0	404.0
4	New algorithm	20% FCD	0.1033	1.0	420.0
4	New algorithm	10% FCD	0.1667	0.8	405.0

6
Synthesis and discussion

6.1. Synthesis

In general the new algorithm is better in detecting the tested incidents than the McMaster algorithm. As expected the McMaster algorithm does have more problems when there was already congestion at the incident location and time. The difference between incident locations 1-4 and 5-8 where expected for the McMaster algorithm. Remarkable is the bad performance in scenario 11, with incident location 7. In this case the algorithm did not detect any of the incidents.

Not only the McMaster algorithm has some difficulties for the incident locations 5-8, also the performance of the new algorithm is lower than for locations 1-4. Probably the lane change behaviour in congested conditions does not vary that much compared to congested circumstances with an incident. Also the lane drop can have an effect on the performance of the new algorithm. The expectation was indeed that the new algorithm would have some difficulties at this location, happily it did detect all incidents.

Also lower penetration rate for the new algorithm are researched, 50%, 20% and 10% FCD. As expected the performance of a lower penetration rate is worse, especially in terms of the false alarm rate. Remarkable is that the detection rate of scenario 13 (50% FCD) is lower than scenario 14 (20% FCD), probably because of the stochastic variance of the traffic and thus the results. Also the detection time is not a consistent ascending series when the penetration rate decreases, especially in scenario 16-18. While a longer detection time was expected for a lower penetration rate.

6.2. Discussion

This section is used to have a critical view on the research and the consequences on the found results.

The most important critical note is usage of a simulation for this research. For the circumstances it is the best choice, but it does come with some restrictions. One of them is that five vehicle types are defined, while there are a lot more different vehicles in real traffic. Another is the fact that the used program FOSIM is using ideal circumstances, which is also not a realistic approach for the real traffic. This means that the data produced by FOSIM is not hundred percent realistic, but these small deviations will probably not influence the results significantly.

But most important for the performance of the new algorithm is the programming of lane changing in FOSIM. In case of a change in road lay out or an incident that required a lane change, the program is adding a 600 meters long section of mandatory lane changing and 600 meters long road section of desirable lane changing. Which in fact means that you are literally programming vehicles to change lanes and in the new algorithm you are measuring that they are changing lanes. This is highly influencing the performance of the new algorithm. The results of this research can therefor only be seen as an first indication that incidents can be detected by comparing the number of lane changes to data without incidents.

Another critical note is that lane changes are assumed to be gathered from floating car data. At this moment the spatial accuracy of FCD is not accurate enough to measure lane changing correctly. This will be the case in close future, probably with a higher penetration rate and thus a better performance of the new algorithm.

The new algorithm uses density of a road stretch of 200 meters, which is measured at the beginning of the road stretch. In reality there are no detectors to measure the density each 200 meter. An solution would be to use the already existing detectors which are located on average each 500 meter. Disadvantage is that in this case no detectors can be removed, a current wish of the road authorities.

Note that the tested 100% penetration rate of the floating car data will never be reached, since always some devices will be broken, switched of or otherwise unavailable. But the scenarios with the 100% penetration rate does indicate that an incident can be detected by using lane changing data.

Although these limitations of the research, this research does indicate that using floating car data, and in particular lane change detection is useful for incident detection.

7

Conclusions and recommendations

7.1. Conclusions

Based on this research with found results, two main conclusions are drawn. These two conclusions do answer the main research question *'How can floating car data be used to detect incidents?'*. The conclusions are closed with the short answers on the research questions.

The first conclusion is that incidents can be detected on lane level by comparing the number of lane changes for a situation without an incident and a situation with a possible incident. The new algorithm is a way to do it, this algorithm has to be improved, but the first results give enough reason to draw this first conclusion.

The second conclusion is that floating car data can be used to detect incidents. This is the case if the floating car data is accurate enough to give information about the number of lane changes for certain road stretches. Then the information from floating car data can be used as input for the new algorithm to detect incidents.

Now the short answers for each research question are presented:

How can an incident be recognised?
The main traffic characteristics of an incident, to make detection possible are stated below:

- Head of the queue is standing still;
- Downstream of the incident location is a vacuum in therms of flow and density, along with a high speed;
- Upstream of the incident location there is a queue with high density and low speed (note: not in case of a flow that is lower than the remained capacity during the incident);
- One or more lanes of the road are blocked at the incident location, results:
  - No or less vehicles just upstream, at and just downstream of the incident location;
  - More lane changes just upstream and just downstream of the incident locations.

What information can loop data and floating car data provide?
The main difference between loop data and floating car data is that floating car data contains traffic information of individual vehicles and loop data contains average about the traffic. Floating car data gives trajectories with location and speed for a vehicle, loop detector data gives average speed and average flow per minute. An important notice is that floating car data gives information about a limited amount of vehicles, the penetration rate, while loops are measuring all vehicles.

What are the requirements of the algorithms to detect an incident?
An incident detection algorithms must detect all incidents, not missing one. Thereby must the incident not give a signal when there is no incident. Finally the algorithms need to detect the incidents quick. These three requirements determine the performance of the algorithms.

Which current detection algorithms are available and what is their performance according to the literature?
Three existing algorithms are theoretically researched: the blokkadedetector, the Presikhaaf algorithm and the McMaster algorithm. In general all algorithms have problems with infrastructural changes, this causes false alarms. Besides these algorithms cannot detect incidents when the incident does not cause congestion. The McMaster algorithm is classified as the best of these three algorithm, but still not very promising.

How can a new algorithm be designed to detect incidents using floating car data?
Floating car data (FCD) can be used for incident detection by providing lane change information. The new algorithm is based on the fact that vehicles have to change lanes in case of an incident, since the road is partly blocked. So the number of lane changes per minute of a road section is compared to a situation without an incident on the same road section. Also the measured density on the road section is taken into account, since it it highly influencing the number of lane changes.

How can incident detection algorithms be evaluated?
Incident detection algorithms can be evaluated by simulating a realistic traffic pattern and calculating the false alarm rate, the detection rate and the detection time.

What is the performance of the new designed algorithm compared to existing algorithms?
The new algorithm can detect incidents with a better performance than the McMaster algorithm if all vehicles provide lane change information. In case of a penetration rate of just 10% FCD, the performance of the McMaster algorithm is comparable to the McMaster algorithm. An increase of the penetration rate will improve the performance of the new algorithm and make is more suitable for real incident detection.

7.2. Recommendations

During the research, various questions are encountered for further research. Some of these recommendations for further research are mentioned below:

- In this research the density is used together with the number of lane changes. Interesting would be to use the speed and the number of lane changes, especially since the speed can also be gathered from floating car data.
- Also further development of the algorithm is desirable to optimise the performance.
- An interesting scenario to test the performance of the new algorithm is in case of a low flow, especially if the flow is lower than the rest capacity after the incident $Q < Q_{restcapacity}$. In this case no queue will build up, so it is a perfect opportunity for the new algorithm to prove its added value compared to the existing algorithms.
- Also fascinating would be to test the new incident detection algorithm in another simulation program, preferably a simulation program with more accurate lane change behaviour.
- Eventually it will be interesting to test the new algorithm with floating car data that is really gathered by actual vehicles to see if really faster detection of incidents is achieved.
- Another new method to detect incidents is by cameras. It would be interesting to compare the new algorithm with FCD with the performance of detection by cameras.
- Another application of the algorithm can be detecting incidents in non signalised regions like urban roads. The performance in this case have to be researched further.

List of Figures

List of Tables

Bibliography

[1] Be-Mobile. Floating car data voor automatische incident detectie, May 2016.

[2] Jorn de Vries. Flitsmeister floating car data. `https://www.flitsmeister.nl/fcd.html`, 2017. Accessed: 2017-04-08.

[3] Ana I Gall and Fred L Hall. Distinguishing between incident congestion and recurrent congestion: a proposed logic. *Transportation Research Record*, 1989.

[4] BS Kerner, C Demir, RG Herrtwich, SL Klenov, H Rehborn, M Aleksic, and A Haug. Traffic state detection with floating car data in road networks. In *Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE*, pages 44–49. IEEE, 2005.

[5] V.L. Knoop. Reader traffic flow theory and simulation cie4821-09. Technical report, Delft Technical University, February 2015.

[6] Xiao-Yun Lu and Alexander Skabardonis. Freeway traffic shockwave analysis: exploring the ngsim trajectory data. In *86th Annual Meeting of the Transportation Research Board, Washington, DC*, 2007.

[7] Emily Parkany and Chi Xie. A complete review of incident detection algorithms & their deployment: what works and what doesn't. Technical report, University of Massachusetts, 2005.

[8] Emily Parkany and Chi Xie. A complete review of incident detection algorithms & their deployment: what works and what doesn't. Technical report, 2005.

[9] Bhagwant N Persaud, Fred L Hall, and Lisa M Hall. Congestion identification aspects of the mcmaster incident detection algorithm. *Transportation Research Record*, (1287), 1990.

[10] Dieter Pfoser. Floating car data. In *Encyclopedia of GIS*, pages 321–321. Springer, 2008.

[11] Verkeer en Leefomgeving Rijkswaterstaat, Water. Publiekrapportage rijkswegennet, december 2016 - april 2017. `https://www.rijksoverheid.nl/documenten/rapporten/2017/06/28/publieksrapportage-rijkswegennet`, 2016. Accessed:2017-12-10.

[12] Ralf-Peter Schäfer, Kai-Uwe Thiessenhusen, Elmar Brockfeld, and Peter Wagner. A traffic information system by means of real-time floating-car data. 2002.

[13] Delft Technical University. Fosim 5.gebruikershandleiding. `https://www.fosim.nl/documenten/fosim5.1manual.pdf`, 2006.

[14] JWC Van Lint and Serge P Hoogendoorn. A robust and efficient method for fusing heterogeneous data from traffic sensors on freeways. *Computer-Aided Civil and Infrastructure Engineering*, 25 (8):596–612, 2010.

[15] L.R.W. van de Ven. Automatische incident detectie met behulp van meetlusgegevens. Msc thesis, Delft University of Technology, May 2007.

[16] Xiao-Yuan Wang, Kai-Wang Zhang, and Xin-Yue Yang. Research of the road traffic incident characteristics. In *Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on*, volume 5, pages 2688–2693. IEEE, 2005.