BRNO UNIVERSITY OF TECHNOLOGY

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF CIVIL ENGINEERING

FAKULTA STAVEBNÍ

INSTITUTE OF ROAD STRUCTURES

ÚSTAV POZEMNÍCH KOMUNIKACÍ

TRAFFIC LANES MERGING ANALYSIS AND POSSIBLE IMPROVEMENT MEASURES

ANALÝZA SPOJOVÁNÍ JÍZDNÍCH PRUHŮ A NÁVRH MOŽNÝCH OPATŘENÍ

DIPLOMA THESIS

DIPLOMOVÁ PRÁCE

AUTHOR
AUTOR PRÁCE
Bc. Igor Mikolášek

SUPERVISOR
VEDOUCÍ PRÁCE
doc. Ing. PETR HOLCNER, Ph.D.

BRNO 2017

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
FAKULTA STAVEBNÍ

STUDIJNÍ PROGRAM	N3607 Stavební inženýrství
TYP STUDIJNÍHO PROGRAMU	Navazující magisterský studijní program s prezenční formou studia
STUDIJNÍ OBOR	3607T009 Konstrukce a dopravní stavby
PRACOVIŠTĚ	Ústav pozemních komunikací

ZADÁNÍ DIPLOMOVÉ PRÁCE

DIPLOMANT	Bc. Igor Mikolášek
NÁZEV	Analýza spojování jízdních pruhů a návrh možných opatření
VEDOUCÍ DIPLOMOVÉ PRÁCE	doc. Ing. Petr Holcner, Ph.D.
DATUM ZADÁNÍ	31. 3. 2016
DATUM ODEVZDÁNÍ	13. 1. 2017

V Brně dne 31. 3. 2016

....

doc. Dr. Ing. Michal Varaus
Vedoucí ústavu

....

prof. Ing. Rostislav Drochytka, CSc., MBA
Děkan Fakulty stavební VUT

PODKLADY A LITERATURA

Zákon č. 361/2000 Sb. o provozu na pozemních komunikacích
ČSN 73 6101 Projektování pozemních komunikací
ČSN 73 6110 Projektování místních komunikací
ČSN 73 6102 Projektování křižovatek na silničních komunikacích
Treiber, Kesting - Traffic Flow Dynamics: Data, Models and Simulation
Gang-len Chang, Zuojin Zhu - A Macroscopic Traffic Model for Highway Work Zones: Formulations and Numerical Results
Saeed, Abolfazl, Ali - A comparative numerical study on Freeway weaving area using a multi-class first-order macroscopic model
Manuály Aimsun
Mapové podklady
Ortofotomapy

ZÁSADY PRO VYPRACOVÁNÍ (ZADÁNÍ, CÍLE PRÁCE, POŽADOVANÉ VÝSTUPY)

Proveďte analýzy dopravních dějů při spojování dopravních proudů v místě zúžení. Zjistěte a popište současný stav poznání o tomto problému v teoretické i praktické oblasti výzkumu.
Analyzujte problémy tohoto dopravního děje a pokuste se posoudit, jestli lze vhodnou organizací a případnými úpravami v místě zúžení docílit zlepšení, pokuste se je kvantifikovat. Pomocí simulací se pokuste ověřit možná opatření.
Použijte k tomu data získaná během studijního pobytu v Dánsku a vlastní dopravní průzkumy. Při dopravních průzkumech se zaměřte se na základní charakteristiky dopravního proudu - hustotu, rychlost a intenzitu - v místě a v blízkém okolí spojování jízdních pruhů na komunikaci. Popište a analyzujte děje při spojování dopravních proudů a pokuste se posoudit možnosti ovlivnění kapacity v místě zúžení.

STRUKTURA BAKALÁŘSKÉ/DIPLOMOVÉ PRÁCE

VŠKP vypracujte a rozčleňte podle dále uvedené struktury:

1. Textová část VŠKP zpracovaná podle Směrnice rektora "Úprava, odevzdávání, zveřejňování a uchovávání vysokoškolských kvalifikačních prací" a Směrnice děkana "Úprava, odevzdávání, zveřejňování a uchovávání vysokoškolských kvalifikačních prací na FAST VUT" (povinná součást VŠKP).
2. Přílohy textové části VŠKP zpracované podle Směrnice rektora "Úprava, odevzdávání, zveřejňování a uchovávání vysokoškolských kvalifikačních prací" a Směrnice děkana "Úprava, odevzdávání, zveřejňování a uchovávání vysokoškolských kvalifikačních prací na FAST VUT" (nepovinná součást VŠKP v případě, že přílohy nejsou součástí textové části VŠKP, ale textovou část doplňují).

....

doc. Ing. Petr Holcner, Ph.D.

Vedoucí diplomové práce

ABSTRACT

The presented thesis deals with lane merging at lane drops. The theory of traffic flow is briefly introduced and put into the perspective of lane merging. Forming of queues at lane drops, the capacity drop and traffic flow behaviour at lane merging is explained. A review of existing measures at lane drops at work zones and elsewhere is provided. Measurements of traffic flow from three different locations are presented. The locations are introduced, the methods used for analysis of the data are explained and the results are provided and discussed. The capacity drop is confirmed and the first proof of concept of the later introduced metering system is presented. The behaviour of the merging drivers was found to have a significant influence on the merging capacity during congestion. Further, an overview of existing applications of traffic light in traffic flow control is provided and ramp metering and mainstream metering is explained. The new metering system for lane drops is presented including several possible modifications and extensions. Finally, the proposed metering system is tested in microsimulation software Aimsun. The simulations further confirm the viability of such systems. It brings significant capacity improvements and consequently even greater improvements of delays and travel times due to shorter queues.

ABSTRAKT

Předkládaná diplomová práce se zabývá spojováním jízdních pruhů. V úvodu je představena teorie dopravního proudu a je aplikována k popisu dějů při spojování dopravních proudů. Je vysvětlena tvorba kolon v místě spojování pruhů, pokles kapacity způsobený kongescí a chování dopravního proudu v místě spojení pruhů. Dále je představen přehled aktuálně běžně využívaných dopravních opatření v místech pracovních míst a jiných míst poklesem počtu jízdních pruhů. Bylo provedeno měření dopravního proudu na třech lokalitách se snížením počtu jízdních pruhů. Jsou popsány jednotlivé lokality, průběh měření a metodika analýzy dat a jsou představeny a diskutovány výsledky měření. Ta potvrzují pokles kapacity v době trvání kolony a poskytují podporu navrhovanému konceptu regulace dopravního proudu. Také se ukázalo, že chování řidičů v oblasi spojování pruhů má významný vliv na kapacitu. V neposední řadě je představeno současné využití svetelné signalizace k účelu regulace dopravního proudu, zejména ramp metering, a také zatím nepraktikovaný návrh využití u poklesu počtu jízdních pruhů. Je navržen zcela nový systém regulace dopravního porudu před zůžením komunikace s pomocí světelné signalizace a to včetně několika možných modifikací a rozšíření. Tento systeém je na závěr otestován v simulačním software Aimsun. Tyto simulace dávají příslib využitelnosti podobných systémů v praxi. Měly by přinést významné zvýšení kapacit a v důsledku toho ještě výraznější zlepšení časových ztrát a cestovního času v návaznosti na snížení délky kolon.

KEY WORDS

traffic flow, lane merging, lane drop, work zones, intensity, capacity, mainstream traffic flow control, mainstream metering, microsimulation

KLÍČOVÁ SLOVA

dopravní proud, spojování jízdních pruhů, snížení počtu jízdních pruhů, pracovní místa na dálnici, intenzita, kapacita, řízení hlavního dopravního proudu, regulace dopravního proudu, mikrosimulace

BIBLIOGRAFICKÁ CITACE VŠKP

Bc. Igor Mikolášek *Traffic lanes merging analysis and possible improvement measures*. Brno, 2016. 86 s. Diplomová práce. Vysoké učení technické v Brně, Fakulta stavební, Ústav pozemních komunikací. Vedoucí práce doc. Ing. Petr Holcner, Ph.D.

Prohlášení:

Prohlašuji, že jsem diplomovou práci zpracoval(a) samostatně a že jsem uvedl(a) všechny použité informační zdroje.

V Brně dne ………………..

…………………………………………

podpis autora

ACKNOWLEDGEMENTS

I thank doc. Ing. Petr Holcner Ph.D for supervising my Thesis as well as to Thomas Kjær Rasmussen who supervised my work at DTU and all those who have guided me during my university studies. A big 'thank you' belongs to my family who supported me through all my studies and made it possible for me to spend a year studying in Denmark.

CONTENTS

INTRODUCTION

With the ever-increasing transport demand and growing number of cars and other types of motorized vehicles on the roads, the capacity of the transport infrastructure is of paramount importance. To keep the level of service at a reasonable level the road network must be continuously maintained and improved to meet this demand. There are three basic ways of doing this. The first is reducing or at least preserving the level of demand for individual automobile transport by supporting other transport modes such as public transport (PT) or bicycles. Another way is to add lanes on the existing roads and introducing new roads. Either of these strategies has its drawbacks – additional lanes generate more traffic, and PT or bicycle is not always a suitable means of transport for individuals – and while they can be helpful and should be utilized appropriately, it is also possible to focus on the existing road network and try to find ways to improve its capacity, which is the aim of this thesis.

Generally speaking, the capacity, especially in urban areas, is given not by the capacity of sections and the number of lanes there, but mainly by the capacity of various bottlenecks, typically intersections. Intersections are naturally not the only type of bottleneck. Another type, which very much resembles a bottleneck even visually, is the reduction of number of lanes on a road called lane drop or lane merge. Lane drop is a typical bottleneck during roadworks on highways (work zones) or on some major urban roads with a low number of intersections and high traffic intensities. Improving capacities at lane drops can offer non-negligible benefits for the overall network performance. Besides that, reducing congestions significantly lowers the costs of road maintenance due to decrease of static traffic, fumes emissions and thus environmental burden, as well as opportunity costs of the road users.

In a previous study (Mikolášek, 2014) it was discovered that the traffic intensity at the lane merging was highest just before congestion occurred and the traffic flow collapsed into a stop-and-go mode. Given the instability of the system before congestion due to the variable demand, the peak performance (intensity) was sustained for short periods of time and the shorter the period, the higher the peak performance. However, even over five-minute intervals the maximum intensity in the pre-congestion state was better than in the congested one, let alone for one-minute intervals. Based on those results a hypothesis was made that the capacity of lane merging is reached before the traffic flow collapses into congestion and that it might be possible to improve the performance during the rush hours when congestion builds up via intelligent transport system stimulating the traffic flow to merge in a pre-congestion manner, i.e. with lower density, higher speed and consequently higher intensity. Review of English-written literature revealed that the phenomenon called '(congestion-induced) capacity drop' has been long known and may reach 5-20 % of the nominal downstream capacity (e.g. Gazis & Foote, 1969; Chung et al., 2007; Carlson et al., 2011).

Three options of flow optimization were suggested in the earlier study (Mikolášek, 2014). Autonomous cars with collective behaviour (platoons) are being developed by automobile companies. Variable speed limits (VSL) are commonly used and the potential of using VSL to improve lane merging is being explored (e.g. Carlson et al., 2011; Radwan et al., 2011). Finally, a rather innovative system which utilizes traffic lights to optimize the traffic flow directly before the lane merging was proposed. Its application should lead to more fluent merging of the traffic flows and consequently to increasing the traffic flow intensity. A similar system was proposed and tested in microsimulation (Lentzakis et al., 2008; Papageorgiou et al., 2008; Tympakianaki et al., 2014) with promising results. There are other similar applications of traffic lights that are already being used ranging from the quite common on-ramp metering, usually using ALINEA feedback control (Papageorgiou et al., 1991), to mainstream traffic flow control (MTFC) very similar to the one proposed, although those are much rarer and are limited to several critical bottlenecks like major bridges or tunnels (Gazis & Foote, 1969; Jacobson & Landsman, 1994; Haboian, 1995).

This study aims at analysing the merging process and suggesting suitable measures to improve the capacity of lane merges. This was done in several steps. At first, additional measurements were performed in order to confirm the congestion-induced capacity drop, to get some data from more ordinary merge to rule out location-specific behaviour due to the pedestrian crossings downstream at the Brno merge, and to compare different lane merges and analyse possible differences and their likely reasons. Two diverse lane merges in Denmark were chosen for the measurements. The measurements and their analyses were followed by a literature review to get deeper understanding of the topic and to find out what is the state-of-the-art in the field of lane merge research. As it was said before, the literature review revealed, among other, that the capacity drop is a known phenomenon and that there are few intelligent transport systems utilizing traffic lights already in use and are constantly upgraded. Finally, a series of microsimulations was performed in microsimulation software Aimsun to test a proposed measure on a two-to-one lane merge. The simulation results are compared with the measurements to evaluate its potential.

1 TRAFFIC FLOW THEORY

Traffic flow has been studied since the dawn of cars in the first half of the twentieth century. The first important theory of traffic flow was the Greenshield's theory (Greenshield et al., 1935), even though it was very simple and inaccurate. As more and more cars were filling the streets, the importance of the research field grew and new, improved, theories were introduced to describe the traffic flow behaviour.

1.1 ELEMENTARY TRAFFIC FLOW VARIABLES

There are three elementary variables describing a traffic flow – intensity, speed and density. These are used to describe the state of a traffic flow at a given time and place. Based on desired level of detail they can be averaged over longer or shorter time periods and distances (e.g. average intensity over one minute / hour / day, density over 100 m or one km, etc.). Different measures are useful for different purposes.

1.1.1 Intensity

Intensity represents the number of vehicles passing through a thought line on a road per a given period of time, most commonly expressed as the number of vehicles per hour (veh/h). It can be measured per lane, per direction or per road and it should always be clear which one is used as it may not be clear from the value. This work uses intensity per lane (or per direction when the section of road before the merge is concerned).

Interval over which the intensity has been measured is very important when interpreting the value. One has to be aware that it is rarely the same value as the one used in the intensity unit (e.g. intensity can be expressed as veh/24h even though it has been measured over a period of several hours or as veh/min while being measured over an hour). The shorter the measurement period the more fluctuations are prone to appear in the results. While one-minute interval can provide information on short peak intensity or estimated capacity (see below) it cannot be used to evaluate long term performance. Likewise, one hour interval does not provide information on potential for capacity increase because it is affected by intensity drops inevitable in real-world traffic flow. Therefore, when estimating real-world capacity and average intensity, longer intervals are more relevant and when estimating potential capacity achievable by e.g. application of ITS shorter intervals are useful.

The highest achievable intensity for a given point on a road is called capacity. Capacity is perhaps the most important variable when designing roads. Since the traffic flow is strongly variable, non-linear and dynamic phenomenon affected by many variables, capacity cannot be found analytically as a single explicit value, it is rather an estimate based on empirical data and generally would express estimated maximal potential intensity per hour under normal conditions. The capacity of similar infrastructure layout can vary depending on conditions. Using intelligent transport systems (ITS) can lead to improving capacity through affecting the traffic flow.

While intensity, or capacity, is important for evaluating transport performance it is not used to measure level of service (LoS) since it is not directly relevant for individual road users.

1.1.2 Speed

Speed, rather obviously, represents a change of position per increment of time. Within the frame of the traffic flow theory, it can be looked upon from two different perspectives.

Firstly as a speed of an individual vehicle, i.e. the distance covered by the vehicle per given period of time, and secondly as an average speed of the traffic flow, i.e. the average speed of vehicles in the traffic flow at a given time and on a given part of the road (average distance covered by the relevant vehicles per given time) also called the mean speed.

For microsimulation, the disaggregate approach is more relevant, while for macrosimulation, road performance evaluation and transport engineering in general, the aggregate version is much more relevant. That being said, while for overall performance the average speed of the traffic flow is important, individual drivers generally care particularly about their own speed or, consequently, travel time. In congested conditions the speed of individual vehicle is determined by the speed of the surrounding vehicles and, if measured over a sufficiently long period of time, is approximately equal to the speed of the traffic flow.

Within the transportation field, speed is usually expressed as km/h rather than of the basic SI unit m/s for practical reasons. Speed, represented by the travel time, is a crucial variable affecting the LoS.

1.1.3 Density

Density expresses the number of cars per certain length of a road, usually vehicles per kilometre of road (veh/km). Just like intensity, it can be expressed per lane, per direction or per road as a sum of cars in both directions and it should always be expressed what it is related to as it cannot generally be guessed by its value. In this work, density is again always expressed per lane. Density is also one of the variables used to measure the level of service as it is directly affecting speed.

Measuring density is more cumbersome than measuring intensity. Static picture has to be used to calculate the number of vehicles over a measured distance. The value is only valid for the given time and place. It is also inevitably burdened with error given the integer nature of the number of vehicles. That error is diminishing with longer stretches of road with higher car count.

1.2 TRAFFIC FLOW VARIABLES RELATIONS

There is a relationship between the three elementary variables but as traffic flow is a non-linear, dynamic, and more stochastic than deterministic phenomenon it cannot be derived or described analytically.

For a stable traffic flow, the following formula holds true

$$q = v * \rho \tag{1.1}$$

where q is intensity, v is speed and ρ is density. However, the speed is also depending on density

$$v = v(\rho) \tag{1.2}$$

Therefore

$$q(\rho) = v(\rho) * \rho \tag{1.3}$$

Density is sometimes considered an independent variable but in reality all three elementary variables are mutually dependent and the relationship is dependent on the conditions.

The traffic flow is not only affected by the internal elementary variables but also by many external variables, ranging from the driver's behaviour, intentions, driving style, or alertness, to road geometry or weather conditions, which affect the relationship between speed and density. The relationship is the integral part or even the definition of many traffic flow theories. There is a wide range of those theories and formulas trying to describe the traffic flow using many different variables based on empirical data. These theories can be divided into two elementary categories: macroscopic and microscopic.

1.2.1 Macroscopic scale – fundamental diagrams

Macroscopic theories work with aggregated data and look at the traffic flow as a whole and are analogy to hydrodynamics in the way they perceive the flow. They generally work with the three elementary variables and due to their relative simplicity they are more robust and easier to put to use in large scale scenarios. It was the macroscopic traffic flow theories that started the field of traffic flow research and the previously mentioned Greenshield's theory was the pioneer (Greenshield et al., 1935).

The fundamental diagrams serve to describe the relationship between the three elementary variables but they are not used in traffic modelling usually. For macroscopic models, travel time-intensity relationship is commonly used to distribute the traffic on the individual transport links in the widely used four stage model while on a microscopic scale, completely different approach is used (see chapter 1.2.2).

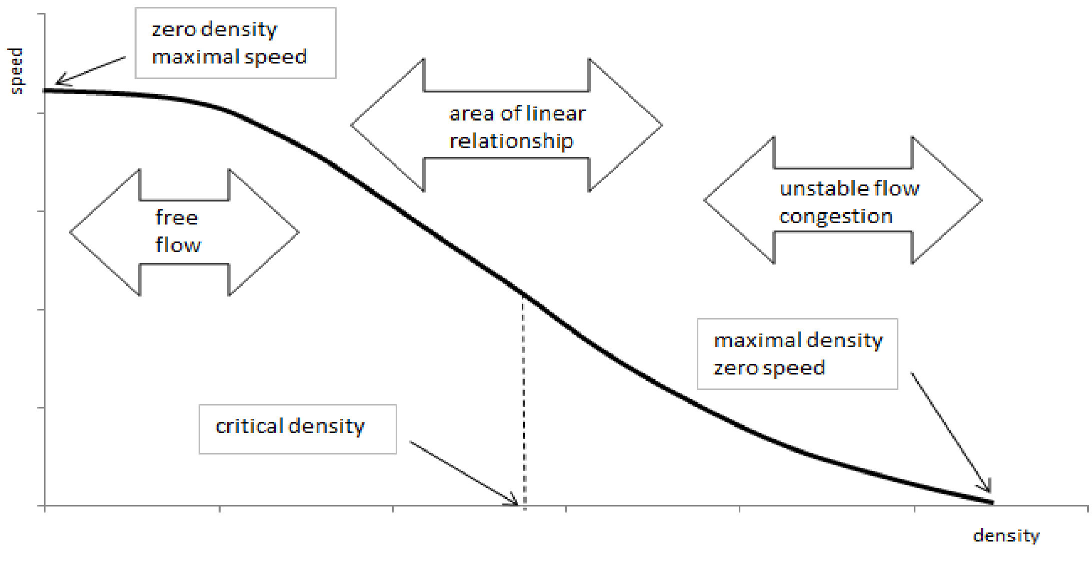


Figure 1.1 General fundamental density-speed diagram (Holcner, 2012)

The relationship among the elementary variables can be expressed in a graph known as the fundamental diagram. Since every model presents a different relationship among the variables, there is virtually an infinite number of fundamental diagrams which more or less accurately represent the real relationship among the variables. The general relation between the traffic flow density and speed is pictured in Figure 1.1. It is the primary relation used to derive the density-intensity diagrams and speed-intensity relations.

When describing the fundamental diagram several assumptions are made. The traffic flow is assumed to converge to a stationary homogeneous state, i.e. the traffic flow variables remain constant over time and space. It is assumed that if the traffic flow is measured for a sufficient amount of time, the measurements will correspond to the fundamental diagram. Any necessary deviations that will be recorded can be attributed to different vehicle- and driver characteristics and conditions, which can be included stochastically (Holcner, 2012).

For low density, the individual vehicles do not, or only rarely, interact and can choose their speed. Therefore, the average traffic flow speed is approaching the limit speed and decreases only slowly with increasing density. This state is called the 'free flow'. With increasing density the interactions between vehicles become more common and the speed starts to decrease more quickly. The middle part of the diagram is characterised with linear decrease of speed with density growth. Maximal intensity is reached at a critical density. As the density further grows, the interactions between vehicles are permanent, the speed of a vehicle is determined by the preceding vehicles and the traffic flow becomes unstable. At a certain point the traffic flow collapses into a stop-and-go kind of movement due to the inability of drivers to sustain their speed low enough for the low headway. The stop-and-go state is characteristic by waves of movement that move against the flow of the traffic.

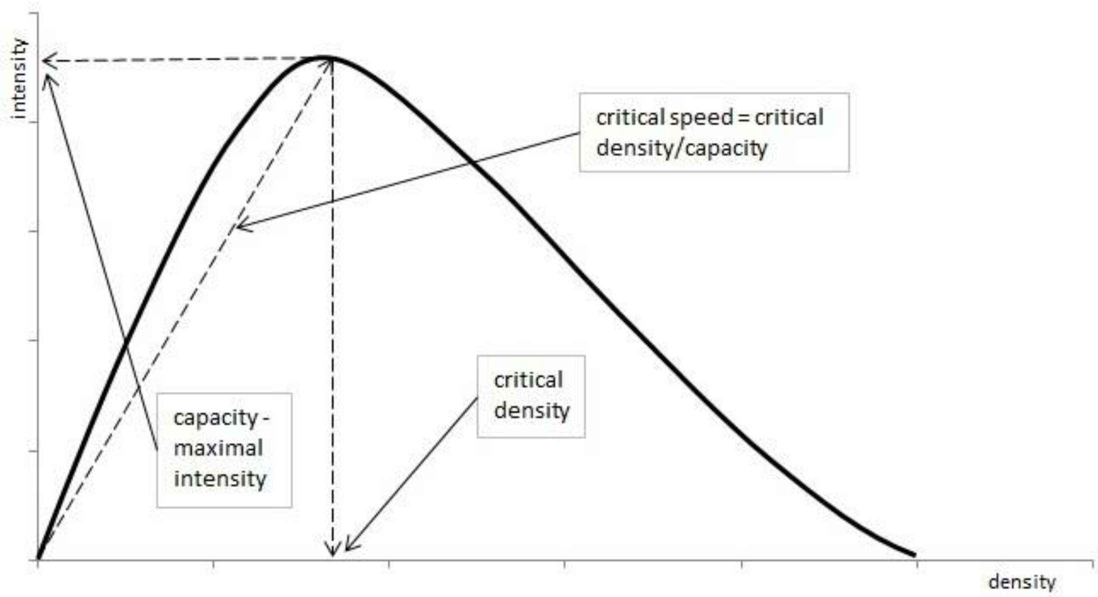


Figure 1.2 General fundamental intensity-density diagram (Holcner, 2012)

Figure 1.2 shows the density-intensity relationship. The maximal intensity, or the capacity, is reached at the critical density. For low density, there are too few vehicles and their speed is limited. As more vehicles are present, the intensity grows while the speed goes down up to the point where the speed drop effect exceeds the effect of a growing number of vehicles where the intensity starts to decrease again.

Real-world data shows that the fundamental diagrams represent reality relatively well but fail at high densities where the traffic flow is congested, becomes unstable and eventually collapses into the stop-and-go behaviour. In reality the density never gets so high or when it does, the traffic flow collapses into the stop-and-go state and is either not moving at all or, when it is moving, it is with much lower density again. Fundamental diagrams, assuming stationary and homogeneous traffic flow, cannot be effectively used to describe this unstable alternating behaviour.

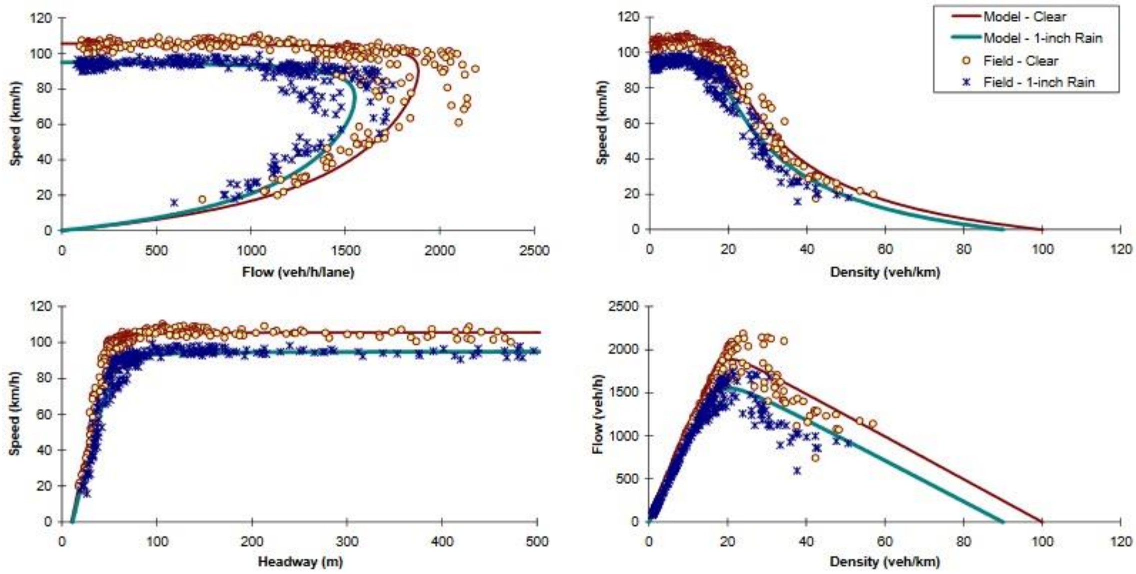


Figure 1.3 Real-world-data based fundamental diagrams (Hranac et al., 2006)

The conservation law can be applied to the traffic flow, i.e. the number of vehicles passing through one point is equal to the number of vehicles passing a point downstream plus the number of vehicles between those two points. It can be used to estimate number of vehicles on a stretch of road and, knowing its length, the density. It is utilized on intersections with intelligent traffic light systems for estimation of vehicle count on a section. See chapter 4.1.2 for another example of potential application in ITS.

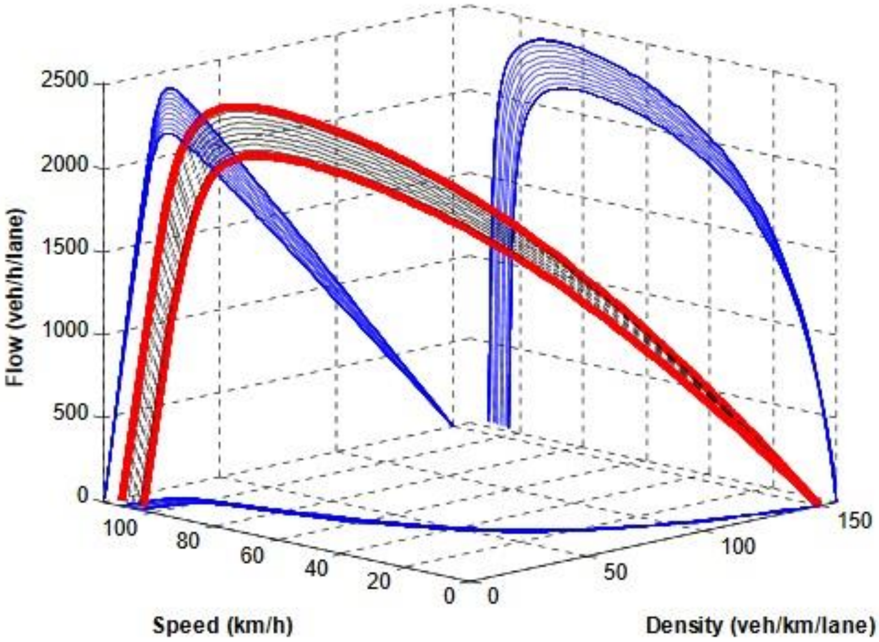


Figure 1.4 3D fundamental diagram based on model exploring impact of weather on traffic flow characteristics (Hranac et al., 2006).

Greenshield's model

The Greenshield's model (Greenshield et al., 1935) is very simple and intuitive. It puts linear or quadratic relations among the three elementary traffic flow variables. While it does not reflect empirical data very well, especially in the free flow and congested traffic flow states, its simplicity and robustness made it widely used for long time despite its flaws and it still serves well for presenting the elementary principles of the traffic flow. Its main importance nowadays lies in being the first traffic flow model on which basis the whole field grew.

The Greenshield's model is characterized by a linear relationship between traffic flow density and speed and parabolic relation of intensity on speed and density as shown in Figure 1.5. The monotonously decreasing function of *v(ρ)* is in line with the general fundamental diagram in Figure 1.1. However, the linear relation at low and high density is not realistic and explains the lack of fit to real-world data. It assumes the stationary and homogeneous traffic flow with constant density and speed of individual vehicles which is never completely true in the real-world traffic flow due to various disturbances.

$$v(\rho) = v_{max} * \left(1 - \frac{\rho}{\rho_{max}}\right) \tag{1.4}$$

Assigning formula (1.4) into (1.3) shows the parabolic relation of intensity on density

$$q(\rho) = v(\rho) * \rho = v_{max} * \left(\rho - \frac{\rho^2}{\rho_{max}}\right) \tag{1.5}$$

The symmetric parabolic shape of the intensity-density relationships is also unrealistic. In reality the capacity is reached at lower density and higher speed. Also the downwards branch of the graph is unrealistic because after the peak the intensity drop is not so dramatic and for high density the traffic flow goes over to the stop-and-go behaviour and never reaches maximal density and zero intensity. See Figure 1.4 for comparison with more realistic diagrams and Figure 1.3 for real-world data.

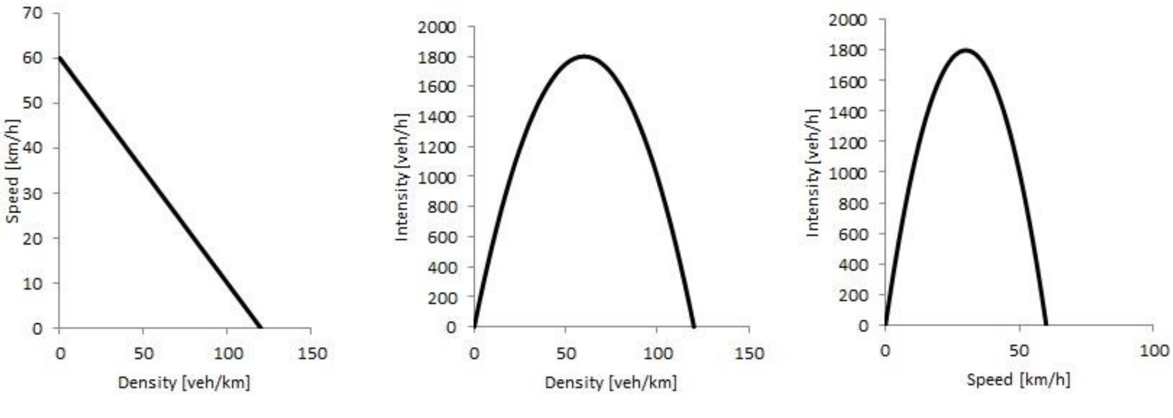


Figure 1.5 Unrealistic Greenshield's model based fundamental diagrams

Greenshield's model is fully defined by two parameters – v_{max} and ρ_{max}, i.e. maximal speed and maximal density. It should be noted that the maximal speed is not necessarily equal to maximum allowed speed which causes further problems.

1.2.2 Microscopic scale – car following models

Microscopic traffic flow theories, or rather models, on the other hand consider each individual vehicle and describe its behaviour based on the state of the traffic flow around it and its ability to perceive and react to it described by many various variables. Microscopic models can be further divided by the approach to the description of vehicle behaviour – Gazis-Herman-Rotery (GHR) models, safety distance models and psycho-physical models (Olstam & Tapami, 2004). GHR models state that the following vehicle's acceleration is proportional to the speed of the follower, the speed difference between the follower and the leader, and the space headway. Safety distance models are based on the assumption that the follower always keeps a safe distance from the vehicle in front. Psycho-physical car-following models use thresholds for determining the driver's behaviour, e.g. minimum speed difference between follower and leader that can be perceived by the follower, and include zones of different behaviour based on relative speed and distance.

Different microscopic models are used in commercial software. For example GHR model is used in MITSIM, AIMSUN uses a safety distance type of model, and VISSIM or Paramics utilize a psycho-physical model. However, the software packages comprise of several sub-models which handle different situations. While the discussed car-following models are used for following cars in a lane, as the name suggests, there are other sub-models for lane-changing, merging, overtaking, and other situations depending on the software. These are modifications or extension of the basic model and they adapt the vehicle behaviour to the given situation in order to reflect the behaviour of real drivers. The model used in AIMSUN is described more thoroughly in chapter 5.1.4.

2 LANE MERGING

Lane merging is one of the bottlenecks on a road and perhaps the most important, excluding the inevitable such as intersections. Bottleneck is a cause of decreased performance of the road network. Besides the infrastructural bottlenecks such as the intersection or lane merging, poor weather conditions, slow vehicles or distractions (e.g. accident in the opposite direction) are sometimes considered as bottlenecks, too. Lane merging is most commonly imposed by temporary work-zone on a road but can also often be semi-permanent as a result of phase construction, decrease of demand as the road gets farther from a major city, or by space restrictions from landscape or surrounding infrastructure limiting the width of a road.

There is a considerable amount of research on lane merging at highway work-zones, however there is much less interest in lane merging at other locations, especially in urban areas where space and phase construction can often be a source of bottlenecks such as lane merging and where there may be specific conditions (lower speed limits, nearby intersections or pedestrian crossings). There are several approaches to lane merging from a legal perspective (see chapter 2.1) and various systems for improving the performance of lane merges (see chapter 2.3).

2.1 LANE MERGING WITHIN THE SCOPE OF THIS WORK

There are several approaches to lane merging from the point of road rules. Either the vehicles in the closing lane are obliged to give way to the vehicles in the mainstream traffic flow (typically on-ramp) or there is no clear right of way (lane drop, typically highway work-zone) where application of the zipper merge is expected in case of high intensity. The circumstances upon which the application of the zipper merge is expected vary from country to country as well as the rules about the right of way (example below). This work is concerning the cases where application of the zipper merge is expected as on-ramp metering is efficient and proven solution for on-ramps. However, on condition that there is regularly a queue both on the on-ramp and on the highway, a modification of the proposed mainstream traffic flow control (MTFC) systems (see chapter 4) could be used there.

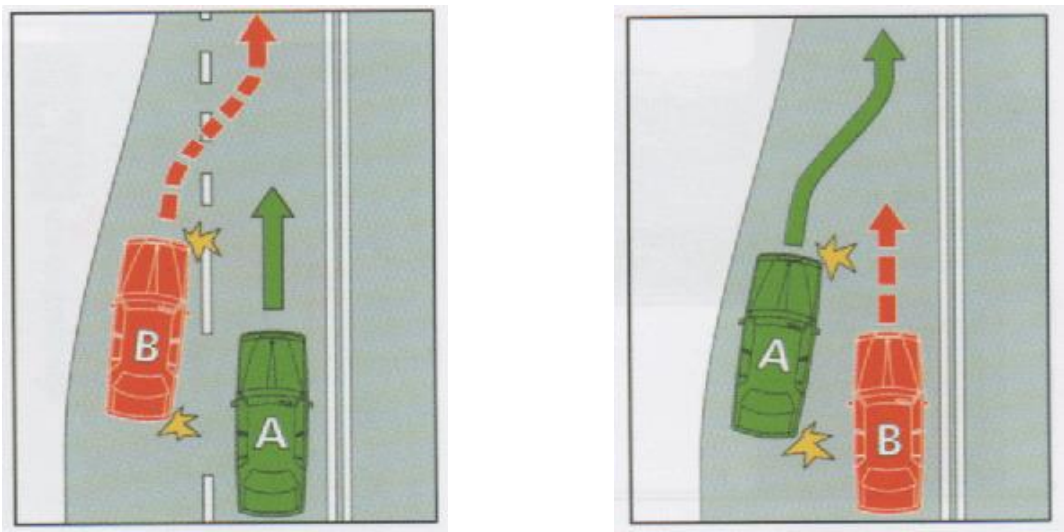


Figure 2.1 Excerpt of the Australian road rules. The vehicle in front has the right of way if there are no horizontal road signs. Otherwise it has to give way to the vehicle in the other lane (Australia, 1999).

There are also at least two different approaches to application of the zipper merge across the world. In some countries the vehicles are supposed to merge one vehicle from each lane at a time, for example in the Czech Republic (Czech Republic, 2000). On the other hand, in some countries (e.g. Australia) the vehicle more forward has the right of way (Australia, 1999). Both approaches have their own pros and cons. While the Australian style may be slightly more fluent, especially when there is higher share of long vehicles, it is more prone to crash or near crash situations as defining the right of way is not always clear. Nevertheless, it is possible that most drivers are not aware of such detailed differences and apply the zipper rule as they see fit, regardless of the exact formulation of the law.

The term truck is used for any vehicles that are long and/or slow (have low acceleration) within this work and the terms such as truck and long/slow vehicle are interchangeable. That is due to the fact that truck is a typical example of such vehicle but the low acceleration and vehicle length are what makes a vehicle problematic when participating in lane merging. Therefore, truck is used as a simple and relatable term to encompass any vehicles with such characteristics.

2.2 TRAFFIC FLOW AT LANE MERGING

Chapter 1.2.1 explains traffic flow behaviour using fundamental diagrams. As it was said, certain idealistic conditions like stationary and homogeneous traffic flow are assumed when describing traffic flow behaviour. While that may be approximately true from long term point of view on a regular road, it fails at lane merging for several reasons, at least during periods with high intensity which are relevant for this work.

The main reason, encompassing most of the others, is the sudden and rather dramatic change in traffic flow behaviour that occurs at the lane merging area, at the congestion head. The traffic prior to the merge is in congested stop-and-go state which on its own is not representable by the fundamental diagrams. At the merging area itself the traffic flow is very chaotic and irregular and there is usually much lower density and higher speed downstream.

That does not mean that the basic principles visualised by the fundamental diagrams do not apply, they rather get more complicated and other variables get into the mix. Acceleration is considered to have major influence (see next paragraph) and drivers' cooperation also plays significant role. The share of trucks grows in importance via the acceleration factor, besides their length which disturbs the regularity of the merging of the vehicles. The behaviour of traffic flow at lane merging is a very complicated and complex phenomenon and it has not yet been fully understood and described. In fact, it may never be described in its complexity due to the amount of variables and the degree of randomness affecting the behaviour.

There is a peculiar phenomenon called '(congestion-induced) capacity drop' which is occurring at lane merge when congestion forms. In fact, it is not necessarily limited to lane merging. Its principle lies, as the name suggests, in capacity drop after the traffic flow congests and a queue forms. The cause of this is considered the need of vehicles to accelerate from low or zero speed as it was said above (e.g. Gazis & Foote, 1969; Tympakianaki et al., 2014). Before a queue forms, the vehicles can drive through a lane merge at certain intensity. However, once the traffic flow collapses and a queue forms the output intensity drops due to the capacity drop (see Figure 2.2 and Figure 2.3).

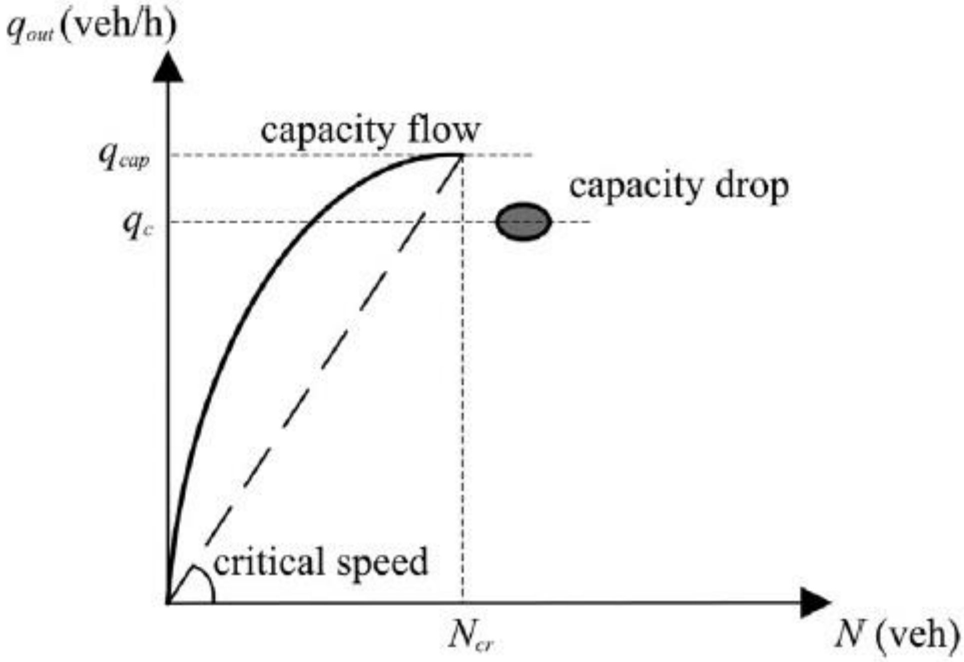


Figure 2.2 Fundamental diagram of merging area illustration the capacity drop. N is number of vehicles in the merging area, i.e. density (Papageorgiou, 2008)

The basic knowledge of the traffic flow behaviour can be used to understand the mechanism of queue forming. The traffic flow density is the key variable determining the traffic flow behaviour. The traffic flow behaviour is in line with the fundamental diagrams before congestion starts. However, the density approximately doubles at the merge as the vehicles from the two lanes are forced into one (or similarly for layouts with more lanes). Therefore, there is a sudden change in the traffic flow state and move in the fundamental diagram. As the intensity increases and there are more interactions and forced merges the fundamental diagrams cease to properly reflect the traffic flow behaviour and are not useful once there is a queue.

When there is low traffic demand and therefore the traffic density is low, there is no real merging going on actually as the individual vehicles do not interfere. As the intensity increases and more vehicles are arriving to the merge the density also increases and the headways among the arriving vehicles become shorter. When the drivers in the closing line merge into the continuous lane the headway gets unacceptably short and they slow down. As the density further grows, this happens more often and the slowing vehicles also start limiting the vehicles arriving from behind. As the headways decrease, the acceptable speed also goes down and these events occur more often.

There is a range of triggers and densities at which the queue forms but the mechanism is always the same in principle. At lower density an outside trigger, e.g. a pedestrian crossing the road or an accident, is needed. It forces the vehicles to stop but as there are constantly new vehicles arriving, they start forming the queue. If the upstream intensity is lower than the merge capacity during congestion (after the capacity drop) the queue can disperse again. However, in the case that the incoming intensity was higher than the capacity during congestion the queue keeps growing, cannot disperse, and the intensity behind the merge keeps at the congestion capacity level. It remains there until the upstream intensity goes below the congestion capacity. Only then the queue starts to get shorter and eventually disperses again. The length of the queue also plays a role as a short queue may disperse during a temporary decrease of incoming intensity.

When the flow does not break down due to some outside trigger, the intensity and density can further grow. At some point the density grows too high, exceeds the critical density, and the output intensity starts to go down (capacity drop). This happens at the point where the speed of the vehicles drops too low and the faster vehicles coming from behind start queueing, further increasing the density at the merging area, similarly to when a pedestrian stops the flow. Just like in that case the queue can disperse again in case there is temporarily lower intensity of the incoming vehicles. However, if the intensity of arriving vehicles keeps growing, sooner or later the number of vehicles and the density gets too high and the vehicles are forced to stop completely. The lower the outgoing intensity and the higher the incoming intensity and the longer the queue is, the higher is the chance that the queue will stay. Once the queue gets long enough, even temporary drop in incoming intensity will not be enough for the queue to disperse and it will stick until the average incoming intensity decreases permanently below the congestion capacity.

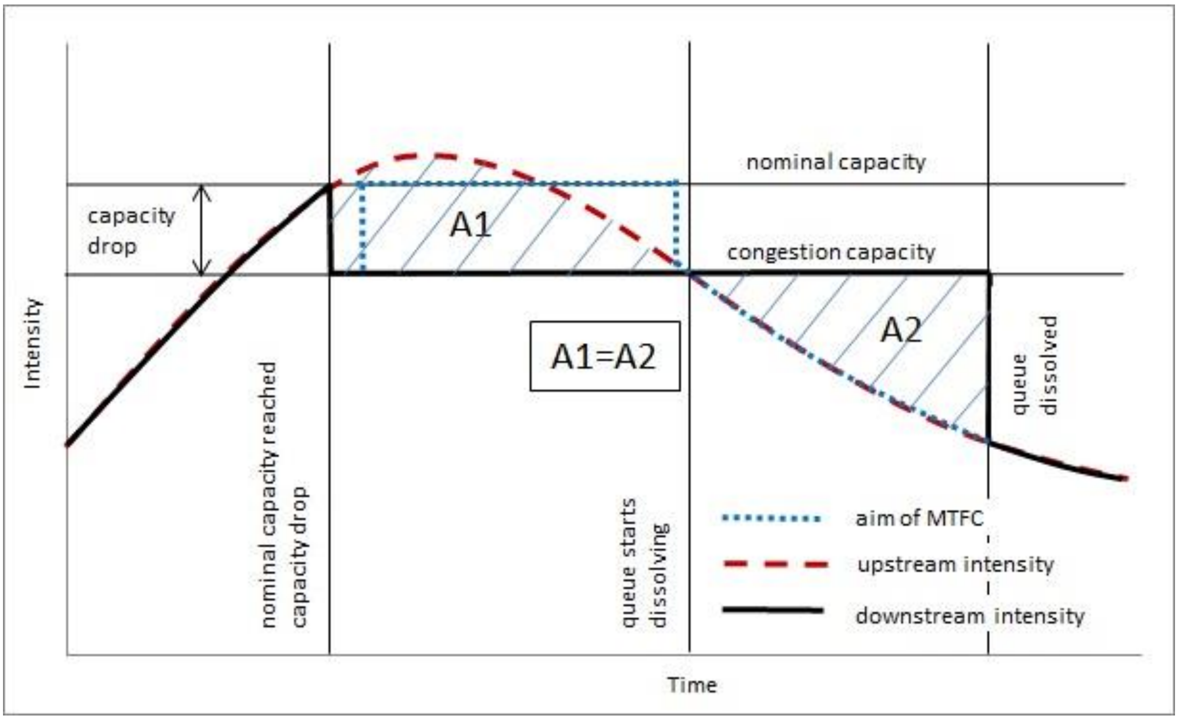


Figure 2.3 Progression of intensity in time up- and downstream of lane merging. Capacity drop and consequent loss of intensity is illustrated. Areas A1 and A2 represent the queue forming/dissolving. MTFC such as the one proposed aims at maximizing capacity and reducing queues.

As explained, the mechanism of queue forming is independent on the cause of the traffic flow breakdown, be it external (e.g. passing pedestrian) or internal (exceeding critical density). However, a strong, long-lasting, external trigger can cause a queue while without it the queue may have never formed or it would form later due to reaching if the critical density. Therefore, it is advisable to reduce any possible sources of outside triggers, such as pedestrian crossings nearby merging areas in urban areas (e.g. merging in Brno, see chapter 3.1.1).

2.3 EXISTING MEASURES USED AT ZIPPER STYLE LANE MERGING

Several measures are used across the world to improve lane merging performance and/or safety. While in the Czech Republic the late merge with zipper rule is used, other systems are used in other countries. (Pei & Dai, 2007) provide comparison of static and dynamic early/late merge and (Hallmark & Oneyear, 2015) provide similar comparison and add VSL and temporary rumble strips. Brief overview of the currently utilized systems is provided below. Some of the systems can be used in parallel, for example VSL can be used together with dynamic merging.

2.3.1 Late merge

Late merge is based on the zipper rule, i.e. vehicles are encouraged to drive in their lane to the lane closure and only there merge into the other lane. The merging should be done by alternating vehicles from the two adjacent lanes (the so called zipper rule). Application of the zipper rule is required by the law in the Czech Republic (Czech Republic, 2000) when a queue is formed. In general, application of the late merge is encouraged by road signs.

The late merge increases capacity and reduces congestion, but there are some drawbacks. Its application is problematic for high-speed low-volume situations, it is difficult to teach drivers to take turns when merging, and the unclear right of way may cause potential collisions (also different countries have different rules for the zipper merge, e.g. in Australia the vehicle 'in the front' has the right of way (Australia, 1999)).

2.3.2 Early merge

Opposing the late merge, early merge encourage drivers to merge early, prior the lane closure by road signs and prohibiting overtaking. The advantage is reduction of rear-end accidents, forced merges, vehicle stops, and aggressive driver behaviour. However, it is only working properly for uncongested flow as it causes long queues once the incoming flow exceeds the capacity. In that case, unruly drivers can overtake the queues and conversely encourage aggressive behaviour.

Figure 2.4 Illustrative picture of early merge

2.3.3 Dynamic early/late merge

Dynamic merge systems utilize detectors to measure current traffic flow and adapt the road signs accordingly. There are different systems depending on the used detectors and signs.

Dynamic early merge can turn the signs encouraging early merging on or off according to the current need and can also employ several independent signs in various distances from the merge. Analogically, dynamic late merge allows turning the signs encouraging late merging on or off. Finally, the most advanced systems can change between early and late merge based on the date from the detectors.

Figure 2.5 Typical dynamic lane merge equipment

The obvious advantage is the adaptability of such systems to current condition of the traffic flow. On the other hand, the constant changes of encouraged merging system may be confusing for drivers. Teaching drivers to adopt certain type of merge may cause them to use it at other merges where another merge system is used (Hallmark & Oneyear, 2015).

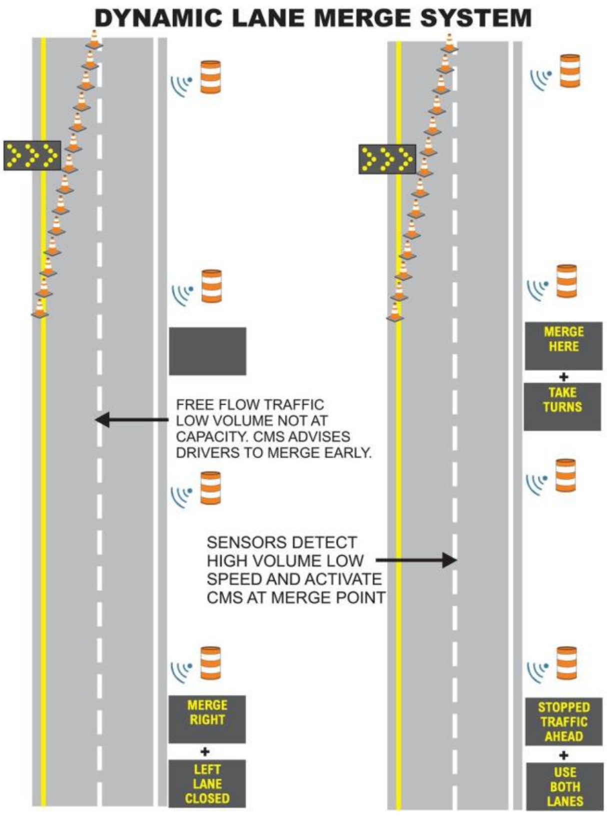


Figure 2.6 Schema of a dynamic lane merge system that allows changing between late and early merge

2.3.4 Variable speed limits

The VSL systems have not been designed specifically for lane merging. Originally, the purpose was to improve traffic flow and road safety by homogenizing speed of the vehicles. The lesser speed variance helps to reduce lane changing manoeuvres and improves safety.

While VSL were proven successful in reducing accidents, their original usage had hardly any positive impact on capacity or travel times. New dynamic systems utilizing VSL as MTFC have been proposed and even tested recently proving the viability of such systems (e.g. Carlson et al., 2011). Also combination of VSL with other traffic control systems is possible, for example combination of VSL with dynamic late merge (DLM) was tested in VISSIM with positive results (Radwan et al., 2011).

The VSL systems can be divided into several categories as Figure 2.7 shows. The scheduled systems can be used only to increase safety during peak periods. Dynamic speed limits allow adapting the speed limit according to current conditions and thus is an example of ITS. The observed variable may be weather conditions, emissions level, or, as in the case of MTFC or specifically lane merging applications, traffic conditions.

Advantage of using dynamic VSL at lane merging is the possibility of adopting speed limits adequate to the current traffic (or weather/emissions) conditions but it does not directly address negative drivers' lane merging behaviour and it may not be respected by drivers as they perceive it as a dynamic temporary limitation without enforcement and usually do not see its benefits. The drivers' respect can be improved by educational campaigns and the combination with DLM can bring the advantages of the systems together.

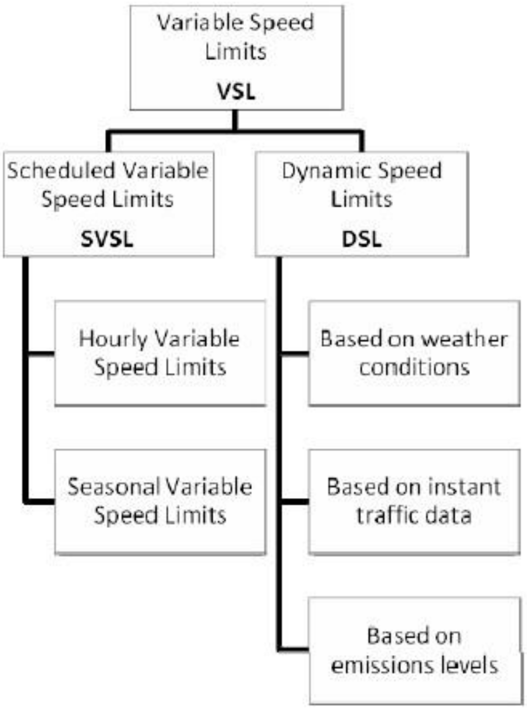


Figure 2.7 VSL classification (Garcia-Castro & Monzon, 2013)

2.3.5 Temporary rumble strips

The purpose of the temporary rumble strips is to alert drivers and encourage them to merge into the continuous lane prior to its end and thus reduce the number of late merges. As such, it is basically a support or alternative to early merge. Because they do not have a single standard use, they need to be used with additional road signs.

Figure 2.8 Example of temporary rumble strips

Due to their nature, they are not very suitable for lane merges with high intensities where congestion forms regularly as early merging is not advisable and efficient in such cases. For other applications they have several advantages as well as disadvantages. They directly address drivers’ negative merging behaviour and encourage early merge where desired at low cost. On the other hand they are noisy and may be dangerous for motorcyclists and cyclists and could also lead to dangerous manoeuvres of drivers attempting to avoid them.

3 MEASUREMENTS

To analyse traffic flow behaviour and capacity at lane merging several measurements at different locations were performed. One had been done previously in Brno, Czech Republic, at Žabovřeská street, part of the urban city ring, with rather specific infrastructure layout including side lane merging right before the actual lane merging and two pedestrian crossings downstream. Some new measurements were also done at the location. Two new locations were included in Denmark – one on the highway 16 by Allerød in the direction to Hillerød where the highway ends and two lanes merge into one, and the other one on Copenhagen Ringvej 4 after the intersection with road 211 in the direction to Kongens Lyngby. This Location is very specific by the signalized intersection short upstream.

3.1 MEASUREMENT LOCATIONS

3.1.1 Žabovřeská, Brno, CZE

The merge at Žabovřeská street is part of the future city ring, at the point where the finished part ends. Just before the merge there is an interchange with one of the ramps entering the city ring just before the merge, effectively making it 3 to 1 lane merge. This has the effect of decreasing the average traffic flow speed in the right, non-continuous, lane and consequently making the queue during congestion slightly longer in the left lane, as the drivers exploit its greater mean queue speed (ca 6 km/h vs. 8 km/h (Mikolášek, 2014)). The allowed speed at the merge is 50 km/h, although it is 80 km/h just few hundred meters before on the finished part of the city ring. Long queues are the rule here every work day during the peak hours.

Figure 3.1 Detail of the lane merging in Brno. The end of the on-ramp can be seen in bottom right.

In the distance of about 100 and 350 m downstream from the merge, there are two pedestrian crossings. The pedestrian intensity is not very high but it significantly affects the performance of the road during congestion as every passenger crossing the road stops the traffic for a period of time. Due to the road geometry it is not possible to see both the lane merging and the crossings, hence it is very problematic to estimate the magnitude of the effect. Also the pedestrian intensity is very variable over daytime and over the year due to the location. This effect has to be kept in mind when comparing the measurements to the other locations. For some of the measurements, namely those focused on the traffic flow breakdown, the measured cross-section was moved slightly downstream, after the bend, so that the pedestrian crossing would be visible and its effect could be considered.

Figure 3.2 Overhead picture of the Žabovřeská lane merge during congestion

Lane merging area geometry is also considered to be one of the possible causes of capacity differences, however due to the other location specific differences, like the pedestrian crossings or slightly different share of heavy vehicles it is not possible to determine its effect on the capacity. The merging area is only about 30 m long and the lane width is about 3 m before and about 4 m after the merge. Some additional buffer space is given by the white-striped areas which vehicles should not enter but may serve to avoid a crash.

3.1.2 Highway 16, Allerød, DK

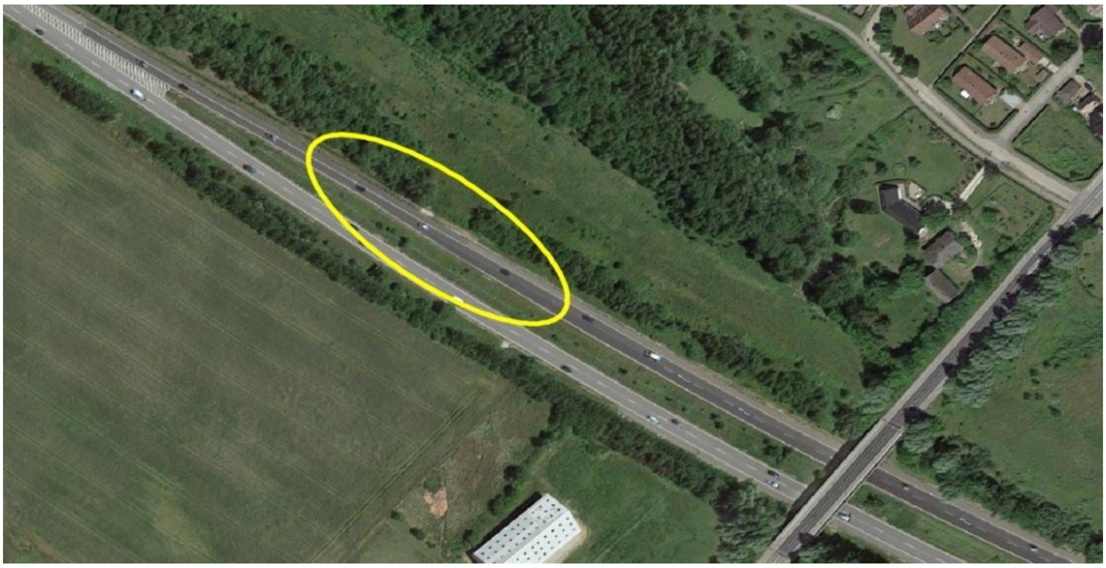


Figure 3.3 Areal shot of the Allerød merging area

The lane merging on highway 16 by Allerød is pretty much the typical highway merge you could see at a work zone. The lane merging is at the end of the motorway reducing the lane count from two to one. The speed limit is 90 km/h from about half a kilometre before the merge. Unlike in Brno, the merge is not affected by any other traffic flows. The merging area is much longer, too, with about 150 m, which might also affect the performance. The lane width is 3.25 m and a wide right hard shoulder stretches along the highway and the merge and extends several hundred meters beyond it. There is a long queue every afternoon on a workday.

Figure 3.4 Allerød merging area from the viewpoint of a driver

3.1.3 Ringvej 4, Ballerup, DK

This lane merge is very specific. The city ring of Copenhagen gets narrower with merge from two to one lane. This merge is greatly influenced by nearby upstream signal-controlled intersection. It causes the vehicles to arrive to the merge in "waves" which is quite important circumstance that greatly influences the merging and the merge performance. The merge can also be occasionally influenced by the following signal-controlled intersection in case of long queue reaching all the way to the merge.

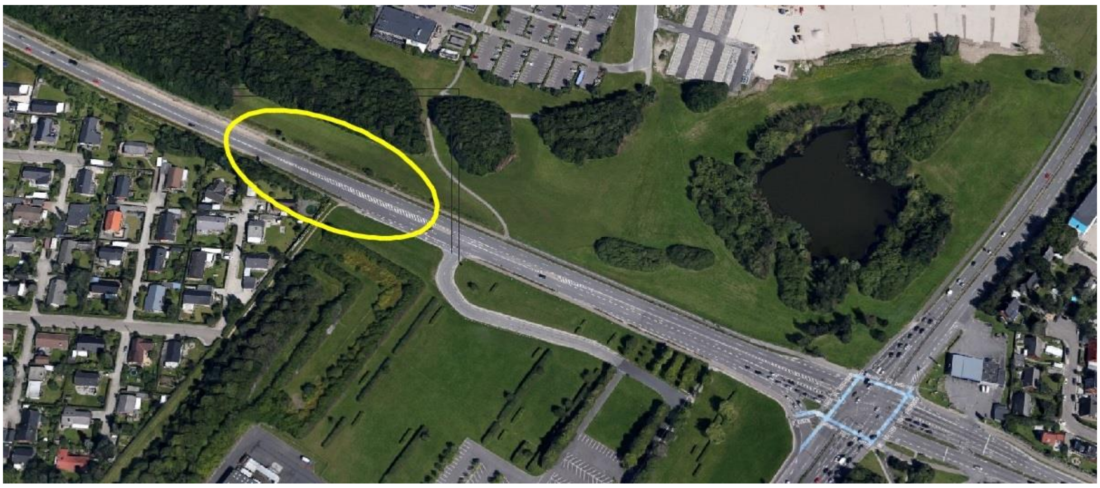


Figure 3.5 Areal picture of the Ballerup merge with the large upstream signal-controlled intersection

The location was chosen specifically for its uniqueness. The controlled intersection was thought to provide an indication on how the signal control affects the lane merging performance. The exact length on the length of the control cycle is unknown but based on the vehicles arrival to the lane merging it is estimated to 75 s.

Figure 3.6 The Ballerup merging area from drivers' point of view

One 'wave' comprised of large number of vehicles from the straight direction. Shorter gap would usually follow, depending on the number of vehicles, before fewer vehicles arriving from the right turn flow. After a short gap, the vehicles from the left turn arrived. A longer gap usually followed before the vehicles from the next wave started arriving.

3.2 MEASUREMENT DESCRIPTION

The primary objective of the data collection was to generate a database of real intensities occurring at the three chosen locations to evaluate the performance, analyse the traffic flow behaviour and to estimate the capacity of the lane merging. Based on the data, possibility of improvements was to be evaluated and possible measures were to be suggested and tested in microsimulation software. Besides the numerical data, observation was very important source of information with regard to the analysis of the lane merging in terms of traffic flow behaviour.

The data was collected using manual vehicle counting. Radars were used in Brno but they proved useless for vehicle counting at lane merging as they could not detect slow-moving vehicles. The intensity was recorded in one-minute intervals which were then aggregated to two-, five-, and ten-minute intervals to evaluate long-term performance.

Part of the observations was also recorded on a camera for later revision. The video records were used to evaluate the precision (see Table 3.1) of on-site counting as well as to estimate the share of heavy trucks or other slow and long vehicles which might affect the fluency of the merging. The total difference between on-site and video-based vehicle count over 212 one-minute measurements was a single vehicle. While it is partially due to negation of the positive and negative errors, along with personal experience, it strongly indicates that most of the errors are actually due to the inaccurate determination of the threshold for the minute interval. A vehicle may have been counted into one interval on-side and into the following interval from the video since it is not possible to differentiate the two consecutive intervals down to fractions of a second.

Table 3.1 Table of differences between on-site and video-based vehicle counting

Error	No error	±1 veh	±2 veh	±3 veh
Count	127	78	6	1
Share	59,91%	36,79%	36,79%	0,47%

Due to the initial idea of using a software to gain the intensities as well as speeds from video recordings (Kristensen, 2016), part of the measurements in Allerød were performed from the bridge about 200-300 m before the actual merge. This may have an effect on the short term intensity as the traffic flow sometimes falls into stop-and-go mode and the intensity in short term is more fluctuated in both directions. Most of the measurements were however performed either at the beginning or at the end of the merging area. The change of measurement profile was forced by tree growth in the spring which blocked the view on the original profile. These two profiles should, however, not be significantly differing in terms of intensity fluctuation, unlike the bridge profile. For the aggregated 5 and 10 min intervals, even the measurements from the bridge should not cause any issues. Eventually the idea was abandoned due to high demands on the recorded video which could not be fulfilled.

In Ballerup, all the measurements at the merge were performed right at the merge at a point where in vast majority of cases the cars would already be in a single file and the lane is too narrow for two cars riding parallel. Most of the measurements here were recorded, too. Short measurement was also performed in the single line section in the opposite direction to evaluate the capacity of a single lane. However, this was significantly affected by the signal-controlled intersections before and after the measurement profile so the actual capacity of single-lane road could not be reached.

In Brno, majority of the measurements were done earlier for the Bachelor Thesis and other variables besides intensity were also measured, including speed, density and queue length (Mikolášek, 2014). Because of this, some of the measurements were not performed directly at the merge point. However, in most cases the different measurements could not be performed simultaneously, so only minority of the recorded intensities are affected by the intensity fluctuation effect explained above. Initially, the intensity here was measured in ten-minute intervals, which did not allow evaluation of short term capacity and it complicates comparison to measurements from Denmark. The second half of the measurements was done in the one-minute interval.

3.3 DATA PROCESSING AND ANALYSIS

Before introducing the analysis and results, it is important to understand some of the terminology used in this study. As the chapter 1.1 explains, capacity is the highest achievable intensity on a given cross-section of a road. However, there can be multiple approaches to that, based on for example what time period is considered when estimating the capacity. In this work, three different kinds of capacity are considered.

First, the average intensity during congestion is considered as the 'road capacity'. While a higher intensity can be reached short-term before the congestion forms and also during the congestion, from the long-term perspective the average congestion intensity can be perceived as the road's capacity. Second, the 'congestion capacity' is estimated for each of the used intervals (one, two, five, and ten minutes, see below). It is intensity achievable regularly over a given period of time. What is considered regularly may be subject of debate. Certain percentile or the number of occurrences within a sample may be both used for the estimation. The third considered capacity is the 'theoretical capacity'. That is again an intensity that has been repeatedly measured only in this case also pre-congestion data is considered. Given the fact that prior to the congestion the intensity is determined by the demand at the given time, rather than by the lane merging process, the percentile approach explained above is not appropriate for estimating the capacity as the percentile is determined by the share of pre-congestion measurements in the sample. Instead, the number of occurrences of a given intensity is more apt for the task. As in the case of the congestion capacity, there is no firm threshold and it is up to the analyst to decide what is feasible in the given case. Estimation of the theoretical long-term capacity is based on the theoretical capacity of the short intervals, from which the long-term capacity is calculated as a multiple.

Taking the road capacity as the base value, it can be compared to the congestion capacity or theoretical capacity. The difference can be considered the estimate of possible improvement. The congestion capacity is based purely on the data gathered during congestion so to achieve stable long-term intensity as high or near as high as the congestion capacity it should be enough to ensure homogeneous traffic flow conditions during the congestion. Sustaining the intensity at the level of the theoretical capacity, long-term, would likely need more severe regulation of the traffic flow, such as significantly decreasing the density of the traffic flow in the merging area. Needless to say, ensuring homogeneous traffic flow without severe regulation is unlikely, so the two cases could also be considered equal, in reality. It is possible that by optimizing the traffic flow properties, intensities exceeding even the theoretical capacity could be achieved. Comparing the congestion capacity and the theoretical capacity may also be suitable approach as both values have similar origin, both being based on few rare occurrences within the sample, while the road capacity has slightly different nature and more accurately reflects the reality.

The measured data were analysed separately after each measurement for preliminary evaluation. At the end, the data were aggregated for each measurement site and the data from during congestion were copied and analysed separately. The analysis used previously (Mikolášek, 2014) was further developed and is now based on histograms which are easier and more straightforward to interpret and compare. Also other plots and tables were created from the data to provide more information. The congested intensities were analysed separately along the complete measurements.

Since the intensities were measured in one-minute intervals, which are not very relevant for long-term behaviour analysis and performance evaluation, the intensities per minute interval were summed into five- and ten-minute intervals and newly also into two-minute intervals which allow more realistic estimation of the theoretical capacity for the case of homogenous traffic flow. Two minutes is relatively long interval and it can be assumed that if a certain intensity was reached several times over two-minute interval, it could be also sustainable for longer periods, whereas one-minute interval is relatively short and the peak values may not be sustainable long-term due to other restrictions such as the capacity of the single lane. It should be kept in mind that the summing was done starting every minute (i.e. summing intervals 1-5, 2-6, 3-7 etc. in the case of the five-minute interval), therefore every real-world minute is considered twice, four times and nine times, in case of two-, five-, and ten-minute interval, respectively. That way, the sample size for longer intervals can be significantly increased and allows recording every thought interval within the measurement period. Even though it also brings the risk of misinterpretation of the data due to recording one exceptional high-intensity event in few consecutive intervals (e.g. there is a seven minutes long interval of high intensity, which will be recorded as three 5-

min intervals with high intensity which will be highly correlated because they are basically part of the same event), the risk may be reduced by careful interpretation of the results (e.g. looking at higher percentile or demanding higher number of occurrences when estimating the capacity). The pros of this approach are considered to highly surpass the cons. The fact that the highest intensities are reached with low or even zero share of trucks which is not sustainable over long period of time also should be kept in mind when estimating the capacities.

The obtained intensities for one-, two-, five-, and ten-minute intervals were plotted into histograms with appropriate intervals. Due to the variable sample size, the histograms are displayed in percentages. The tables with absolute values are also provided and were, in case of the complete data, used for estimation of the theoretical capacity. There were also other supplemental analyses performed, such as drawing a graph of intensity over time around the forming of congestion or calculating share of intensity over 1800 veh/h as that was considered as a target value. It may not be desirable to aim for even higher intensity, even though the theoretical capacity might be even higher, as it would lead to low headways and therefore increased risk of accidents and consequently possible complete blockage of the road (besides the obvious safety risks). Also, increasing the density too much could lead to congestion on the single-lane section of the road. The actual value of desired intensity might be higher and is a matter of discussion and further research, the value 1800 veh/h was based on the commonly cited two second rule for headway.

All the capacity estimations are location-specific (and partially time-specific as the composition of traffic flow may also vary over time) as circumstances such as share of slow/long vehicles (interchangeable with trucks in this text) or incline of the road can play significant role in reducing the capacity. This should be considered especially when comparing capacities on different locations with significantly different road geometry and/or traffic flow composition.

3.3.1 Brno

According to the traffic counting at 2010 (Celostátní sčítání dopravy, 2010) there is about 20-percent share of freight vehicles on work days on average (on AADT), but only about five percent of that are listed in the categories that are considered slow and long vehicles, while the rest are light and medium freight vehicles. That is in line with the about one-hour-long on-site measurement which showed 4.8-percent share of vehicles considered slow/long.

About eight hours of intensity measurements was done in total in Brno, over five of which was during congestion. However, at the beginning the measurement interval was set to ten minutes which significantly reduces the amount of data it could provide.

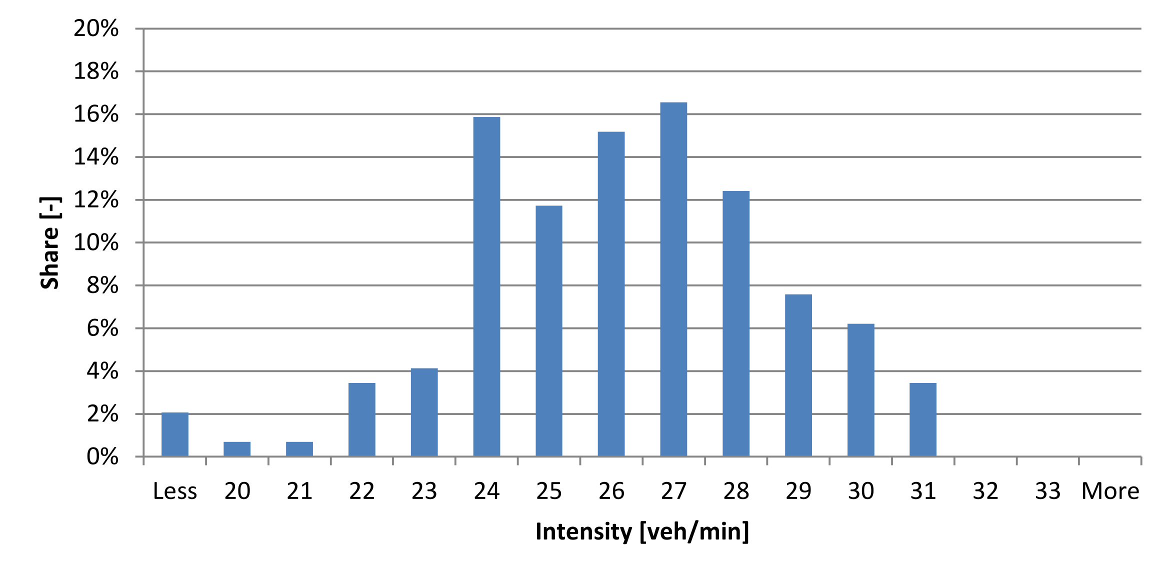


Figure 3.7 Histogram of Brno one-minute interval intensity during congestion

Figure 3.7 shows the measured intensities per one-minute interval in a histogram. Usually, 24-28 vehicles passed the cross-section per minute and the intensity never exceeded 31 veh/min. Intensities below 24 veh/min are also quite rare and most of them could probably be accounted to extraordinary situations, like a crossing pedestrian or a traffic jam downstream. Given the low share of 31 veh/min and the 95th percentile being 30 veh/min, the congestion capacity per one-minute interval is estimated to 30 veh/min (equals 1800 veh/h). The intensity exceeds the desired intensity 30 veh/min (based on 1800 veh/h) about ten percent of time.

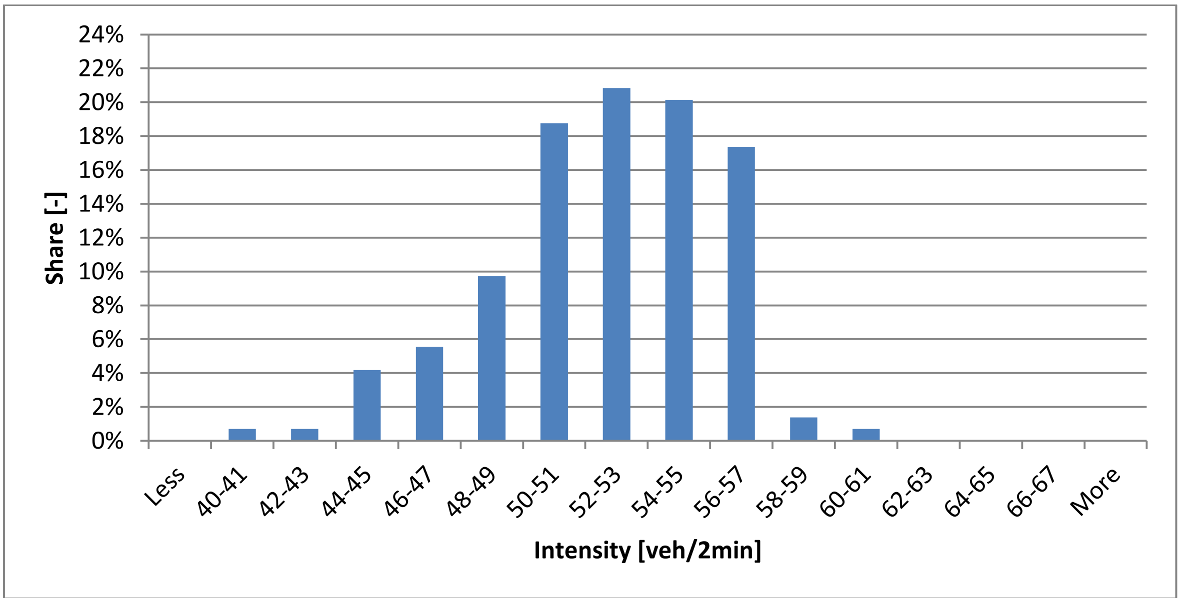


Figure 3.8 Histogram of Brno two-minute interval intensity during congestion

Figure 3.8 shows the vehicle count per two-minute interval. The capacity is very clear in this case – 57 veh/2min or 1710 veh/h/. There is also less extreme values since the short peaks get smoothened. Intensity exceeding equivalent of 1800 veh/h is only reached at less than one percent of the time for the two-minute interval. For the longer intervals, the threshold has not been exceeded even once during the measurements.

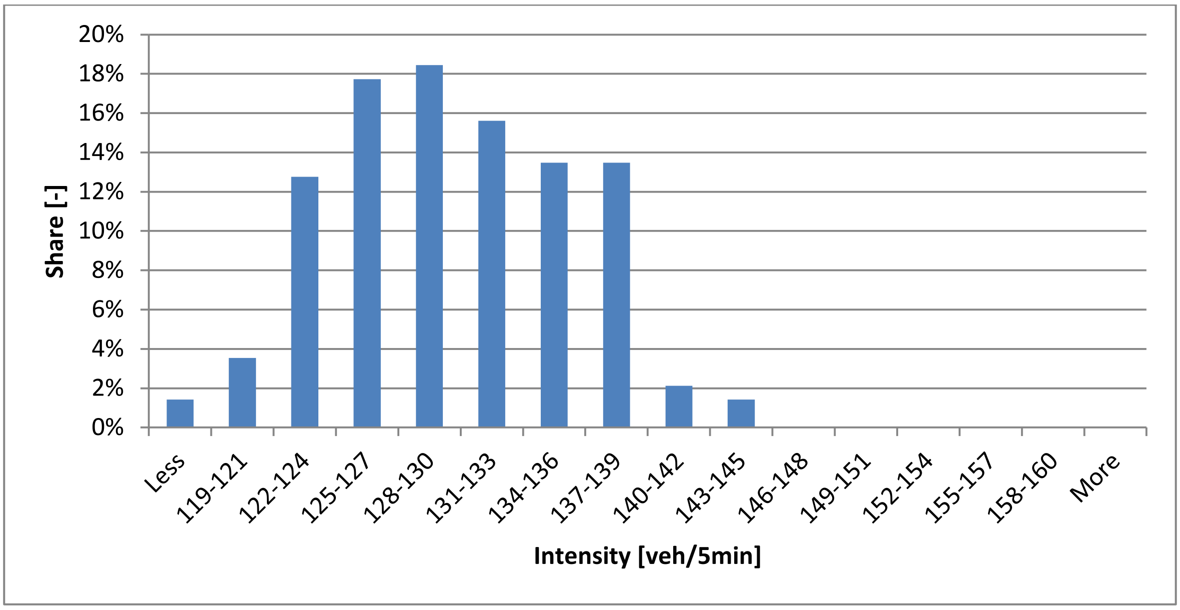


Figure 3.9 Histogram of Brno five-minute interval intensity during congestion

The congestion capacity for five-minute interval is also quite clear, as the Figure 3.9 shows. Intensities over 140 veh/5 min are quite rare and also the 95th percentile supports the estimated capacity of 139 veh/5min (i.e. 1668 veh/h). There is relatively high variance for this interval with intensities ranging from 122 to 139 have all quite high shares. The cause of this is probably the fact that for short intervals, even though the relative variability may be higher, the vehicle count is relatively low and thus is also the variance, while for the longer intervals the peaks get smoothened.

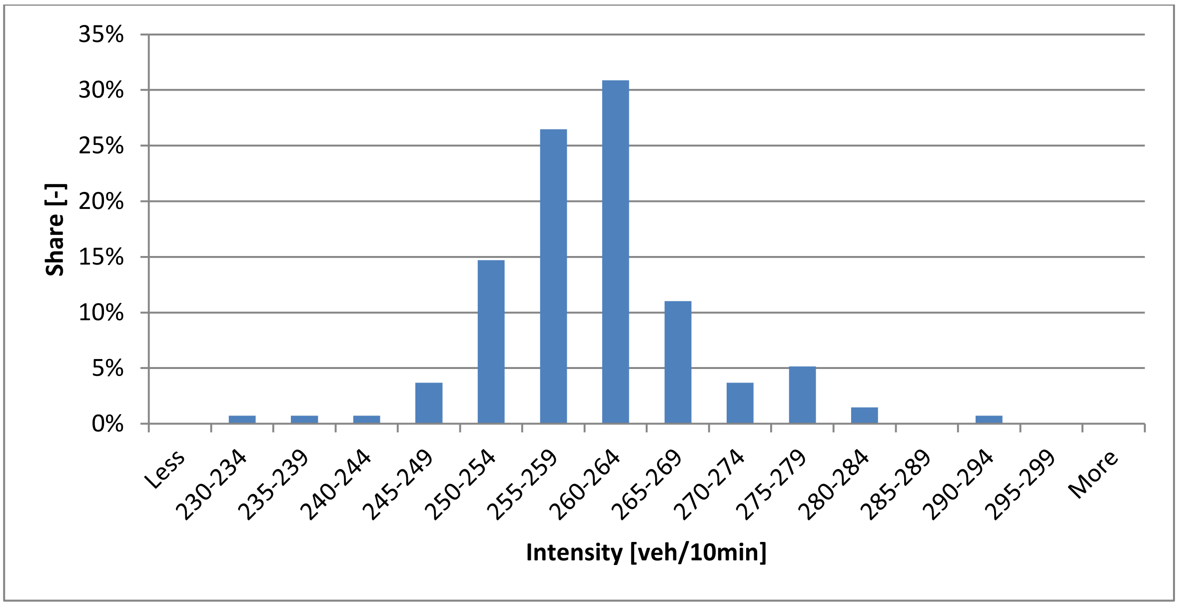


Figure 3.10 Histogram of Brno ten-minute interval intensity during congestion

Finally, Figure 3.10 shows the measured intensities over 10-min interval. In few cases, very low intensities below 250 veh/10min were recorded, most likely caused by series of unpropitious events. Oddly, there was not a single interval with 285-289 vehicles but once 290 vehicles were recorded. The 95th percentile suggests 277 veh/10min as possible capacity estimate but given the length of the interval and repeated occurrence of higher intensities, higher capacity can be assumed. Larger sample with more high-intensity records would be needed to estimate the capacity more accurately.

Table 3.2 Share of intensities exceeding the equivalent of 1800 veh/h during congestion in Brno

Share of intensity ≥ 1800 veh/h				
Interval	Threshold	Total	Above threshold	Share
10 min	299	136	0	0,00%
5 min	149	141	0	0,00%
2 min	59	144	1	0,69%
1 min	29	145	14	9,66%

In the real world traffic flow, the desired intensity (or its equivalent) of 1800 veh/h is only reached for very short periods. As the Table 3.2 shows, only 10 percent of the time the intensity is equal to or above the 1800 veh/h threshold. This intensity has never been sustained for five or more minutes in Brno.

Table 3.4 shows, the average intensity during congestion is only 1560 veh/h and even the 95th is only 1662 veh/h. This suggests a decent potential for improvement, especially when compared to Allerød.

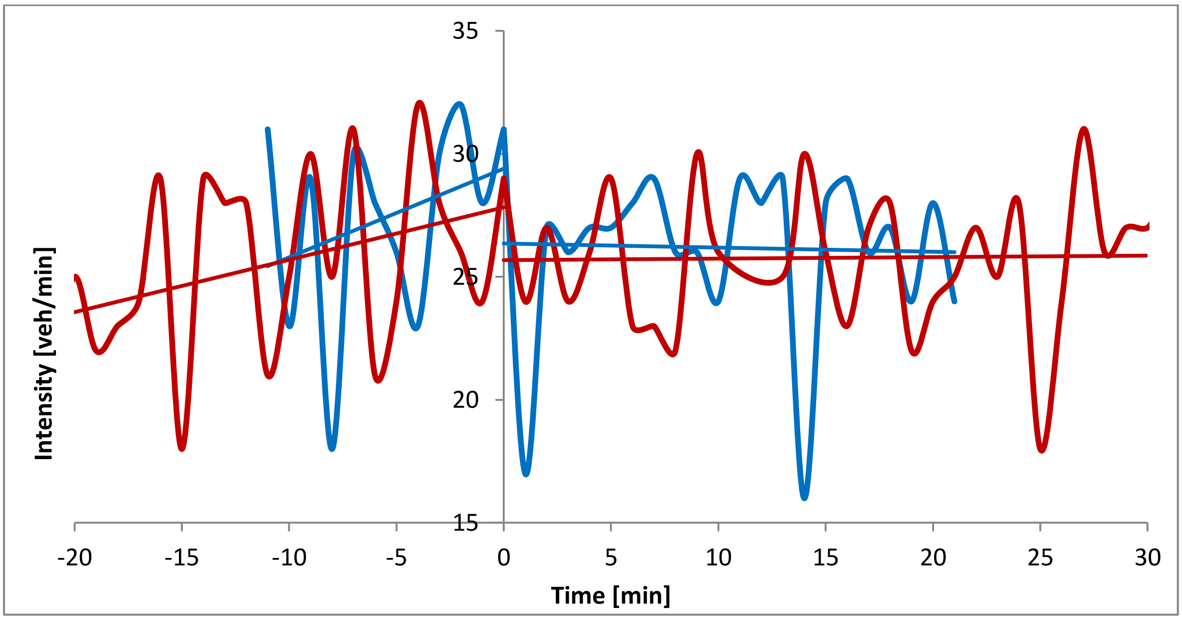


Figure 3.11 Intensity over time before and after congestion formed in Brno with linear interpolation curve included. Two measurements differentiated by colour.

Plotting one-minute intensities measured around the time when the congestion formed shows the congestion-induced capacity drop (Figure 3.11). Adding trendlines separately for pre-congestion and congestion reveals the increasing trend before the flow collapses and the queue forms. In both cases, pedestrians on the crossing causing halt to the flow were the cause of the traffic flow breakdown and creation of the queue. After that the intensity drops and is sustained at the same level, on average. The fluctuations are caused by the stochastic nature of the whole traffic flow phenomenon and the sharp drops are most likely caused by lack of demand prior to the congestion formed and by a pedestrian using the downstream crossing during the congestion.

Complete distribution of the measured intensities during congestion can be seen in Table 3.3. It is the table that was used to draw the histograms above, however the table shows the actual counts and not just the shares and also shows the sample size on the bottom line, which is important when evaluating the accuracy and interpreting the results. Even though the measurements during congestion were quite long in Brno, due to the ten-minute interval measurements mentioned earlier, the sample size is not as large as it could be if one-minute interval had been used from the beginning.

Table 3.3 Table of intensities during congestion in Brno, used for drawing of the histograms above, with the actual counts

1 min interval		2 min interval		5 min interval		10 min interval	
Vehicles	*Frequency*	*Vehicles*	*Frequency*	*Vehicles*	*Frequency*	*Vehicles*	*Frequency*
Less	3	Less	0	Less	2	Less	0
20	1	40-41	1	119-121	5	230-234	1
21	1	42-43	1	122-124	18	235-239	1
22	5	44-45	6	125-127	25	240-244	1
23	6	46-47	8	128-130	26	245-249	5
24	23	48-49	14	131-133	22	250-254	20
25	17	50-51	27	134-136	19	255-259	36
26	22	52-53	30	137-139	19	260-264	42
27	24	54-55	29	140-142	3	265-269	15
28	18	56-57	25	143-145	2	270-274	5
29	11	58-59	2	146-148	0	275-279	7
30	9	60-61	1	149-151	0	280-284	2
31	5	62-63	0	152-154	0	285-289	0
32	0	64-65	0	155-157	0	290-294	1
33	0	66-67	0	158-160	0	295-299	0
More	0	More	0	More	0	More	0
Total	145	Total	144	Total	141	Total	136

Table 3.4 provides a statistical overview on the distribution of the congestion intensities including minima and maxima, means, and percentiles. The column 'Measured' has the actual measured vehicle count over the one-minute interval. The 'Real' column provides two, five and ten minute interval intensities obtained as a sum of two, five and ten consecutive one-minute intervals, respectively. Even though the consecutive one-minute vehicle counts were not necessarily measured over two consecutive minutes (the measurement had to be occasionally interrupted to set the camera, measurements from different days are include, etc.), the share of such cases is not high and is not expected to cause significant systematic error. Eliminating those cases would also greatly reduce the sample size of the long intervals and consequently reducing the credibility of the results much more. The column 'Theoretical' is only multiplication of the corresponding sub-column of the 'Real'

column to one-hour intensity which is easily comparable among the different time intervals.

While the previous graphs and tables dealt with congestion intensity and capacity, the Table 3.5 below provides distribution of the complete data from Brno, including that pre-congestion. As explained at the beginning of chapter 3.3, the distribution of intensities is dependant of the share of pre-congestion data and cannot be effectively used for many analyses. Nonetheless, it can be used to estimate the theoretical capacity of the merging area, or any road cross-section.

Table 3.4 Overall statistics of intensity during congestion for the Brno merging

Congestion intensity Brno	Intensity						
	Measured	Real			Theoretical		
	[veh/min]	[veh/2min]	[veh/5min]	[veh/10min]	[veh/h]	[veh/h]	[veh/h]
Min	16	41	117	233	1230	1428	1398
Max	31	60	145	290	1800	1716	1740
Mean	26	52	130	260	1560	1559	1562
Median	26	52	130	260	1560	1560	1560
10th percentile	23	47	123	251	1380	1476	1504
20th percentile	24	49	125	254	1470	1500	1524
80th percentile	28	55	136	265	1650	1620	1590
90th percentile	29	57	138	271	1710	1644	1626
95th percentile	30	57	139	277	1710	1667	1662

Repeated occurrence of a given intensity (or intensity range) is deemed necessary to provide sufficient confidence when estimating the theoretical capacity as even when regulating the traffic flow perfect homogeneity cannot be achieved in real world and the highest measured intensities are likely to be measured with zero or low share of trucks, especially in the case of short intervals. Based on these assumptions, following estimates of theoretical capacity can be done: 31 veh/min, 60 veh/2min, 143 veh/5min and 280 veh/10 min. Since the longer intervals are inevitably laden with phases of lower intensity, the shorter intervals are more suitable for estimating the theoretical capacity. Given that intensity 31 veh/min was twice recorded including a single slow/long vehicle and the share of such vehicles and given the share of trucks (implying roughly one truck per minute on average), it might be possible that such intensity may be sustainable over longer periods by regulating the traffic flow. The two-minute interval capacity can be used to get more conservative estimate for long-term theoretical capacity. In that case, it can be estimated to 1800 veh/h based of 60 veh/2min, which is equal to the intensity considered desirable.

The theoretical capacity estimates above correspond approximately to 98th percentile. That is slightly more than the 95th used as guidance for the congestion capacity, however the sample size is almost double and many of the pre-congestion intervals had low intensity due to the low demand which necessarily increases the share of low intensities in the sample and thus a higher percentile should be used as guide when estimating the capacity.

Table 3.5 Table of intensities in Brno including pre-congestion measurements

1 min interval		2 min interval		5 min interval		10 min interval	
Vehicles	*Frequency*	*Vehicles*	*Frequency*	*Vehicles*	*Frequency*	*Vehicles*	*Frequency*
Less	35	Less	28	Less	67	Less	56
20	6	40-41	8	119-121	7	230-234	3
21	9	42-43	9	122-124	28	235-239	8
22	10	44-45	15	125-127	27	240-244	2
23	18	46-47	19	128-130	38	245-249	18
24	30	48-49	26	131-133	35	250-254	25
25	25	50-51	36	134-136	27	255-259	45
26	32	52-53	40	137-139	26	260-264	52
27	29	54-55	41	140-142	4	265-269	26
28	28	56-57	32	143-145	4	270-274	12
29	17	58-59	7	146-148	1	275-279	8
30	16	60-61	4	149-151	0	280-284	2
31	9	62-63	2	152-154	0	285-289	0
32	3	64-65	0	155-157	0	290-294	2
33	1	66-67	0	158-160	0	295-299	0
More	0	More	0	More	0	More	0
Total	268	Total	267	Total	264	Total	259

3.3.2 Allerød

The AADI is unknown here but the measured share of trucks based on two measurements over two hours long in total, the average share of slow/long vehicles is about 3.1 %. Close to six hours of intensities was recorded in Allerød, almost five of which was during congestion. All the measurements in Allerød were recorded in one-minute intervals which helped to increase the sample size given the way the data was processed (see chapter 3.3 for explanation).

Figure 3.12 shows histogram of one-minute intensities during congestion in Allerød. Much greater dispersion can be noticed in comparison to Brno. There is quite large share of intensities exceeding 30 veh/min, which were only rarely reached in Brno. On the other hand, there is also much higher share of low intensities below 25 veh/min. Based on observation of the traffic flow, the cause appears to be its different drivers' behaviour compared to Brno (see chapter 3.4 for details). It can be said with confidence that the congestion capacity in Allerød is higher than in Brno, reaching 33 veh/min regularly while in Brno 31 veh/min was the maximum. In Allerød the recorded maximum was 35 veh/min.

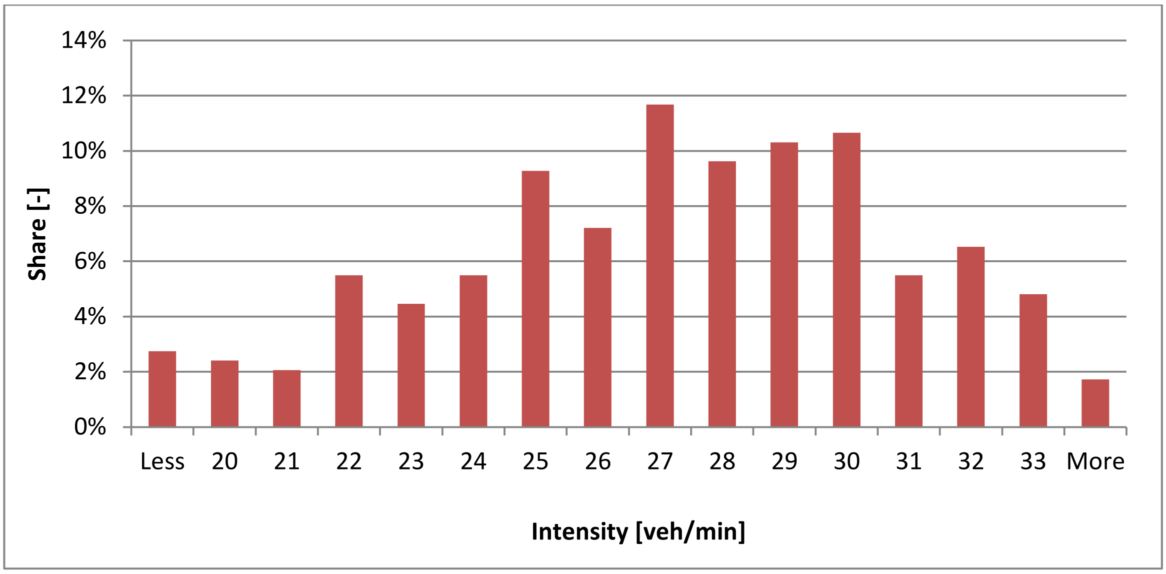


Figure 3.12 Histogram of Allerød one-minute interval intensity during congestion

The two-minute interval shows similar trends – higher dispersion and higher capacity. The difference is perhaps even more apparent here. While the congestion capacity in Brno was estimated to 57 veh/2min, in Allerød almost one third of the records exceed that value. Even though Brno had lower share of low intensities below 24 veh/min, it has similar share of low intensities below 50 veh/2min for the two-minute interval given that the low peaks get smoothened over the longer interval. The capacity is estimated to 63 veh/2min.

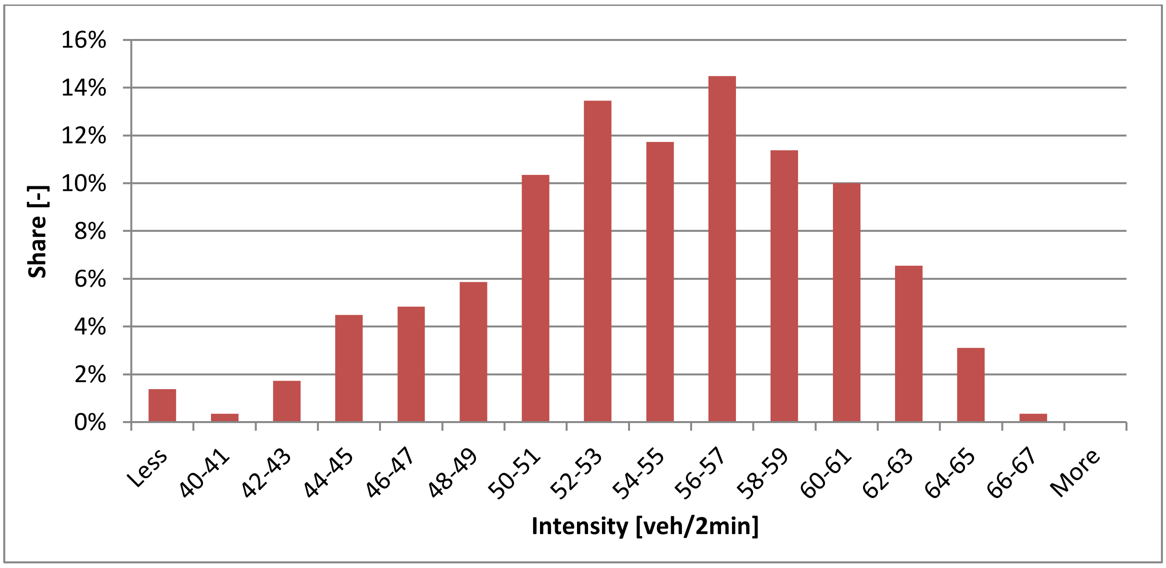


Figure 3.13 Histogram of Allerød two-minute interval intensity during congestion

The histogram of five-minute interval intensities gives the same indications as the previous. Although there is slightly higher share of intensities below 122 veh/5min in Allerød, there is much higher average and maximal intensity. Again, in more than one third of the cases, the intensity in Allerød exceeds the estimated capacity (139 veh/5min) in Brno. The maximal intensity per five minutes reached 157 vehicles in Allerød and the estimated capacity is 149 veh/5min.

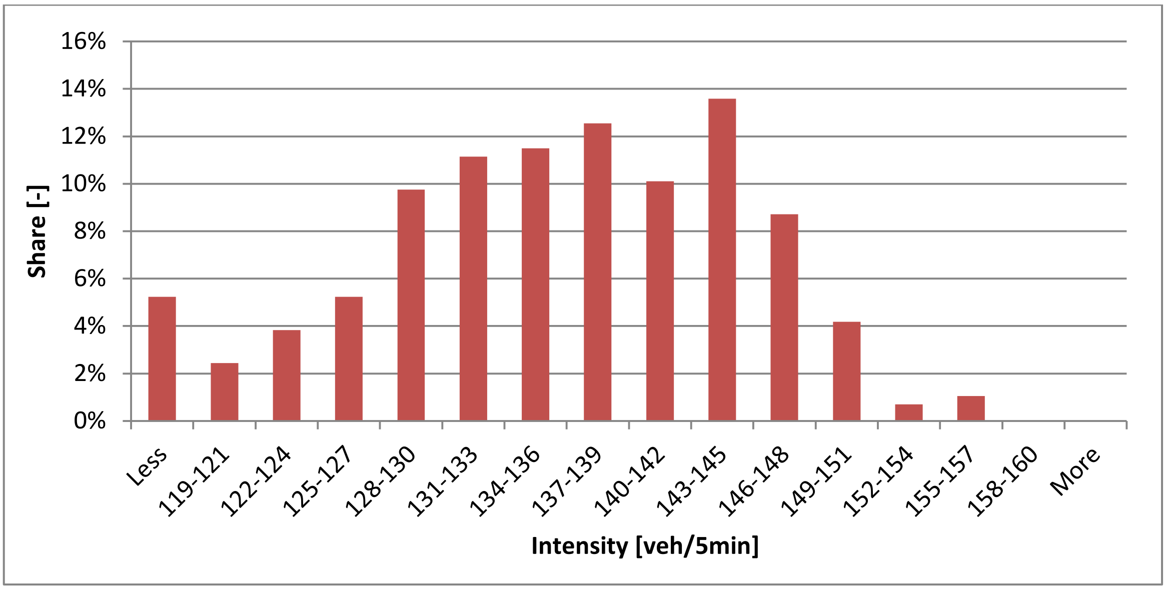


Figure 3.14 Histogram of Allerød five-minute interval intensity during congestion

The ten-minute interval is most relevant for long-term performance and the distribution of intensities unsurprisingly mirrors the shorter intervals. Yet again, third of the recorded intensities exceed the capacity of the merge in Brno. Besides the obvious confirmation of higher capacity of the Allerød merge, the histogram in Figure 3.15 shows noteworthy frequency drop between 270 and 280 veh/10min creating sort of double-peak distribution. The size of the sample and the magnitude of the drop hints rather on some systematic cause than on purely game of luck and similar, albeit less severe, drop can also be seen in the shorter intervals. However, the cause of this oddity is unknown – perhaps it has to do with the behaviour of the traffic flow an Allerød during merging. The estimated capacity over ten-minute interval is 294 vehicles.

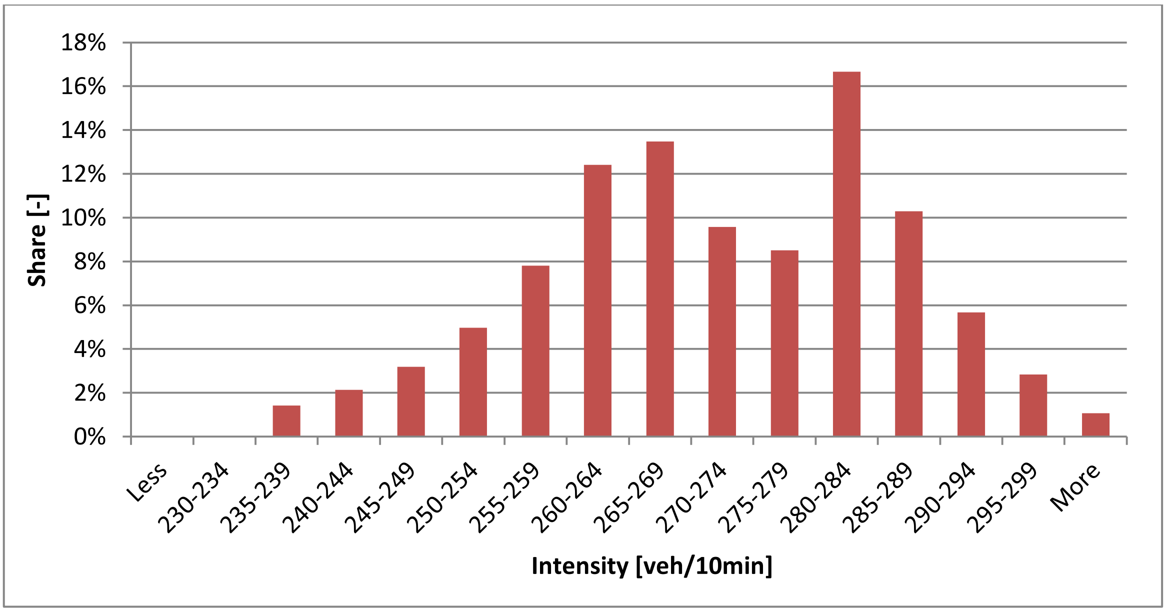


Figure 3.15 Histogram of Allerød ten-minute interval intensity during congestion

Just like in Brno, there were two measurements done at the time of congestion forming to capture the process and evaluate the capacity drop. The two measurements are plotted in the Figure 3.16. Similarly as in Brno, the intensity grew, on average, until the flow collapsed and a queue formed. After that the intensity oscillates around a constant due to the fluctuations in the traffic flow. As the histograms showed, the fluctuations are slightly higher compared to Brno but there are no pedestrian-induced drops and all the fluctuations come purely from the traffic flow.

Whereas in Brno the queue was formed after a pedestrian crossed, the process was completely natural here. At the moment when too many cars arrive to the merging area within a short period of time, the traffic flow collapses. The vehicles at the front are forced to slow down due to the increased density but are able to merge without stopping. However, when the number of incoming vehicles is too high and the difference between the 'output' and the 'input' gets too high, the density grows too high and the incoming vehicles are forced to stop and a queue forms. Due to the extremely high density at the merging area and even after it, the flow often comes to a complete stop for a moment before the single-lane clears and the vehicles can start leaving the merging area again. This causes severe intensity drop which is clearly visible on the red curve in the Figure 3.16.

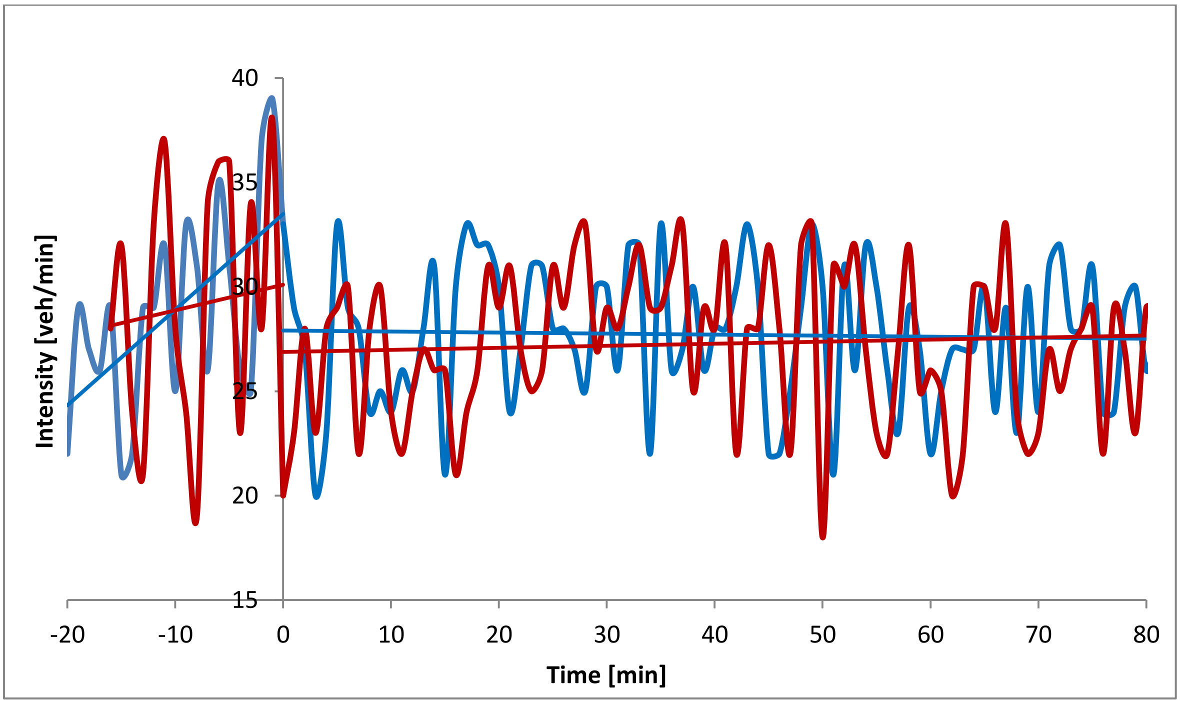


Figure 3.16 Intensity over time before and after congestion formed in Allerød with linear interpolation curve included. Two measurements differentiated by colour.

Similar effect stands behind the more severe intensity fluctuations in Allerød. Given the drivers behaviour (see chapter 3.4), the intensity can get quite high compared to Brno, even during the congestion. However, as also the following single-lane has a limited capacity and is constantly saturated by the merging vehicles, the high merging intensity regularly causes breakdown on the single-lane. Due to the high density the vehicles in the single-lane cannot accelerate fast enough to make room for the merging vehicles and those are forces to stop again, similarly as when the queue forms before the merge. Analogically, a drop in intensity follows. High short-term intensity is needed for this to happen and a short period of very low intensity is the consequence, hence the fluctuations over the short intervals are much higher in Allerød compared to Brno.

Table 3.6 Table of intensities during congestion in Allerød, used for drawing of the histograms above, with the actual counts

1 min interval		2 min interval		5 min interval		10 min interval	
Vehicles	*Frequency*	*Vehicles*	*Frequency*	*Vehicles*	*Frequency*	*Vehicles*	*Frequency*
Less	8	Less	4	Less	15	Less	0
20	7	40-41	1	119-121	7	230-234	0
21	6	42-43	5	122-124	11	235-239	4
22	16	44-45	13	125-127	15	240-244	6
23	13	46-47	14	128-130	28	245-249	9
24	16	48-49	17	131-133	32	250-254	14
25	27	50-51	30	134-136	33	255-259	22
26	21	52-53	39	137-139	36	260-264	35
27	34	54-55	34	140-142	29	265-269	38
28	28	56-57	42	143-145	39	270-274	27
29	30	58-59	33	146-148	25	275-279	24
30	31	60-61	29	149-151	12	280-284	47
31	16	62-63	19	152-154	2	285-289	29
32	19	64-65	9	155-157	3	290-294	16
33	14	66-67	1	158-160	0	295-299	8
More	5	More	0	More	0	More	3
Total	291	Total	290	Total	287	Total	282

Table 3.7 Overall statistics of intensity during congestion for the Allerød merging

Congestion intensity Allerød	Intensity						
	Measured	Real			Theoretical		
	[veh/min]	[veh/2min]	[veh/5min]	[veh/10min]	[veh/h]	[veh/h]	[veh/h]
Min	15	32	99	237	960	1188	1422
Max	35	66	157	301	1980	1884	1806
Mean	27	54	136	272	1630	1630	1630
Median	27	55	137	272	1635	1644	1632
10th percentile	22	47	123	252	1410	1476	1512
20th percentile	24	50	128	260	1500	1536	1560
80th percentile	30	60	144	284	1794	1728	1706
90th percentile	32	62	147	289	1857	1764	1734
95th percentile	33	63	149	294	1890	1788	1763

Table 3.6 above provides complete data on intensity distribution in Allerød during congestion. The sample size is over double compared to Brno making it, along with the pedestrian crossing issue in Brno, more reliable and representable of a general lane merging.

Table 3.8 Share of intensities exceeding the equivalent of 1800 veh/h during congestion in Allerød

Share of intensity ≥ 1800 veh/h				
Interval	Threshold	Total	Above threshold	Share
10 min	299	282	3	1,06%
5 min	149	287	13	4,53%
2 min	59	290	58	20,00%
1 min	29	291	85	29,21%

Table 3.7 provides some basic statistics such as percentiles, mean and extremes of the Allerød intensities. The 95th percentile was used as an important guide for estimating the congestion capacity. The average intensity during congestion is 1630 veh/h compared to 1560 veh/h in Brno. Table 3.8 shows that intensities over 1800 veh/h are commonly reached short-term in Allerød and sometimes even over periods exceeding ten minutes. One third of one-minute intervals has intensity at least 30 veh/min and one fifth of two-minute interval at least 60 veh/2min even in unregulated traffic flow.

Table 3.9 Table of intensities in Allerød including pre-congestion measurements

1 min interval		2 min interval		5 min interval		10 min interval	
Vehicles	*Frequency*	*Vehicles*	*Frequency*	*Vehicles*	*Frequency*	*Vehicles*	*Frequency*
Less	11	Less	5	Less	28	Less	4
20	8	40-41	2	119-121	11	230-234	2
21	8	42-43	9	122-124	13	235-239	9
22	20	44-45	15	125-127	19	240-244	14
23	17	46-47	17	128-130	31	245-249	10
24	20	48-49	19	131-133	34	250-254	19
25	32	50-51	38	134-136	31	255-259	21
26	26	52-53	45	137-139	41	260-264	36
27	35	54-55	40	140-142	35	265-269	42
28	35	56-57	48	143-145	44	270-274	28
29	36	58-59	36	146-148	31	275-279	29
30	32	60-61	34	149-151	17	280-284	52
31	19	62-63	22	152-154	5	285-289	34
32	22	64-65	14	155-157	6	290-294	26
33	16	66-67	2	158-160	0	295-299	11
More	14	More	4	More	1	More	5
Total	351	Total	350	Total	347	Total	342

As it was explained in chapter 3.3.1 above, the two-minute interval is assumed to have the optimal length for estimating the theoretical long-term capacity. Table 3.9 sums the entire recorded intensities in Allerød. From the frequency of the recorded two-minute intensities the two-minute capacity can be estimated to 65 vehicles, which equals 1950 veh/h.

3.3.3 Ballerup

The lane merging in Ballerup is very specific thanks to the signal-controlled intersection about 200 m upstream. It serves as a real-world experiment of a very specific kind of traffic flow regulation. It has been chosen for that reason to explore how the signal-controlled intersection affects the merging process and the capacity of the merge.

Given the specific circumstances the data processing has been slightly altered to provide more meaningful information. The intensities were measured twice in Ballerup. The first time, there was a queue from the intersection downstream and after that the demand was already too low to measure capacity, thus the data was not used in the analysis. The second measurement was about one and half hour long and about half of it was also with sub-capacity demand (the queues at the intersection resolved during each green phase). For the reminder of the measurement, the capacity of the intersection was reached and was the limiting factor of the lane merge capacity. The data from that period is used in the following analysis. Most of the measurement was recorded on camera and the share of trucks was calculated to the average 1.4 % during the measurement.

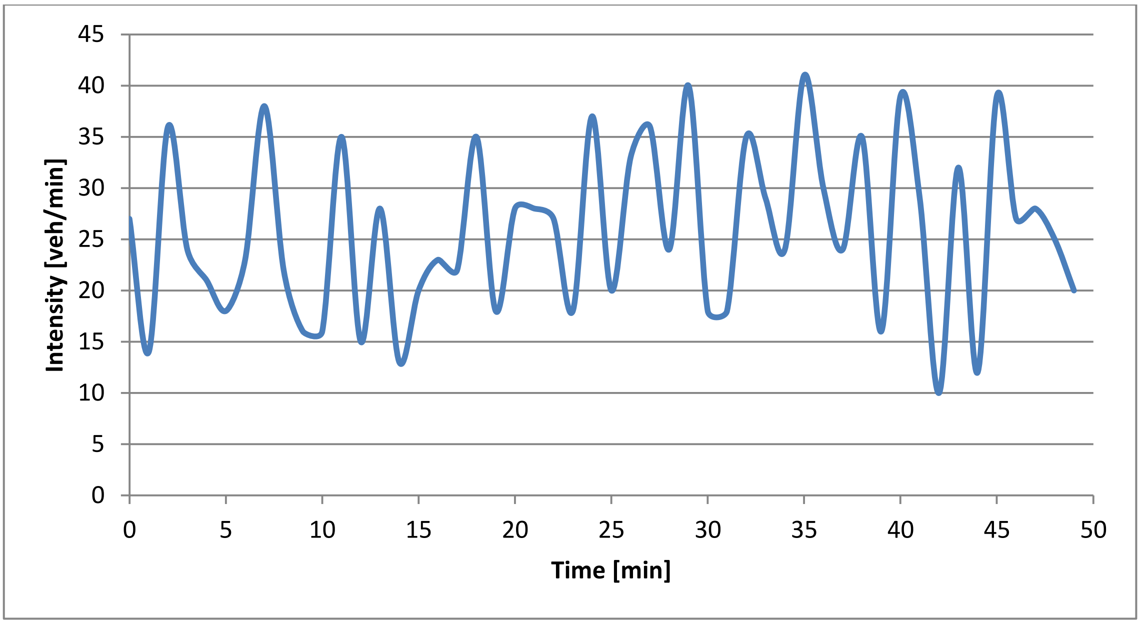


Figure 3.17 Progression of one-minute intensities at the time the capacity of the upstream intersection was reached

As can be seen on Figure 3.17, the intensity at the lane merge is determined by the vehicle 'waves' or platoons coming from the intersection. When the start of the wave is about the same as the start of the measured interval, the number of passing vehicles regularly exceeds 35 veh/min and reaches as high as 41 veh/min. It has to be noted that during the wave there are two small gaps in the incoming vehicles between the different turns at the intersection. However, then a gap without any vehicles follows and the intensity is low due to the lack of demand. Therefore, similarly to pre-congestion in Brno or Allerød, the intensity is not limited by the merging but rather by the amount of the incoming vehicles and thus the actual road or congestion capacity at the lane merge is unidentifiable. However, the maximal reached intensities can be used to estimate the theoretical capacity.

Table 3.10 provides more thorough information on the merging intensity. Using the 95th percentile as the estimate of the theoretical capacity it can be seen that it exceeds the theoretical capacities at Brno and even in Allerød. The difference is not large for longer intervals due to the very significant gaps between the waves but there is very significant difference for the shorter intervals.

Table 3.10 Statistics on Ballerup merging intensity at the intersection’s capacity

Ballerup intersection on capacity	Intensity						
	Measured	Real			Theoretical		
	[veh/min]	[veh/2min]	[veh/5min]	[veh/10min]	[veh/h]	[veh/h]	[veh/h]
Min	10	32	99	210	960	1188	1260
Max	41	71	159	302	2130	1908	1812
Mean	26	52	129	259	1546	1551	1557
Median	25	53	128	266	1590	1536	1596
10% percentile	15	39	109	225	1170	1309	1351
20% percentile	18	42	116	229	1260	1390	1373
80% percentile	35	60	144	282	1800	1728	1693
90% percentile	38	65	150	293	1950	1804	1760
95% percentile	39	69	154	298	2055	1844	1786

Still, the values in Table 3.10 do not actually show the true one-minute capacity. As the measurement intervals were not synchronized with the signal controls cycle due to the different length, the waves would often divide into two consecutive intervals. To estimate the real short-term capacity, another measurement was performed on the recorded video where the one-minute interval was set to start at the time when the first vehicle of the wave would reach the chosen cross-section. Occasionally, the length of the measurement interval was extended by several seconds to capture the whole wave.

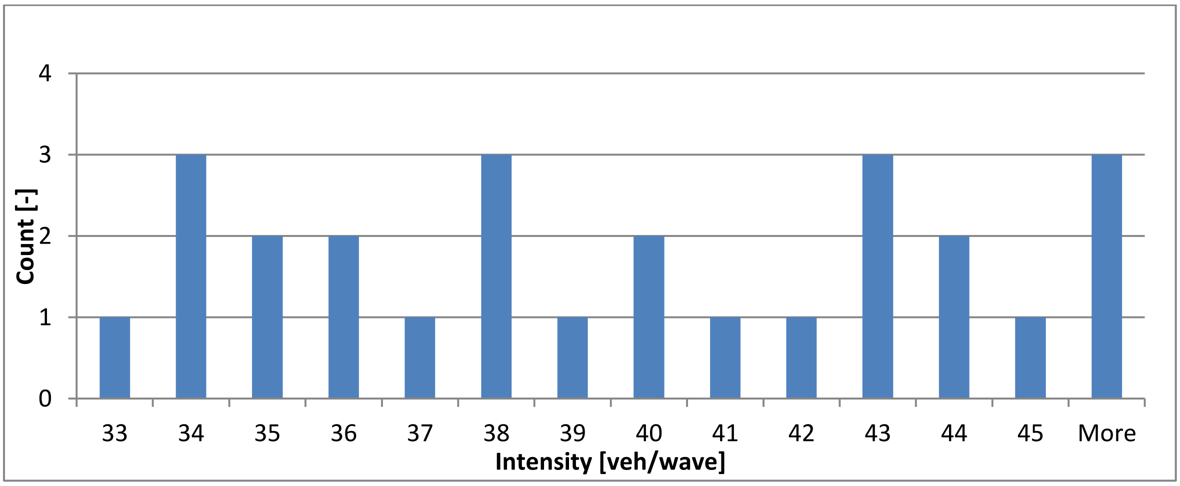


Figure 3.18 Histogram of ‘wave’ intensity in Ballerup (measurement interval 60-70 s)

The eight lowest waves were cut out from the Figure 3.18 to save space. The maximal number of vehicles per one wave was 48. It is obvious that measuring the intensity per wave renders even higher records. Yet, even those waves usually included small gaps between the individual directions on the intersection.

However, one must not think that without the gaps the intensity would increase. On the contrary, the gaps are what make it possible to reach such high intensities. The gaps allow each wave (or sub-wave) to leave the merging area and the following single-lane before the next wave arrives. Further, each wave contains limited number of vehicles which are able to merge without stopping or only several of the last vehicles are forces to stop. The merging is also enhanced by the higher speed at which the wave arrives to the merge and by the fact that the vehicles often have longer headway because they started only about 200 m downstream at the intersection and the faster vehicles did not have enough time to reach the slower ones. Together, those circumstances allow higher speed and lower density when reaching the merging area and together with the free road downstream and limited wave size make possible the higher merging intensity.

The wave intensity is not sustainable over long periods of time as the gaps would have to be dramatically reduced and there would not be enough time for the vehicles to leave the merging area before another wave of vehicles arrives. However, the real-world intensities over the five- and ten-minute intervals are real-world proof that by regulating the stream of incoming vehicles the capacity of a merging area can be significantly improved. Based on the rather limited measurements, the estimated theoretical capacity is 68 veh/2min, i.e. 2040 veh/h. It is however assumed, that by optimizing the gap length and wave/platoon size, even higher intensity might be reached as the gaps were often up about 20 s and the waves had very unequal number of vehicles.

3.4 MEASUREMENT SUMMARY AND CONCLUSIONS

There are two main conclusions from the measurements. First, there is significant difference between the different merging areas, the reasons for which are discussed below. Second, the lane merging in Ballerup served as a kind of real-world test of the proposed performance improving measures discussed in chapter 4 and proved its high potential.

The merging in Ballerup is very specific thanks to the signal-controlled intersection and its different performance and behaviour is not surprising. The quite different road/merge capacities in Brno and Allerød are much more interesting. There are several possible (and on the first glance perhaps even obvious) reasons behind this difference. Observations of the traffic flow, along with the data analysis, allow identifying the most important reason.

Obviously, the pedestrian crossings play a large role in Brno, both as a trigger of congestions and as capacity reducer. Another, albeit less important, reason for the lower intensities in Brno is the slightly higher share of trucks. However, the probably most important reason is the different drivers' behaviour. That is also possibly connected to the different merging area geometry, which is the last important difference.

Pedestrians and trucks negatively affect the average intensity in Brno but since the highest intensities are reached under optimal circumstances (i.e. there are any pedestrians or trucks) they should not have large impact on the estimated theoretical capacity. That leaves the drivers' behaviour and the merging area geometry as the reasons for the eight-percent difference in the theoretical capacity (1800 veh/h vs. 1950 veh/h).

Observation of the merging vehicles during congestion in Brno and Allerød revealed important difference in the drivers' behaviour. In Brno the queue density was very high all the way to the merging point as the drivers were driving without any large gaps, likely in order to prevent drivers in the other lane to merge in front of them. While most of the drivers followed the rules of the zipper merge, due to this high density of the traffic flow the vehicles were forced to accelerate from low or zero speed. The need to accelerate is considered the main reason for the congestion-induced capacity drop (Tympakianaki et al., 2014). On the other hand, the drivers in Allerød would usually leave much larger gaps between the vehicles as they approached the merging area in order to allow the vehicles from the other lane to merge fluently come the merging point. This allows them to merge without stopping and with lower density which also enables faster acceleration. In fact, the merging was often so fluent that the single-lane got 'saturated' by vehicles and the traffic flow stopped there. Inevitably, the density also increased in the merging area and the merging stopped until the way cleared. This behaviour also eliminated stop-and-go for most of the time, at least in the low-density area.

Figure 3.19 Illustration of the change of density about 200 m before the merging area in Allerød

It is possible that the longer merging area in Allerød, together with the different horizontal road signs, more or less encourage this behaviour. Nevertheless, the main reason is probably higher discipline, and consequently higher willingness to cooperate, of the Danish drivers.

Measuring density is little bit more complicated, time consuming and requires view on sufficiently long stretch of road in order to be meaningful and so it was not properly performed in Denmark. Moreover, the density was very variable which would further complicate any measurements. Nonetheless, the difference in density between Brno and Allerød is obvious when observing the merging. Calculation of vehicles on several videos from Allerød revealed that when the described behaviour is taking place, the density is ranging circa from 60 to 90 veh/km. For comparison, in Brno the average queue density was about 110 veh/km and never dropped below 100 veh/km, although the sample was also quite small (Mikolášek, 2014). The difference was also visible when comparing density close to the merging and further up the stream as can be seen on Figure 3.19.

Figure 3.20 Lower traffic flow density before the merging area in Allerød

The measurements in Ballerup, however short, put solid background for the hypothesis that using traffic lights to regulate traffic flow could bring nonnegligible performance improvement. Even the adjacent intersection with the signal plan not designed with any intention to improve the merging regulated the traffic flow in such a way that the average intensity was equal to Brno, despite the long gaps between the waves. More importantly, the theoretical capacity is over 2000 veh/h. Assuming 2000 veh/h as the average intensity, it would mean increase of 440 veh/h in Brno and 370 veh/h in Allerød (i.e. increase by about 28 and 23 percent, respectively).

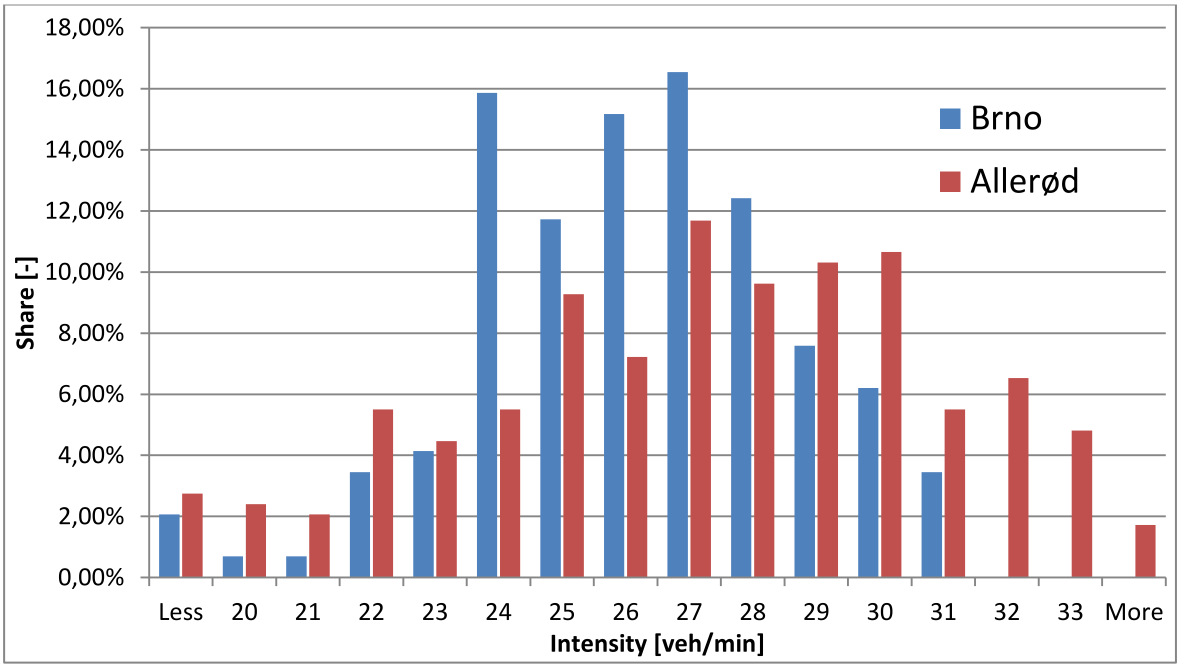


Figure 3.21 Comparison of intensities over one-minute interval in Brno and Allerød

To sum up the measurements – the merge capacity in Brno is 1560 veh/h, in Allerød it is 1630 veh/h and it is 1560 veh/h in Ballerup. The congestion capacity is very abstract number and does not serve any particular purpose. It is mostly good for comparing the potential of the merges and proved useful for identifying the drivers’ behaviour as important reason for the lower capacity in Brno. As there was no congestion in Ballerup, the congestion capacity could not be estimated there. The estimates for individual intervals for Brno and Allerød can be found in their respective chapters. The theoretical capacities were estimated based on two-minute intervals to 1800, 1950, and 2040 veh/h in Brno, Allerød, and Ballerup, respectively. Most important in this regard is Ballerup is it has the highest theoretical capacity. It is assumed that under similar conditions the capacity should be approximately equal at all lane merges, therefore the highest estimation can be considered the overall theoretical capacity. Perhaps most important is that this highest theoretical capacity has been estimated in Ballerup, where the traffic flow is regulated by signal-controlled intersection which resembles the principle of the proposed system of traffic lights aiming at improving the capacity of lane merges.

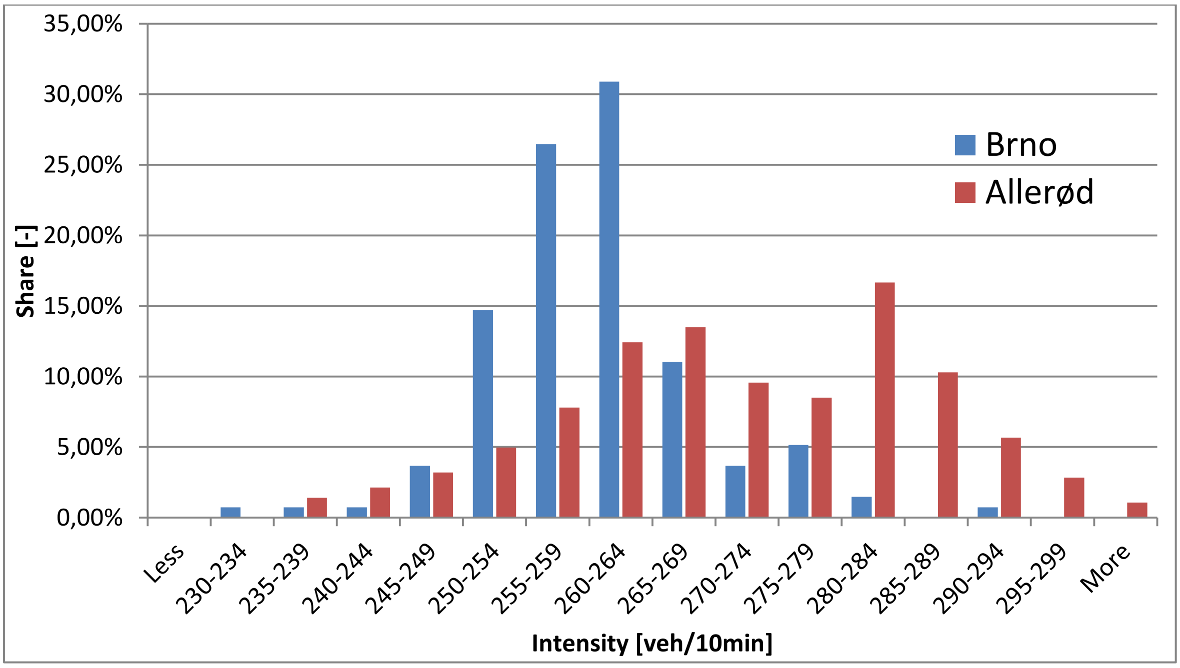


Figure 3.22 Comparison of intensities over ten-minute interval in Brno and Allerød

The measurements confirmed the capacity drop and the room to improve the lane merging capacity. Drivers’ behaviour was proven to have significant impact on the lane merging intensity or capacity. The measurement in Ballerup proved the potential of the proposed lane merging traffic lights system (introduced in the next chapter).

4 UTILIZING TRAFFIC LIGHTS AT LANE MERGING

There are several existing measures used across the world to improve the performance of lane merging which are reviewed in chapter 2.3. Traffic lights are not commonly used to treat lane merges at lane drops but they are used for on-ramp metering and freeway-to-freeway metering. Innovative system utilizing traffic light for removing the capacity drop at lane merging has been proposed earlier (Mikolášek, 2014) and other similar systems utilizing traffic lights has been proposed and tested in simulations (Lentzakis et al., 2008; Tympakianaki et al., 2014). There are even several similar existing systems at few specific locations, like 22-to-5 lane merge between a toll plaza and a major bridge (McCalden, 1984).

4.1 STATE-OF-THE-ART

Traffic lights are nowadays used in three basic situations. The first and most common situation is merging of on-ramp onto highway or any more important road where the vehicles on the on-ramp has to give way to the vehicles in the main traffic flow. On the other side stands the mainline metering, type of MTFC. It is only used on few important bottleneck locations. The last situation, which is borderline between the other two, is freeway-to-freeway metering. Borderline because it is still sort of on-ramp but the two connected roads are equal in their importance. As far as frequency of use is concerned, freeway-to-freeway metring stands in the middle. Due to the similarity with the other types of use it will not be discussed in detail.

4.1.1 On-ramp metering

On-ramp metering is a system used on ramps to reduce congestion on highways and therefore decrease travel times and increase safety by controlling the rate at which the vehicles from the ramp enter the highway.

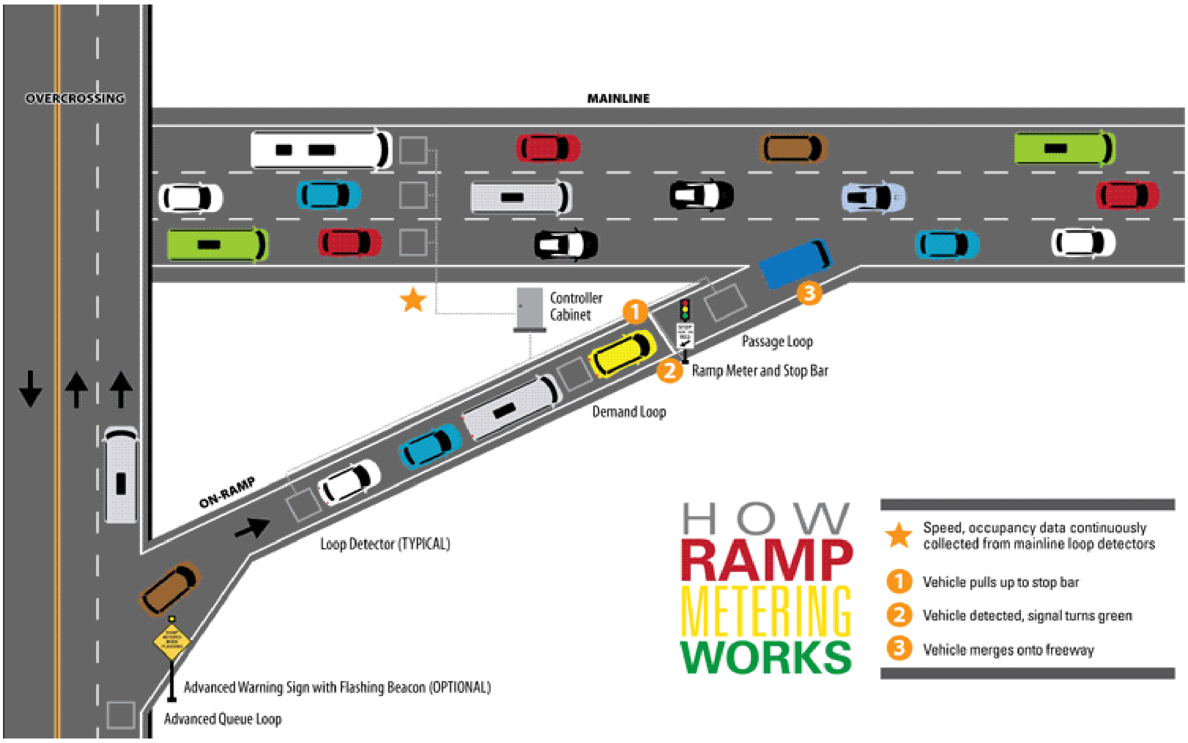


Figure 4.1 Typical configuration of ramp meter (Mizuta et al., 2014)

Figure 4.1 shows a typical configuration of ramp metering. The elementary part of each ramp meter is the traffic light (or traffic lights in case of multi-lane ramp) and the controller cabinet. The other parts are optional and their inclusion depends on what type of ramp metering is used on the given ramp. The optional parts are various detectors (usually built-in loop detectors). The detectors on the highway measure intensity, speed and occupancy (i.e. percentage of time a vehicle occupies the detector, another measure of density). The detectors on the ramp serve to detect waiting vehicles and queue length, and passage of vehicle after green light is turned on so that desired number of vehicles can be let on the highway per one green signal (usually one or two).

There are several types of ramp metering. The basic categorization is to pre-timed (fixed interval) and traffic responsive metering on one hand and to local and system-wide metering on the other hand. Pre-timed metering is much cheaper as it does not require detection but it cannot adapt to current conditions. Therefore it is not suitable for locations with (non-regularly) variable traffic. On the other hand the traffic responsive metering is more costly but yields better results thanks to its adaptability. Local metering is restricted to the single meter and cannot react to situation further in the surrounding area, whereas the system-wide metering provide opportunity to optimize the traffic in the whole area by coordinating the individual meters. It is very expensive to deploy but renders highest benefits.

Additionally, the ramp metering can be divided to single- or multi-lane and to single- or dual-release metering. For multi-lane metering the lanes should merge before entering the highway. Single- and dual-release metering is differentiated by the number of vehicles in the platoon allowed through per green signal. Some ramps may also provide bypass lanes for high occupancy vehicles or other priority vehicles.

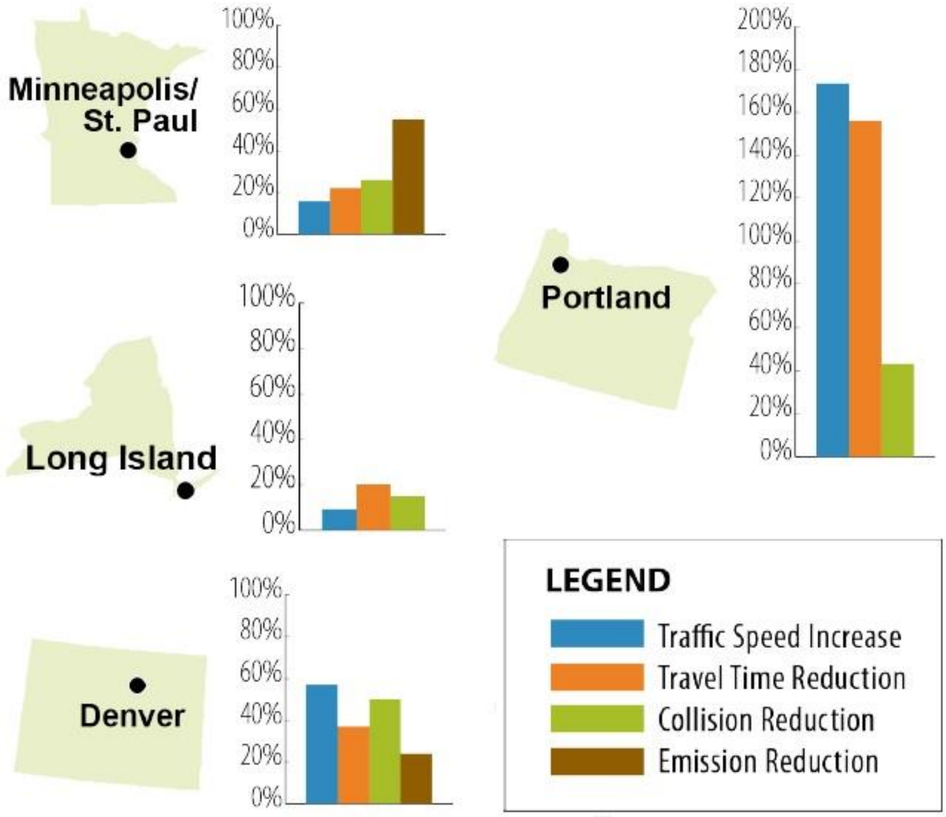


Figure 4.2 Estimated benefits of ramp metering (Mizuta et al., 2014)

Due to the reduced congestion on highways, the delays and travel times are reduced for most of the drivers even when the queueing on the ramp is included. It may not hold true for drivers only driving short distance on the highway. The metering breaks the platoons of vehicles trying to merge onto the highway and allows smoother merging of vehicles into the mainstream flow. Drivers on the highway are not forces to slow down and the risk of crashes is reduced. Eliminating the stop-and-go conditions and reducing the travel times is also considered associated with reduced fuel consumption, emissions, and environmental impacts overall, although it is difficult to measure and clearly connect to the metering.

Figure 4.2 shows estimated benefits of ramp metering in some metropolitan areas. Travel speed, travel time and collision rate improve usually by 10-30 % but in Portland the travel time and travel speed improved by about 150 % which is extreme improvement almost hard to believe. Cost/benefit ratio is claimed to be up to 1:15 as concluded in the Twin Cities Ramp Meter Evaluation (Mizuta et al., 2014).

Ramp length and geometry can be restrictive in use of ramp metering as there may not be sufficient visibility of the queue which increases the risk of collision (especially relevant for freeway-to-freeway metering due to the high travel speed). There may also be limited space for regulation due to short ramps as the queue should not block adjacent roads. Also public is often critical to expansion of ramp metering due to misunderstanding the concept and benefits. Public campaigns are advisable when deploying ramp metering or any similar traffic flow control system both to gain wider public support and to learn drivers how to behave.

4.1.2 Mainline metering

Mainline metering is a modification of the on-ramp metering where the mainstream flow is metered. As such, it is a type of MTFC. The point of MTFC is avoiding the capacity drop when the upstream intensity exceeds the capacity of a bottleneck and a queue forms. Examples of MTFC include VSL (see chapter 2.3.4), traffic lights (mainline metering), and in the future also vehicle-infrastructure interaction, IntelliDrive or similar systems (Carlson et al., 2011). The point of all the systems is to move the congestion upstream from the bottleneck and to 'batch' the vehicles in such a way that there are optimal traffic flow conditions at the bottleneck (e.g. lane merging) and the capacity drop is removed.

Mainline metering is still very limited nowadays and only few cases of mainline metering exist. However, the existing cases provide proof of concept as they significantly improved the throughput. After implementing mainline metering in San Francisco-Oakland Bay Bridge where 22 lanes at toll plaza merge into five lanes on the bridge the throughput increased by 15 percent (McCalden, 1984; Habioan, 1995). The first known application of traffic responsive mainline metering is at the tunnel under the Hudson River where it brought improvement of estimated nine percent (Gazis & Foote, 1969; Carlson et al., 2011). (Jacobson & Landsman, 1994; Haboian, 1995) provide further examples of application of mainline metering and also freeway-to-freeway metering.

There is an innovative control system for mainstream lane merges being researched and tested (Lentzakis et al., 2008; Papageorgiou et al., 2008; Tympakianaki et al., 2014). It is based on ALINEA metering control system (Papageorgiou et al., 1991). ALINEA is integral (I-type) feedback regulator for on-ramp metering. It regulates the rate at which vehicles are let go on the highway based on the occupancy. It is given by formula

$$q(k) = q(k-1) + K_R[\hat{o} - o(k-1)] \tag{4.1}$$

where k is discrete time step, *q(k)* is intensity to be implemented during the next period k, K_R is a regulator parameter (> 0), *o(k-1)* is the occupancy in the previous interval (%) and $\hat{o}$ is the desired occupancy for the downstream occupancy on the motorway. To maximize the capacity, it is set to critical occupancy o_{cr} (analogy to critical density) at which the intensity is maximal. Alternatively, the number of vehicles N can be used instead of the occupancy, as is utilized in the innovated proportional-integral (PI-type) regulator given by equation

$$q(k) = q(k-1) - K_P[N(k) - N(k-1)] + K_I[\hat{N} - N(k)] \qquad (4.2)$$

where K_P and K_I are regulator parameters and $\hat{N}$ is the desired number of downstream vehicles. Min and max values for *q(k)* may be set to avoid the wind-up phenomenon.

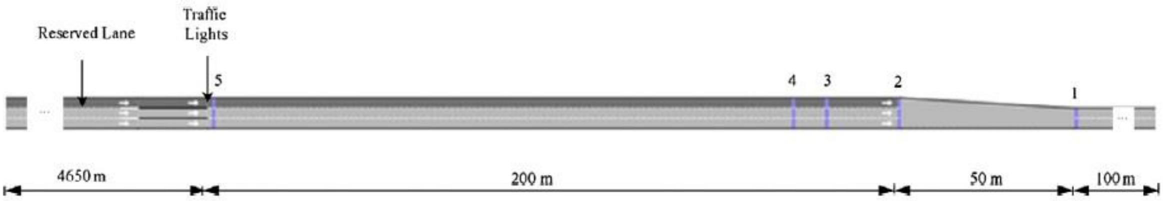


Figure 4.3 Hypothetical work zone used for simulation (Tympakianaki et al., 2014)

Different metering policies may be used to achieve the desired flow using the traffic lights, e.g. one-car-per-green, n-cars-per-green, full traffic cycle or discrete release rates. In the paper (Tympakianaki et al., 2014) the full cycle system is employed and tested in microsimulation on a 3-to-2 lane merge. They further tested adaptive fine tuning (AFT) procedure to optimize the regulator parameters which can be used also in field applications to improve performance of the control system based on real measurements. Both the regulator and AFT were found to be useful in removing the capacity drop and maximizing the throughput at lane merging.

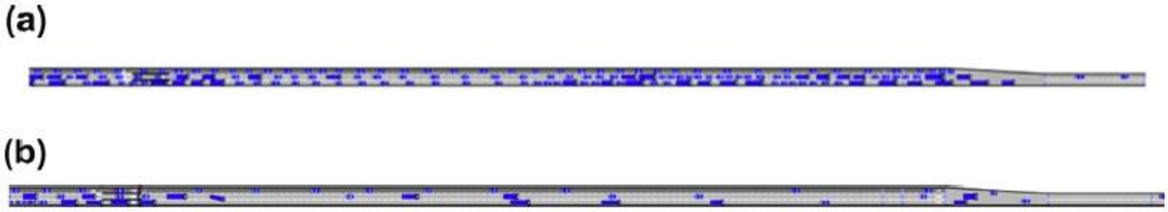


Figure 4.4 The simulated work zone with and without the tested merging traffic control (Tympakianaki et al., 2014)

The feedback regulator is based on the number of vehicles in the merging area. Detectors are placed before and after the merging area and based on the difference in the number of detected vehicles the number of vehicles in the merging area is calculated. The optimal value needs to me defined for each merging area based on parametrical simulation study. The distance of traffic lights from the merging area was also found to have influence on the capacity.

4.2 HOW DOES METERING INCREASE CAPACITY

Regardless of what kind of traffic flow control system is used (speaking of systems aiming at increasing intensity during congestion), the principle and aim is always the same – decreasing density during congestion. Figure 4.5 illustrates the principle of all the on-ramp and mainstream metering systems.

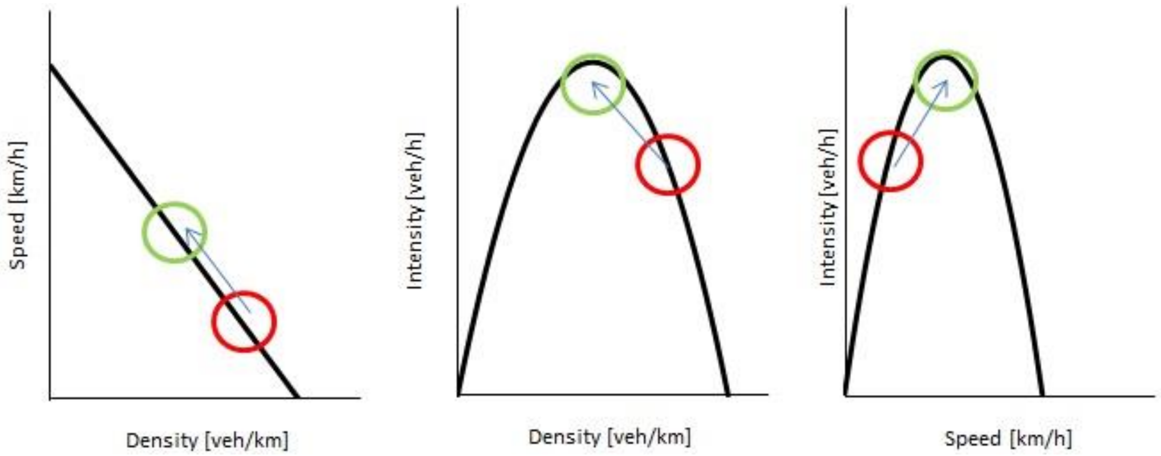


Figure 4.5 Simplified illustration based on Greenshield's fundamental diagrams of how traffic flow metering increases capacity. Red circles show non-controlled traffic flow during congestion, green circles the state of flow the control system aims to achieve.

During congestion the density exceeds the critical density at which the highest intensity is reached. The traffic flow control aims at decreasing the density near the critical density and thus maximizing the intensity/capacity.

Any measure that decreases density during congestion, in case of the lane merging at the merging area, can increase the road/merge capacity. In fact, if drivers give each other more space to merge, limiting the talk to lane merging, the intensity or capacity of the merging increases even without any external regulation system as the example of Allerød proves (see chapters 3.3.2 and 3.4). However, even in Allerød the density is still far higher than the critical density as drivers never intentionally leave gaps long enough to maintain the optimal density as they would perceive them as way to big. Therefore traffic flow control is necessary to achieve optimal performance of given infrastructure during congestion.

4.3 DESIGN OF ALTERNATIVE SYSTEM

Even though it turned out that the traffic control system proposed earlier (Mikolášek, 2014) is not as innovative as it was considered at the time, it is still something that has not been used or even tested, as far as the literature review revealed, in exactly the way it has been proposed. The proposed system is basically the same as double-lane fixed-interval ramp meter except the ramp meter is not calibrated to maximize the throughput on the ramp (or single-lane) but on the highway. However, in principle it is the same system and by proper setting of the interval the system can be calibrated to maximize the merge capacity. Another problem with the existing metering configurations is that they are generally non-transferable, excessively complicated and expensive. The proposed system only needs traffic lights with a controller and preferably two simple occupancy or speed detectors to detect queues.

To achieve optimal traffic flow at the merging area during congestion, the vehicles are metered shortly before the merging area using a separate traffic light for each lane. The traffic lights turn colours alternately in fixed intervals to let vehicles form one lane to pass while the vehicles in the other lane wait. Two different basic intervals of two and ten seconds are tested in this work. For the two-second interval, one vehicle is expected to pass per green signal. Shorter intervals were also tested and showed that possibly even shorter intervals may be feasible allowing higher intensity. The length of the interval needs to be calibrated in real-world conditions as the simulation is not able to reflect the behaviour of real vehicles and drivers sufficiently accurately. Also the length of the ten-second interval can be subject to changes. Ten seconds allow several vehicles from one lane to pass before the vehicles from the other lane are allowed to go and thus it eliminates the merging process altogether. On the other hand the interval is not too long as that could upset the drivers in the other lane and reduce compliance with the metering. Too long interval would also cause significant shortening of the queue in one lane and consequently unwanted lane changes.

Alternatively, in case of more advanced, and expensive, system the vehicles could be allowed to pass in platoons of one or more vehicles. This would require passage detector just like at ramp metering. However, as is explained below, some detectors for detecting queues are highly recommended anyway. The main advantage of the two second interval is fixed intensity at desired level (assumed the cycle-length could be adjusted by fractions of a second). On the other hand it may not be respected by the drivers or it may not be sufficiently long for some of the vehicles causing potential issues. There are several options about how to meter in field applications and the variability is based on number and types of detectors settled in the proximity of the merging area. Some of the options are presented in Figure 4.6.

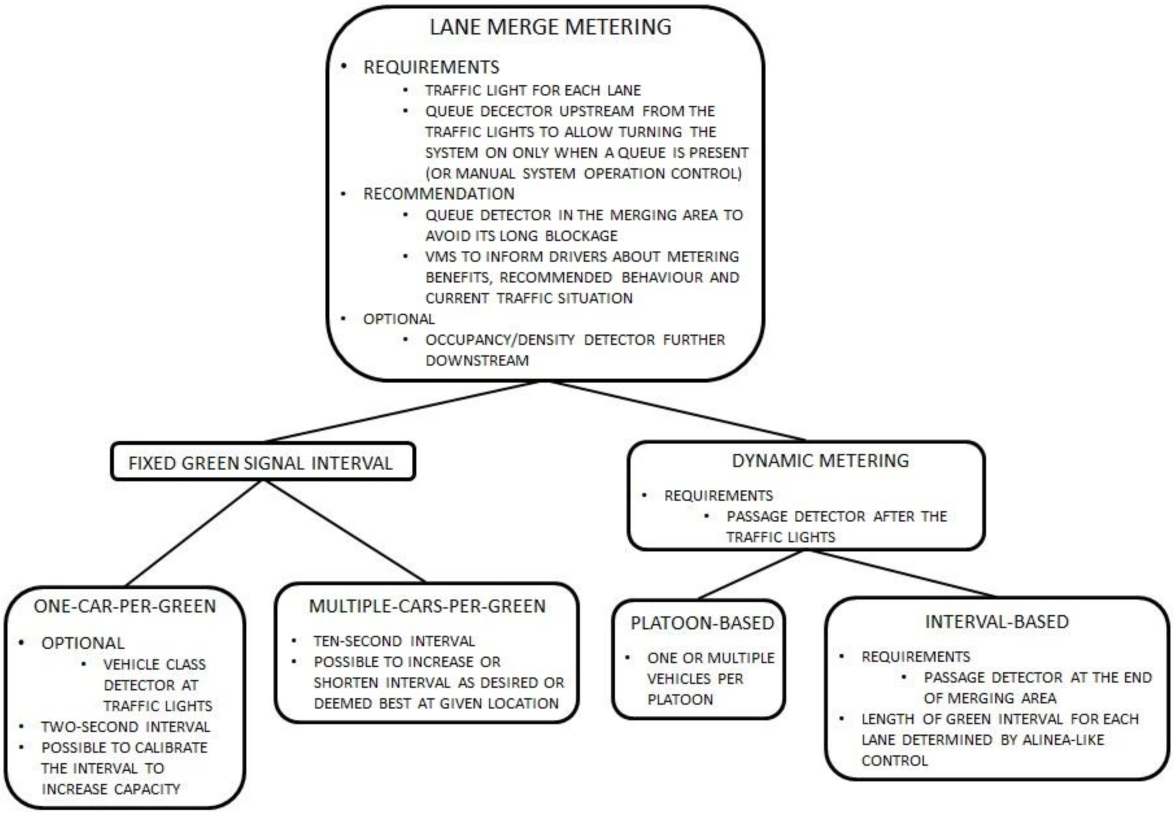


Figure 4.6 Overview of possible metering configurations at lane drop

The proposed system belongs to the left part of the classification tree while the traffic responsive metering is based the existing ramp metering systems and the Tympakianaki's ALINEA-like PI-type regulator. Even though the proposed system also becomes partially traffic responsive when the recommended queue detectors are implemented, the vehicle release rate (green time interval and cycle length) remains fixed.

Figure 4.7 shows a schema of the lane merge with positions of all the possible optional detectors. The purpose of the queue detector further upstream is to detect queue long enough so that the metering should be turned on. Given the locations where the proposed system is expected to be potentially used and its nature it is inadvisable to keep it running 24/7. Therefore a queue detection system is necessary to allow turning the system on when a queue forms and turn it off again once the queue disperses. In justifiable cases the queue detection might be substituted by manual control (directly in field or from distant dispatching/headquarters with use of video surveillance). To prevent turning the system on and off repeatedly in short period due to the queue length oscillating around the upstream queue detector, additional precautions are recommended (e.g. two detectors or time period over which the queue must be (not) detected).

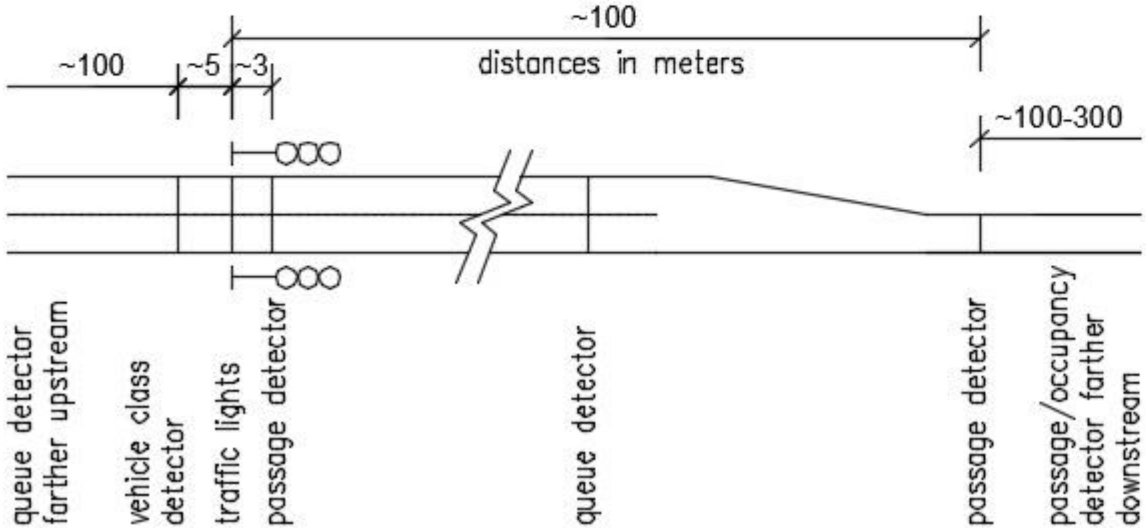


Figure 4.7 Schema of the metering system with all the optional detectors. The distances are only illustrative and need to be adjusted to local-specific conditions and/or further research. The two queue detectors are highly recommended for any configuration (except manual control).

The queue detector in the merging area serves to detect any possible congestion within the merging area while the metering system is operating since any queue in the merging area makes the metering useless and long-lasting queue would return the intensity to the uncontrolled level. While the metering should prevent forming of congestion in the merging area, the possibility cannot be excluded, e.g. due to trucks as happened regularly in simulations when the two-second interval was tested (see chapter 5.3) or progression of queue from the single-lane section. After careful calibration the queue detector may be excluded in some situations but the potential of the system could not be fully utilized (the system could not run at its capacity to prevent downstream congestion).

The rest of the detectors are optional and their need depends on desired functioning of the metering. The vehicle class detector just before the traffic lights serves to detect trucks and adjust the green signal length accordingly when fixed green-signal intervals are used (e.g. double the green time). Possibly, another car (not truck) from the adjacent lane might be allowed to the merging area as the low acceleration of the truck would allow it to pass the truck. However, this option would require further testing and it also may be desirable to leave longer gaps occasionally to prevent traffic flow breakdown farther downstream in the single-lane section.

Sensor technology	Count	Presence	Speed	Output data	Classification	Multiple lane, multiple detection zone data	Communication bandwidth	Sensor purchase cost[a] (each in 1999 U.S. $)
Inductive loop	✓	✓	✓ b	✓	✓ c		Low to moderate	Low[i] ($500–$800)
Magnetometer (two axis fluxgate)	✓	✓	✓ b	✓			Low	Moderate[i] ($900–$6,300)
Magnetic induction coil	✓	✓ d	✓ b	✓			Low	Low to moderate[i] ($385–$2,000)
Microwave radar	✓	✓ e	✓	✓ e	✓ e	✓ e	Moderate	Low to moderate ($700–$2,000)
Active infrared	✓	✓	✓ f	✓	✓	✓	Low to moderate	Moderate to high ($6,500–$3,300)
Passive infrared	✓	✓	✓ f	✓			Low to moderate	Low to moderate ($700–$1,200)
Ultrasonic	✓	✓		✓			Low	Low to moderate (Pulse model: $600–$1,900)
Acoustic array	✓	✓	✓	✓		✓ g	Low to moderate	Moderate ($3,100–$8,100)
Video image processor	✓	✓	✓	✓	✓	✓	Low to high[h]	Moderate to high ($5,000–$26,000)
	a Installation, maintenance, and repair costs must also be included to arrive at the true cost of a sensor solution as discussed in the text.							
	b Speed can be measured by using two sensors a known distance apart or estimated from one sensor, the effective detection zone and vehicle lengths.							
	c With specialized electronics unit containing embedded firmware that classifies vehicles.							
	d With special sensor layouts and signal processing software.							
	e With microwave radar sensors that transmit the proper waveform and have appropriate signal processing.							
	f With multidetection zone passive or active mode infrared sensors.							
	g With models that contain appropriate beamforming and signal processing.							
	h Depends on whether higher-bandwidth raw data, lower-bandwidth processed data, or video imagery is transmitted to the TMC.							
	i Includes underground sensor and local detector or receiver electronics. Electronics options are available to receive multiple sensor, multiple lane data.							

Figure 4.8 List of types of traffic flow sensors and their capabilities (Klein et al., 2006)

The passage detector after the traffic lights is needed for platoon-based metering. It could also be potentially used with the fixed intervals to extend the green signal when a truck would be occupying the detector. In that case vehicle class detector would be required (it could also be combined with the passage detector as the green signal length would be adjusted only after the vehicle would pass the lights).

For application of the Tympakianaki's metering control, both passage detectors are necessary to calculate the number of vehicles in the merging area. This system is expected to be less effective for two-to-one merges but might be useful for merges with higher lane count.

Finally, the occupancy detector further downstream makes it possible to control the traffic flow downstream for congestion and if necessary close both lanes temporarily before the flow downstream returns to stable condition. VMS are highly recommended in this case to inform the stopped drivers about the reasons why they are not allowed to continue. Only short complete closures are recommended, too. Alternatively, the rate at which the vehicles are allowed to pass can be decreased until the downstream situation gets better.

Since the system needs to be portable should it be usable in work zones, the detectors need to be portable, too. Hence, the conventional induction loops build-in in the pavement are not usable, generally speaking. However, any system able to detect queues that is somewhat portable can be used. To detect a queue a reliable (at low speeds) speed or occupancy detector is needed. FHWA provides and extensive overview of existing detectors with their capabilities in their handbook (Klein et al., 2006). Excerpt with the list of basic types and their capabilities is provided in Figure 4.8. As was said before, even manual control of the system based on wireless video transmission into some dispatching or headquarters or even directly in field could be suitable in some cases (e.g. short-term work zones).

Using (variable) message signs informing drivers about recommended behaviour, metering benefits and also about downstream queue ahead of the lane drop is advisable. As was said in chapter 4.1.1, some drivers do not fully understand the reason behind seemingly excessive traffic lights, do not see the benefits, and are critical of metering systems. Also they may not know how to behave as similar systems are not commonly used and there is even no ramp metering in the Czech Republic. Examples of recommendable signs include 'One vehicle per green', 'By abiding the metering you increase the throughput of the merge by xx percent', or any other relevant message. Different messages are suitable for different metering systems.

The proposed system was designed primarily for two-to-one lane merges but there is no reason why it should not be applicable on other types of merges with few simple modifications to the signal control. However, one possible issue is the need to raise a frame holding the traffic lights in case of three or more lanes as each lane needs a separate traffic light. Additionally, some types of detectors would not be usable or would have to be suitable adjusted. In cases where there is more than one downstream lane (after the merge), more lanes should be opened at a time in order to fully utilize the downstream capacity and not to make it starve for flow.

Combination of metering with VSL and VMS is recommended, especially at long-term work zones. The VSL and VMS increase safety and delay congestion and queues and the metering increases capacity once the flow collapses and a queue forms. That way the throughput and safety can be maximized at all times.

5 SIMULATION

Simulation is the imitation of the operation of a real-world process or system over time (Banks et al., 2001). In other words, simulation is trying to replicate events of real world in using mathematical methods. Usually, simulation is used to predict future events (e.g. weather forecast) or estimate effects of a certain change on a given system (e.g. traffic modelling).

5.1 INTRODUCTION TO SIMULATION

5.1.1 Types of traffic modelling

There are three different approaches to traffic modelling – macroscopic, mesoscopic and microscopic. The difference is in the detail of the simulation. Macroscopic simulation is the most aggregated, least detailed, which makes it very useful for large-scale simulations but also makes it the least precise (assuming the model is properly verified and validated). Microscopic models are analogy to hydrodynamics by regarding traffic flows as a particular fluid process whose state is characterized by aggregate macroscopic variables: density, volume, and speed (Barceló, 2010).

Microscopic models model behaviour of each individual vehicle in discrete time steps. They are most detailed and accurate if designed and validated properly but they are also most demanding in terms of computational performance. Therefore they are used mainly in smaller-scale simulations where high level of reality reflection is desired. That makes it perfect type of simulation to test the proposed metering systems.

Mesoscopic models are somewhere in between the two extremes. They model individual vehicles and its properties similar to microscopic models but the level of detail of behaviour and interaction is lower, e.g. lane change or turn is performed as a single action. One example of mesoscopic models are so called cellular automata (also used in other fields) which divide the network into discrete 'cells' which act as a chess board or as a grid-board in old-school RPGs. They are often used where higher level of detail is necessary but the model is too large for microscopic simulation.

5.1.2 Fundamentals of traffic simulation

Simulation is performed on a model that is mimicking real world. The point of simulation (or simulation experiment) is to eliminate real-world experiment where it would be impossible or too costly or to test different hypothesis before executing a real-world experiment. The key part of any simulation is creating a model that is able to sufficiently represent reality with desired level of accuracy so that the results of the simulation experiment are predicative.

The methodological framework of any simulation experiment is presented in Figure 5.1. As the real-world is too complex to be simulated in its full extent, the simulated model has to be simplified. The degree of simplification depends on type of simulation, questions to be answered, etc. First step in building a simulation model is creating a conceptual model. It is a simplified model with level of abstraction adequate to the research question and hypothesis. The conceptual model then needs to be translated into a computer model and verified (checked for errors). To calibrate and validate the model, collection of real-world data is needed. The process of calibration and validation is a process where the model parameters are adjusted so that the model is able to replicate the collected real-world data.

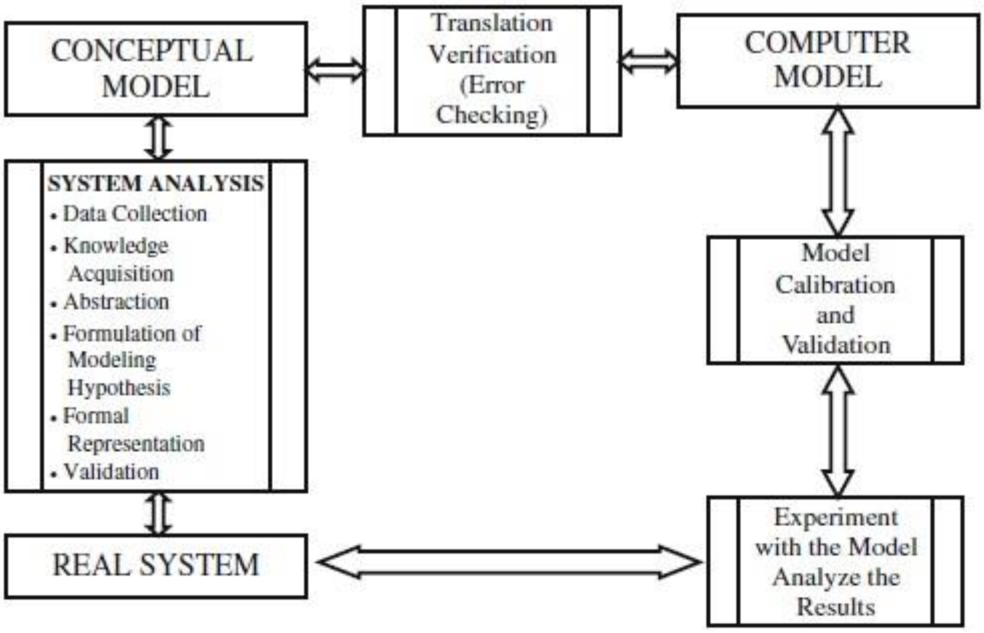


Figure 5.1 Methodological steps of the model-building process (Barceló, 2010)

The calibration process is very demanding as it is very dependent on the available data. Only the variables that have been recorded in the real system can be used in calibration. That makes really thorough calibration extremely expensive and time-consuming and thus it is rarely done in real-world applications. The extent of calibration is limited to the necessary extent appropriate for the desired accuracy. The most common variable used for calibration is intensity (and capacity). Further the relationships among the elementary or other (e.g. occupancy) variables can be used to further refine the model, for example speed-intensity or occupancy-intensity diagrams can be used.

The calibration and validation is an iterative process. The model parameters are set to certain values and validation test is run, i.e. the results of the simulation are compared to the real world data. If the simulated and real-world data fit sufficiently, which is given by formula 4.3, than the model is validated. *P* is probability, *d* tolerable difference and α is the level of significance.

$$P\{\,|'reality' - 'simulation'| \;\leq d\} > \alpha \qquad (4.3)$$

The validated model than can be used for experiments where the effects of considered measures or alterations to the existing system (traffic network) are estimated and later evaluated.

Simulations within the transportation field are usually based on stochastic distributions of the model parameters and event occurrence (e.g. vehicle entering the system). Therefore each simulation replication renders different results. That is secured by the so called random seed which needs to be different for each replication as the randomness of the parameters is based on this random seed. Replication of the simulation with identical random seed would lead to identical results. The simulation result is the average of the simulation replications. Sufficient number of replication is needed to get large enough sample so that the average is reliable, i.e. the variation of the sample is sufficiently small. The necessary accuracy depends on the desired level of confidence and on the desired minimal detectable difference between the base and experimental model. The necessary number of replications to achieve the desired accuracy can be estimated by the following formula (Larsen & Larsen, 2016):

$$n \cong Z^2_{1-\frac{\propto}{2}} * \frac{\sigma^2}{\delta^2} \tag{7.1}$$

where n is the estimate of necessary replications, α is the desired level of confidence, Z is the Z-score at the given level of confidence and σ and δ are the standard deviation of the initial sample and the desired standard deviation, respectively.

5.1.3 Measures of effectivness

To evaluate the effects of the changes to the system, set of tracked variables is chosen. Those variables are called the measures of effectiveness (MoE). The MoE are compared between the base validated model and the experimental models or scenarios. If the MoE improve compared to the base scenario, the tested measure can be evaluated as positive and conversely if the MoE worsen, the tested measure is probably not very effective in addressing the researched issue. The MoE need to be chosen appropriately for the tested measure, hypothesis or research question. Different MoE can be used within the same simulation model for different simulation experiments and multiple MoE are usually used. Delay, travel time, capacity, or the number of stops or lane changes are among the commonly used MoE. The total or average values are used as some drivers may be affected negatively by the measure but on average the measure is beneficial.

5.1.4 Aimsun simulation software

Aimsun in its current state provides a complex package of traffic simulation tools which include microscopic, mesoscopic and hybrid simulation with the four-stage model. The microsimulation module was used for the purpose of this work.

The flowchart of the microsimulation module depicted in Figure 5.3 provides insight into the functioning of the simulation process. Most important part is the ‘Update vehicle’ box, which represents the actual traffic flow modelling. It is based on the basic behavioural models – car-following model and lane-changing model. Those models also have optional extensions such as two-lane car following model which improves the vehicle behaviour on multiple-lane roads by taking into account the traffic in the adjacent lanes, or two-way overtaking model. The choice of model used in the current simulation step is given by the following condition

```
if (it is necessary to change lanes) then
        Apply Lane-Changing Model
endif
if (the vehicle has not changed lanes) then
        Apply Car-Following Model
endif
```

Figure 5.2 Condition for vehicle behaviour model choice (TSS – Transport Simulation Systems, 2014)

The car-following model is based on the Gipps model (Gipps, 1981 and 1986). The model parameters are not global but are determined by the influence of ‘local’ conditions, e.g. ‘type of driver’, road geometry or vehicles in surrounding lanes. The vehicle is trying to accelerate to its desired speed and is limited by the preceding vehicle. Details of the car-following model are provided in Aimsun Dynamic Simulator User Manual (TSS – Transport Simulation Systems, 2014).

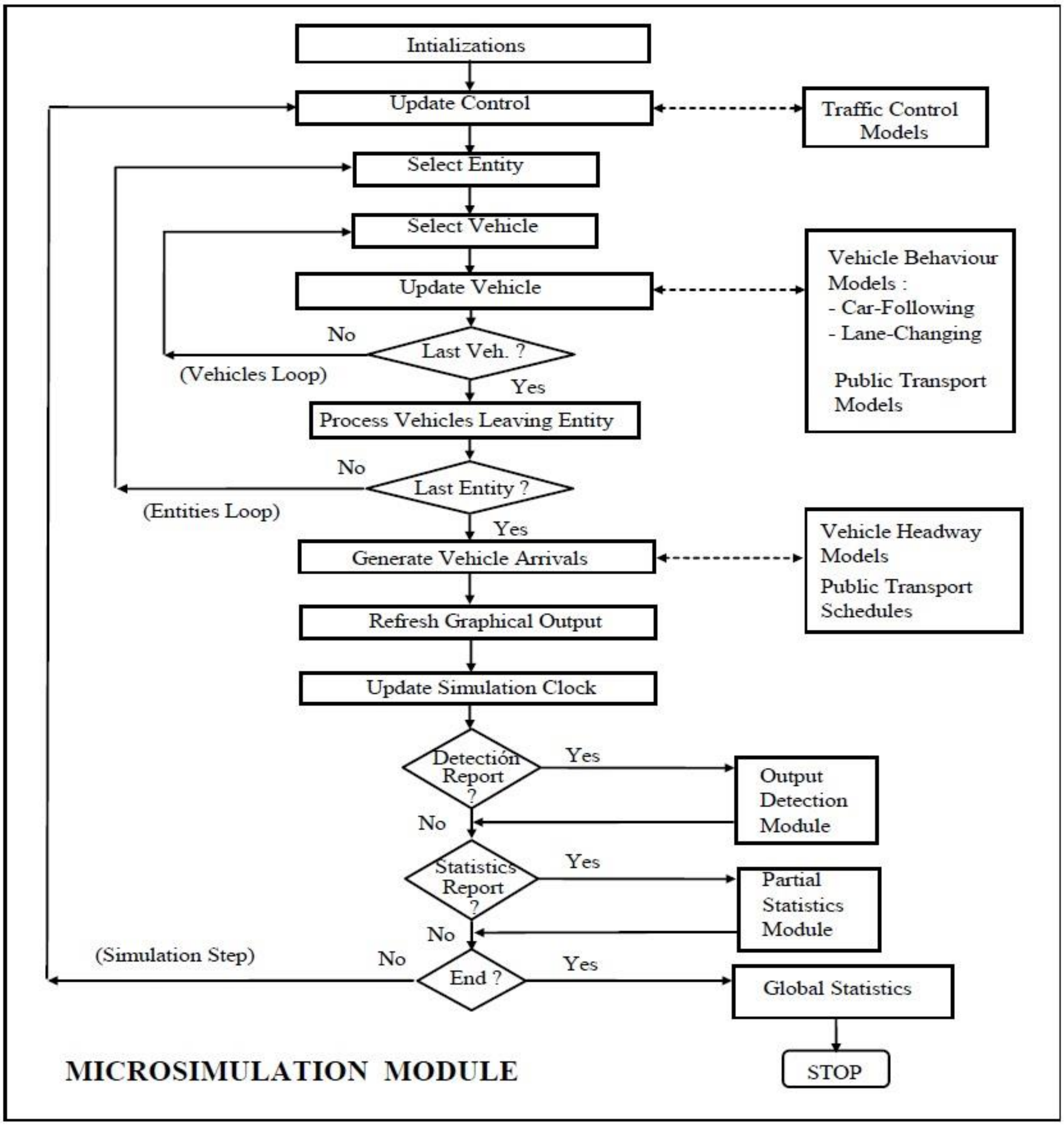


Figure 5.3 Aimsun microsimulation flowchart for traffic states based traffic demand (TSS – Transport Simulation Systems, 2014)

The lane-changing model is applied when the vehicle wants or needs to change lanes. Again, it is an extension of the Gipps lane-changing model. It distinguishes among three zones at which slightly different behaviour is applied – the closer the end of lane or intersection, the more forced the driver is to change lanes and behaves accordingly. The lane changing process consists of several elements: lane changing zone distance calculation, target lanes calculation, vehicle behaviour considering the target lanes, gap acceptance model, and target gap and cooperation. These sub-models aim at more realistic driver behaviour, e.g. seeking sufficient gaps in adjacent lanes, vehicle cooperation or acceptance of smaller gaps closer to the end of lane and with growing waiting time. More thorough explanation of the models is again provided in the Aimsun Dynamic Simulator User Manual (TSS – Transport Simulation Systems, 2014).

5.2 SIMULATION OF THE FIXED-INTERVAL MERGE METERING

5.2.1 Modelling the base scenario

The base scenario is used for validation and calibration of the simulated model and for comparison with the proposed metering system (with two different configurations). Hypothetical two-to-one lane merge is considered in this study. Three kilometres long entrance section with two lanes is used to accommodate any possible queues. The merging area is considered 50 m long. The last 200 m long segment of the road is single-lane. The total length of the modelled road is 3230 m. The original plan to base the model on the Allerød lane merge was rejected as the merging area geometrical properties appear not to play significant role in the simulation, at least within the tested range. The maximal allowed speed is 90 km/h on the whole modelled road as in Allerød and represents a typical highway work-zone lane merge.

Implementation in Aimsun

The simulation was performed in Aimsun v. 8.0.8 (R33472) using 'dynamic scenario' which is a microscopic simulation. The validated (see below) model comprises of two sections and a node. The modelled road begins with a two-lane three kilometres long section through which the vehicles enter the system. The last part is a single-lane 200 m long section. Both of these sections had their lane widths and shoulders based on the Allerød merge area, i.e. the lane width of 3.25 m, the left hard shoulder of 0.25 m and the right hard shoulder of 2.0 m. The width of the right hard shoulder is only estimated but according to the Aimsun manuals (TSS – Transport Simulation Systems, 2014) it is only decorative and has no effect on the vehicle behaviour anyway. The entrance two-lane and exit one-lane sections are connected with a node with a single 'turn' where the two lanes act as an entrance and the single-lane as an exit. There is a collision point at about half of the merging area where the vehicles are forced to merge. A detector was placed directly after the merging area to measure the traffic flow intensity and other variables.

Verification, calibration, and validation

One hour long simulation duration with 10-min warm-up period, enough to get stable congestion with higher traffic loads, was used for the calibration process. At the beginning, a shorter two-minute warm-up period was used in order to check the congestion-forming process, too. Four different demand scenarios (intensity of the vehicles entering the road) ranging from 1600 to 1900 car/h with 100 car/h increments were used to evaluate the ability of the model to replicate real-world traffic flow behaviour and performance. Also, a model with 1740 cars/h and 60 trucks/h was used to calibrate the capacity during congestion to the Allerød lane merging measurements.

Different demand scenarios and different parameter settings we used to verify and validate the model. Each setting was run 5 times with different random seed to get a moderate statistic while keeping reasonable computation time. In most cases the standard deviation was within acceptable range. In few cases the variance got a bit higher but it could be accounted to unstable conditions during the simulation (e.g. delayed queue formation caused by temporarily lower intensity at the beginning). The accuracy was deemed sufficient for the input uncertainties like model parameters and vehicle behaviour.

The base system was verified visually using the animation in Aimsun. The animation proved very useful for the purpose and was indeed necessary as at the beginning, the vehicles started merging several hundred meters before the merging. While they were merging relatively well in terms of the merging process itself, the merging was done well before the actual merging area which is very unrealistic and would not allow to reasonably test the proposed signal control systems.

Trial-and-error testing was performed in order to identify the cause of the peculiar behaviour. The initial implementation of the merging area in Aimsun turned out to cause the issues. After adopting a different solution (described above) to the design of the modelled system, the merging behaviour changed. The vehicles started to ride in their lanes until the merging area, where they merged very much like in real life during the free flow traffic state.

For the congested state the simulated vehicles were replicating the zipper merging process reasonably well, with usually 1-2 and occasionally 3 vehicles driving from one lane before a vehicle from the other line. Inclusion of trucks in the model worsened the situation given their lower acceleration – usually about five (but even up to 10) cars drove from the adjacent lane before the truck could merge. After calibration the number dropped significantly and usually only about 3 cars passed the truck before one let it merge. Given the less realistic merging behaviour with trucks and its unknown effect for the different modelled systems only cars were used for comparing the systems, primarily. To get an idea about the performance of the tested systems with trucks present, some simulations were also run with different shares of trucks.

Even extensive testing of various settings could not completely eliminate the slightly non-realistic merging behaviour due to the lack of a complex cooperation model within the simulation software. While there is an option determining willingness of drivers to cooperate when merging lanes, it did not appear to have strong effect on the merging behaviour. Parameters such as reaction time and vehicle acceleration were much more important when dealing with this issue. At the end of the day, the default vehicle parameters were used in the model (partially because the simulation software started crashing when the vehicle parameters window was opened).

While not entirely realistic, this merging behaviour was considered sufficiently good as it was not possible to achieve completely realistic behaviour and the model was able to replicate realistic capacity and capacity drop after calibration. Slight congestion waves were also visible in the queue, even though the model did not replicate the stop-and-go phenomenon very well and the vehicles in the queue were moving slowly all the time, except right at the merge when giving way. Nonetheless, waves of slightly lesser density could be identified in the queue.

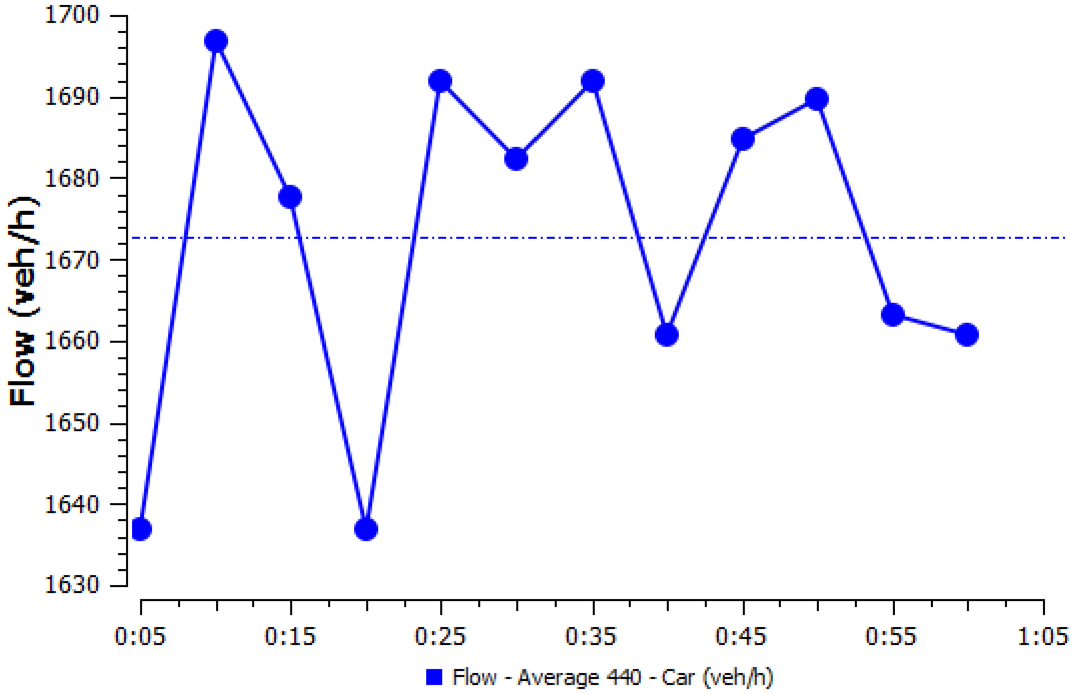


Figure 5.4 Progression of intensity (flow) over the course of the simulation of the calibrated base model for average demand of 1700 car/h

The model was able to satisfactory simulate the traffic flow behaviour near the capacity. For the demand scenario of 1600 car/h the traffic flow was quasi-stable and occasionally a short queue formed. Usually, after a certain period of time, when the immediate demand decreased due to its stochastic nature the queue dissolved again and this pattern would repeat over time. For the 1700 car/h demand the cars would merge fluently at the beginning but once a denser wave of cars arrived to the merge, the flow would collapse and a queue would form. Due to the constantly higher number of the incoming cars the queue would not dissolve, unless immediately after forming and only temporary, and over the course of the simulation the queue would grow in length, with some fluctuations. This behaviour is also in line with real world observations. For 1800 and 1900 car/h demand scenarios the queue would form even sooner and grow faster due to the higher density and demand.

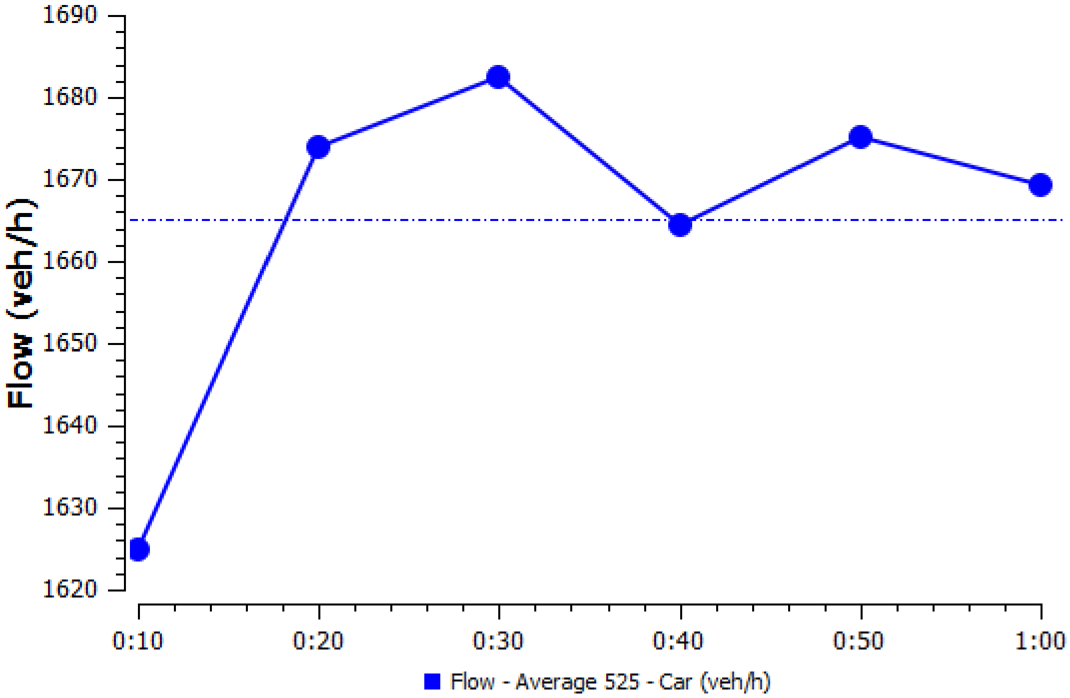


Figure 5.5 Progression of intensity (flow) over the course of the simulation of the calibrated base model for average demand of 1800 car/h

After the queue formation, the intensity dropped for all the demand scenarios due to the congestion-induced capacity drop (see chapter 2.2) to about 1670 car/h (obviously with the exception of the 1600 car/h scenario) which very well corresponds to the average congestion capacity measured in Allerød, i.e. 1630 veh/h, and fits reasonably well to the merging in Brno with 1561 veh/h if the higher share of trucks and the pedestrian crossings there are taken into consideration. For the scenario with trucks included (3.33-percent share based on the Allerød measurements) the average intensity dropped to 1634 veh/h which fits the average intensity during congestion in Allerød perfectly. This is the parameter by which the model was mainly calibrated, besides applying reasonable, realistic values and keeping the merging behaviour as realistic as possible. The figures 5.3, 5.4, and 5.5 show the history of average intensity over the course of the simulation for the base model with different traffic loads. The averages represent the merging capacity which is not dependant on the demand as there was always a queue present.

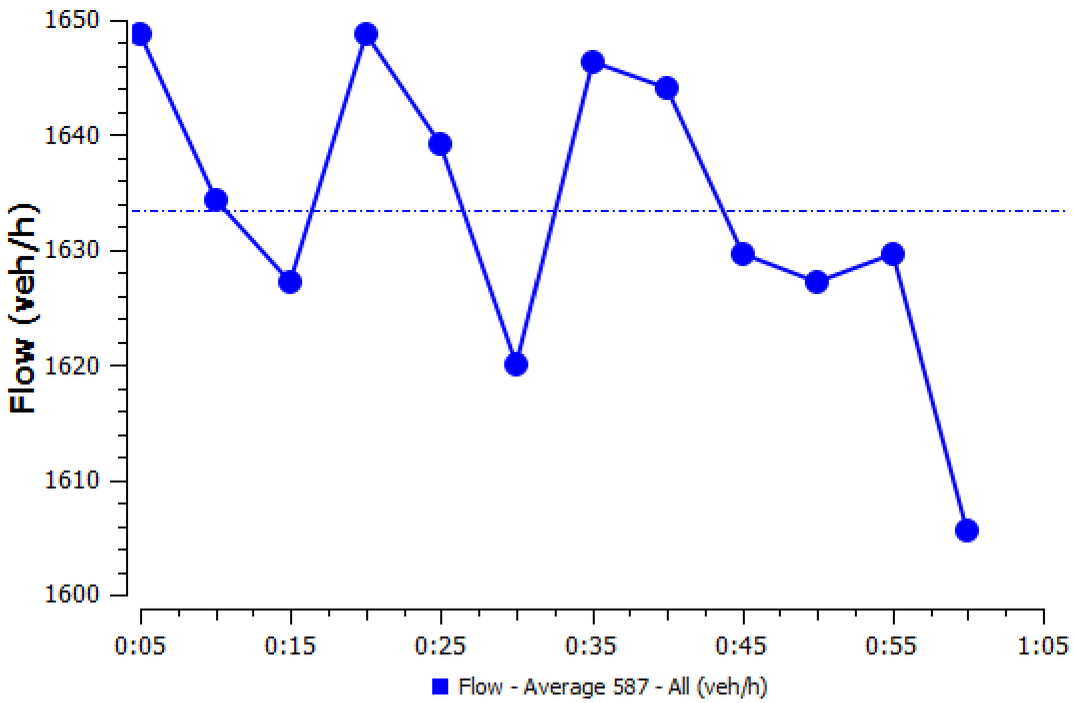
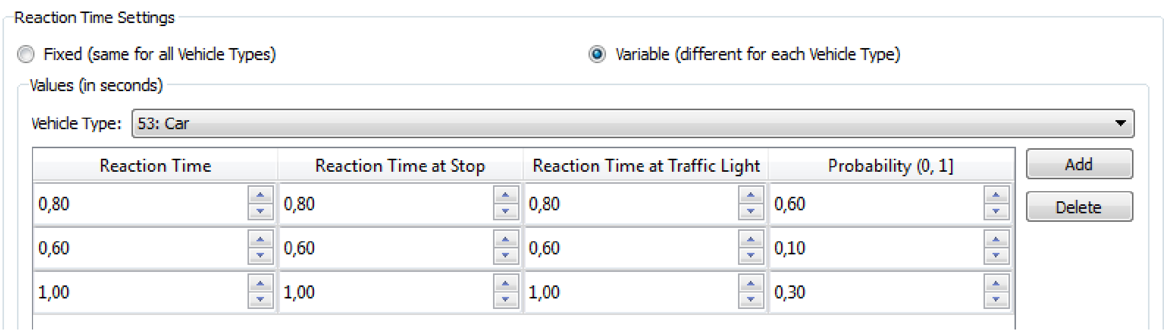


Figure 5.6 Progression of intensity (flow) over the course of the simulation of the calibrated base model for average demand of 1740 car/h and 60 trucks/h (1800 veh/h)

After trying various different settings, most of the parameters were left at the default values. However, some were changed more or less. Most notably the simulation step was set to 0.2 s and also the reaction times were changed and a stochastic distribution was applied to them. The reaction times for cars and trucks were considered identical. Aimsun also allows different reaction times at stops and traffic light but the regular pattern is expected to keep drivers alerted, hence the reaction times were kept identical. Further, drivers are expected to drive without actually stopping completely at the traffic lights if the two-second green interval is used. The used reaction times used in the calibrated model can be seen in Figure 5.7. The vehicle parameters were left at default values as was explained before.



Reaction Time	Reaction Time at Stop	Reaction Time at Traffic Light	Probability (0, 1]
0,80	0,80	0,80	0,60
0,60	0,60	0,60	0,10
1,00	1,00	1,00	0,30

Figure 5.7 Reaction times in the calibrated model

5.2.2 Modelling of the experimental models

The aim of the simulation was to compare two different lane merge metering systems (see 4.3) that are proposed as means to increase lane merging capacity with the base scenario or model, which represents an uncontrolled lane merge. One of the two systems uses fixed green-signal time interval of two seconds and the other one uses also fixed interval but ten seconds long. Additionally, the two-second interval was modified by 0.2 s to test if shorter intervals which would allow even higher capacity are feasible. The ‘two-second interval’ term can also encompass this possibility of the shorter intervals in some cases.

Set of tests was performed on the three models (base, two-second interval, and ten-second interval). First, all three models were loaded with constant traffic demand of 1800, 1900, an 2000 car/h. The ten-second interval model was further tested with 2100, 2200, and 2300 car/h to find its capacity. All the simulations were run for one hour.

To further illustrate the benefits of the proposed system with regard to queue length and merging capacity, a more sophisticated and realistic traffic demand scenario was used to load all three models. The demand grew every ten minutes in increments of 100 cars from 1700 car/h to 2000 car/h where it stayed for 20 minutes before it started to decrease again in similar manner to 1500 car/h and then less evenly down to 1000 car/h to allow faster dispersion of the created queues.

The simulation was also run with trucks included in the traffic flow to test performance of the metering systems in more realistic conditions despite the slightly less validity of the model when the trucks are included. About 3.5 % of the traffic flow was made up with trucks for each demand scenario in this test. The base and two-second interval model were loaded with 1800 and 2000 veh/h. The ten-second interval model was also loaded with 2200 veh/h.

Finally a simulation with ten-percent share of trucks in 2000 veh/h traffic flow was run to test how the performance scales with high share of trucks in the flow.

The metering systems were implemented in Aimsun by using metering modules immediately before the merging area. The metering type was set to ‘fixed’ in both cases. Figure 5.8 shows the setting of the signal control of the metering. Two-second yellow time was added to the ten-second metering cycle for safety reasons to allow arriving drivers stop since after ten seconds the arriving vehicles are already arriving at rather high speed.

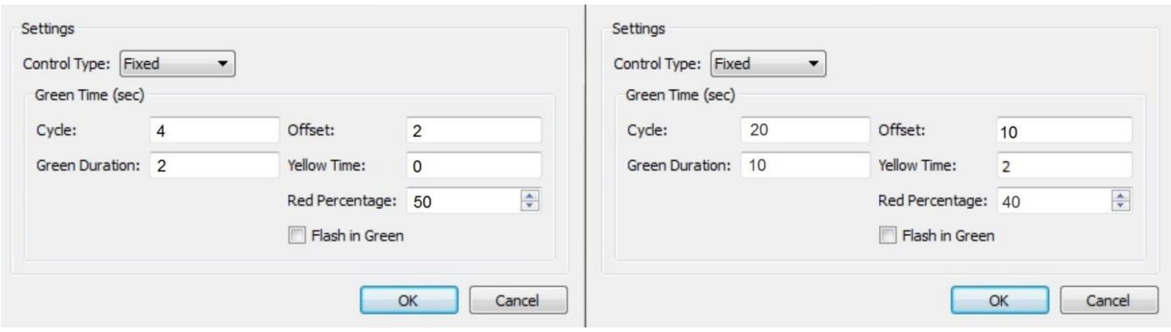


Figure 5.8 Setting of the signal control of the metering for the two- and ten-second interval models

Intensity, total travel time (TTT) and average delay (AD) were used as MoE. Intensity, or rather capacity, alone is not necessarily very important for drivers (and strictly speaking neither for the overall performance, at least directly). Or more precisely, they often do not perceive it as important because it does not directly affects them and they cannot measure it. Also achieving intensity 3000 veh/h is meaningless if the average speed is 10 km/h. Instead, the travel time and the delay are much more important variables both for drivers and the overall road network performance. The advantage of measuring delay is that it is independent on the trip length, unlike the travel time. Increased capacity during congestion decreases the queue length and consequently decreases delays and travel times but increased intensity may also lead to lower travel speeds due to higher density depending on traffic flow conditions. Therefore, controlling delay and travel time is important when evaluating any measures affecting traffic flow. Reduction of the queue length may also positively affect nearby roads as on- and off-ramps upstream do not get blocked by the jammed vehicles.

5.3 SIMULATION RESULTS

The first and most important task was comparing the three models using the chosen MoE and estimating capacity of each of the model. Results are presented in Table 5.1. The estimated capacities are circa 1675, 1775 and 2170 car/h for the base, two-second, and ten-second interval model, respectively. To estimate the capacity of the ten-second model, the tested traffic demand range had to be extended up to 2300 car/h since 2000 car/h did not cause any significant queues (only short and temporary when a dense wave of cars arrived). Results of the ten-second interval model are presented in Table 5.2.

Table 5.1 Estimates of MoE of the three tested models under different traffic loads without truck

Variable	Base model			two-second interval model			10-second interval model		
Demand [car/h]	1800	1900	2000	1800	1900	2000	1800	1900	2000
Intensity [car/h]	1678	1674	1679	1763	1779	1786	1786	1905	1987
Total travel time [h]	136,3	182,7	217	90,4	124,9	185,1	66,1	71,1	75,8
Average delay [s/km]	52,8	83,8	106,1	19,7	40,7	77,7	3,9	4,2	5,2

It can be seen in Table 5.1 and Table 5.2 that not only capacity but delays and travel times improved significantly for both tested metering configurations compared to the base scenario. In fact, the AD for the ten-second interval model with demand of 2300 car/h is lesser than for the base model with 1800 car/h (40.2 s/km vs. 52.8 s/km). Similarly, the TTT for base model with 1800 car/h is higher than the TTT for ten-minute interval model with 2200 car/h.

Table 5.2 Performance of the ten-second interval metering configuration (no trucks)

Variable	10-second interval model					
Demand [car/h]	1800	1900	2000	2100	2200	2300
Intensity [car/h]	1786	1905	1987	2083	2174	2170
Total travel time [h]	66,1	71,1	75,8	83,9	126,8	151,3
Average delay [s/km]	3,9	4,2	5,2	7,5	27,5	40,2

It can be said quite clearly that metering with the ten-second green-signal reaches better results than with the two-second interval. However, the matter of discussion is whether intensity exceeding 2100 car/h is desirable from the safety perspective. Strong argument for the positive answer to that question is the fact that by slight decrease of the average headway the congestions can be greatly reduced. Additionally, drivers seldom keep the commonly recommended two-second headway so the safety may not even be decreased at all (with lower intensity, there would be congestion waves with high density and low headways anyway). More important argument against such high intensities might be the capacity of the single-lane but the regulated flow at the entrance should be helpful in this regard. Furthermore, in reality there are trucks and other slow vehicles within the traffic flow which would naturally decrease the intensity and density.

In fact, alternating between the two modes may be utilized to keep the system operating constantly. In situation when the upstream flow would be between the capacity of an uncontrolled merge and the capacity of the ten-second interval metering, the queue would disperse and the system would turn of only for the queue to start growing again. By modifying the length of the green signal the downstream intensity could be modified in such a way that only short queue would be present so that the system does not turn off and on constantly and the traffic flow is kept as fluent as possible. Perhaps even more elegant solution would be inserting a short red signal into the cycle.

Some additional information and observations could be obtained from the simulations. For the demand of 1900 car/h the queue length for the base model was almost 1000 m, for the two-second interval model 600 m and there was no queue for the ten-second interval model at the end of the one-hour simulation in a randomly chosen simulation replication.

Due to the inability of replicating the queue detectors that would regulate the merging the metering was operation throughout the whole simulation. That caused that sometimes vehicles would pass the red light because they approached the metering at high speed as there was no queue ahead. This was an issue especially at the beginning of the simulation with lower demand at the two-second interval metering and at all times at the ten-second interval metering, although there the interval was long so only the first vehicle would pass the red and the others managed to stop. In field application this behaviour would be highly unwanted and potentially dangerous and would have to be accounted for by turning the system on and off with use of the queue detectors, or by decreasing speed in the approach section, or by any other effective measure. This behaviour was the likely cause for the higher standard deviation in some of the modelled scenarios.

In few rare cases a car did not drive through on the green light in the two-second interval model. It can be expected that real drivers would continue driving on the red light instead of breaking again if they did not manage to drive through on time. Therefore this should not cause issues in potential field applications.

Finally, a test with four- or close-to-four-second interval that was expected to allow two vehicles to pass the green light was performed. However, the different acceleration abilities of different vehicles are too big an obstacle and the metering could not be fully utilized of this issue should be addressed by increasing the interval length. Also the same or better results can be achieved with the two- or ten-second interval.

Fixed two-vehicle interval was found not to be effective as sometimes two vehicles made it through and sometimes only one and by increasing the interval further the capacity would be decreased. While a suitable interval could possibly be found, the same results can be achieved with the one-vehicle interval or with longer intervals. ALINEA-style or PI-type regulator would be likely more appropriate method if two-vehicle platoons were desired for some reason.

To test the models in more realistic conditions a variable demand scenario was created and applied to the three models. It also better illustrates the benefits of the metering more clearly. Figure 5.9 shows comparison of the progression of the intensities for the three models in a chosen replication which better illustrates the capacity drop and the progression overall. The same graph but with average intensity over the five replications and with the demand included are provided in Figure 6.1.

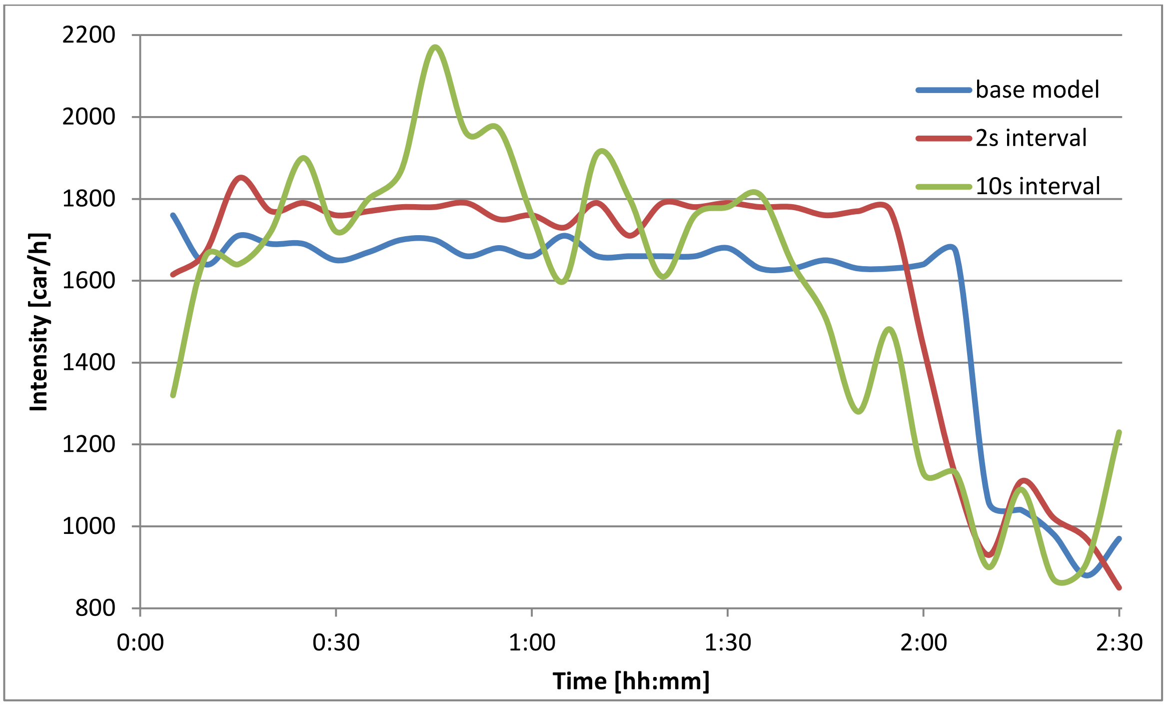


Figure 5.9 Course of intensity in chosen replications (one from each model)

This demand scenario was also used to compare average queue length. Detectors were plotted in 100-meter distances upstream from the merging area and the average speed at the detector was used to estimate the queue length at given time. Figure 5.10 shows progression of the average speed in a detector 500 m before the merging area in the base model. It is obvious that from 0:55 to 1:45 there was always a queue at least 500 m long in the base model. Using these graphs from all the detectors, the average queue length can be estimated. In some cases interpolation was used to estimate the queue length progression more realistically. The resulting graph of progression of the average queue length is shown in Figure 6.2.

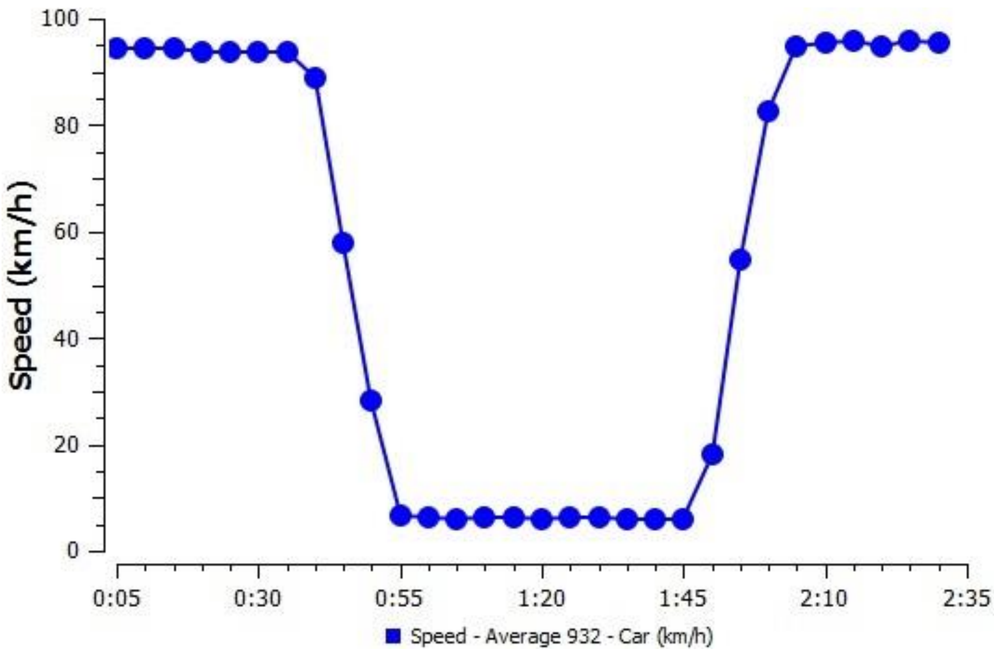


Figure 5.10 Course of average speed on a detector 500 m before the merging area in the base model

The last test without trucks included in the flow was aimed at possibility of shortening the two-second interval to achieve higher capacity. The simulation step was limiting the options of the green signal length. The simulations suggest that shorter than two-second interval might be feasible, at least at merges with low share of trucks. With the length of green light mere 1.4 s the average intensity reached 2568 car/h which is given by the fact that one car passed the green per one signal (3600/1.4=2571). The intensity for the other intervals was given by the same formula. For interval 1.2 s the metering stopped being functional as the cars did not have enough time to accelerate and would drive through on red light and a queue formed at the merging area.

Table 5.3 Estimated capacity of different fixed-intervals one-vehicle-per-green configurations

Variable	Modified two-second interval model					
green time length [s]	1,2	1,4	1,6	1,8	2,0	2,2
Capacity [car/h]	-	2568	2250	2000	1779	1573

The shorter intervals could be used in field applications to reach higher capacity close to, or even exceeding, the one of the ten-second interval configuration. Careful calibration would, however, be necessary as the simulation is not representative enough to draw conclusions on such detailed and fine differences.

Despite the slightly problematic truck behaviour in the merging area several test were run with trucks in the flow to test how the different metering configurations would fare with trucks which are always present in real world. Table 5.4 presents comparison of the models with circa 3.5-percent share of trucks included in the flow with two different levels of demand. The capacity of the base model remained at about 1620 veh/h to which it was calibrated. The two-second interval model fares slightly better with the estimated capacity of 1690 veh/h. The ten-second interval model proved yet again to be the best of the three with capacity about 1950 veh/h. However, shorter than two-second interval might bring improved results compared to the two-second interval.

Table 5.4 Performance of the three tested models with included ca. 3.5-percent share of trucks

Variable	Base model		Two-second interval model		Ten-second interval model		
Demand [veh/h]	1800	2000	1800	2000	1800	2000	2200
Intensity [veh/h]	1629	1614	1690	1686	1764	1951	1941
Total travel time [h]	145	245,6	120,3	226,9	67,4	105	220,3
Average delay [s/km]	61,3	131,3	41,5	111,9	5,1	22,4	88,4

As expected, the TTT and AD also improved significantly for the two metered models. For the demand of 2000 veh/h (1930 cars and 70 trucks), the AD was 22.4 s/km compared to 131.3 s/km, i.e. almost six times less. For the demand of 1800 veh/h the improvements are even bigger.

The capacity of the two-second interval model with trucks is significantly affected by queues in the merging area. The trucks often caused congestion in the merging area due to their low acceleration and lack of cooperation. If detectors were used to detect the queue and turn both lights to red until the queue disperses the capacity would increase. In field application the drivers could often cooperate better to avoid the traffic flow breakdown but running any type of metering system without any queue detection cannot be recommended as the benefits would be severely limited.

Table 5.5 Comparison of the three tested models with 10-percent share of trucks in 2000 veh/h flow

Variable	Base model	Two-second interval model	Ten-second interval model
Demand [veh/h] (10 % trucks)	2000	2000	2000
Intensity [veh/h]	1516	1546	1706
Total travel time [h]	296	274	229,9
Average delay [s/km]	178	159	111,7

The last of the performed simulations was with ten-percent share of trucks. It was only run with demand scenario of 2000 veh/h, i.e. 1800 cars plus 200 trucks. The efficiency of the proposed metering systems substantially decreased but still fared better than uncontrolled merge, especially the ten-second interval configuration. Given the problematic behaviour of trucks at merging and lack of cooperation, slightly better results can be expected in field applications (given drivers behave as expected and abide the traffic lights). Increasing the length of the green signal at roads with very high share of trucks, such as the D1 freeway with about 20-percent share of trucks (Celostátní sčítání dopravy, 2010), could be considered, including the drawbacks (longer waiting at the red light and higher speed of vehicle approaching the metering when the light turns red), as the results suggest that it might improve the merging capacity.

5.4 EVALUATION OF THE PROPOSED SYSTEM

The results of the simulations presented above prove the potential of the concept of the suggested metering system with fixed green signal interval. It appears that the ten-second interval would be a better choice compared to the two-second interval and possibly even longer interval could be used, especially on roads with high share of long and slow vehicles (trucks). However, if real-world experiments would prove the possibility to have shorter than two-second interval or with additional detectors installed before and after the traffic lights allowing more traffic responsive metering (see chapter 4.3 for examples of use of the detectors), the capacity of the one-vehicle-per-green metering might be significantly increased.

The presented results (values) cannot be taken too seriously as traffic simulation is not very precise and especially merging of traffic flows is its considerable weakness. Keeping that in mind, the results are so outright, that it can be said that the proposed metering system or systems are viable measure addressing the capacity drop at lane merges and should be further tested in simulation and also in field.

Even when trucks were included in the model and with high traffic demand the performance improvements compared to the base uncontrolled model were in the order of tens of percents when looking at delay reduction. The capacity increased by about 10 percent even in the worst scenario with ten-percent share of trucks. Total travel time is dependent on the length of the trip and therefore is not that well suited for such comparison.

6 CONCLUSIONS AND DISCUSSION

This thesis presents a brief introduction to the basics of traffic flow theory, behaviour, and modelling. The elementary traffic flow variables and their relations are explained using fundamental diagrams. The knowledge of traffic flow behaviour is then applied in explaining traffic flow behaviour at lane merging, queue formation and dissipation, and capacity drop. Different approaches to lane merging are presented. The last part of the theory deals with existing measures used at lane merges.

The practical part starts with description and analysis of intensity measurements in field. Three different lane merges were used for data collection. The data was analysed and compared among the locations. Several conclusions were drawn from the data analysis. The data confirmed the occurrence of capacity drop at lane merges when the capacity is reached. The different estimated capacities in Brno and Allerød suggest that different conditions may lead to different capacities and thus appropriate traffic flow control could lead to increased capacity and overall merging performance. The estimated theoretical capacities further confirm this idea. The drivers' behaviour, along with possibly the merging area geometry, was identified as the main cause of the merging performance. Obviously, circumstances like incline play significant role, too, but that is difficult to affect. Also the pedestrian crossing in Brno decreases the capacity and moreover is more often than not the trigger of a queue formation. The measurements in Ballerup where the merge is 'metered' by the preceding signal-controlled intersection further proved the hypothesis that by affecting the traffic flow the merging capacity can be improved.

The next chapter provides an overview of current use of traffic lights in traffic flow control. Ramp metering is proven by years of use as an efficient measure in dealing with highway congestions and there are also few existing examples of mainline metering at lane drops which improved the capacity by almost 10 percent. An innovative metering system for lane drops, also suitable for work zones, is introduced (including several possible configurations and extensions) and compared to other possible metering systems based on ALINEA metering control or platoon based metering. The proposed system is based on fixed-interval metering using traffic lights and two queue detectors with possible extensions aimed at improving the metering at the cost of higher price. The functioning of the metering is based on improving traffic flow state in the merging area by decreasing density and consequently speed, which improves intensity and removes the problematic acceleration.

The last part deals with simulation of the proposed metering system in two different configurations – two-second green light interval and ten-second green light interval. In both cases the green signal is alternating between the two merging lanes so that only one lane is opened at a time. The simulations provided solid proof of concept for the proposed systems. The improvements were different for the two configurations and strongly dependant on the demand and the share of trucks within the traffic flow but in all the simulated scenarios both the metering models performed significantly better than the base model with uncontrolled traffic flow. The capacity improved by 2-57 percent depending on demand scenario and type of metering. The improvements for field application are estimated to about 10-20 percent. The improved capacity leads to great decrease of delays and travel times. The delays were close to zero in some of the simulation scenarios.

Figure 6.1 illustrates the benefits of the two metering configurations in a typical rush hour situation. In the base model, once the flow collapsed, the downstream intensity dropped to the congestion capacity level and stayed there until the queue dispersed long after the actual demand (i.e. upstream intensity) decreased below the capacity. The two-second interval model performed much better, increasing the capacity from ca. 1670 car/h to ca. 1770 car/h (the traffic flow was made up purely from cars in this scenario due to the less realistic truck merging behaviour). Therefore, the intensity increased and the queue dispersed much sooner than in the base model and also grew significantly shorter. Comparison of the progression of the queue length over the course of the simulation is provided in Figure 6.2. The third model with the ten-second green light interval performed even better. No sustained queue formed as the estimated capacity is exceeding the highest demand used in this scenario. Only short temporary queues formed when a wave of vehicles with high local density arrived to the merge. For most of the time all the vehicles waiting for the green signal in their lane would be able to drive through during the ten seconds of the green signal.

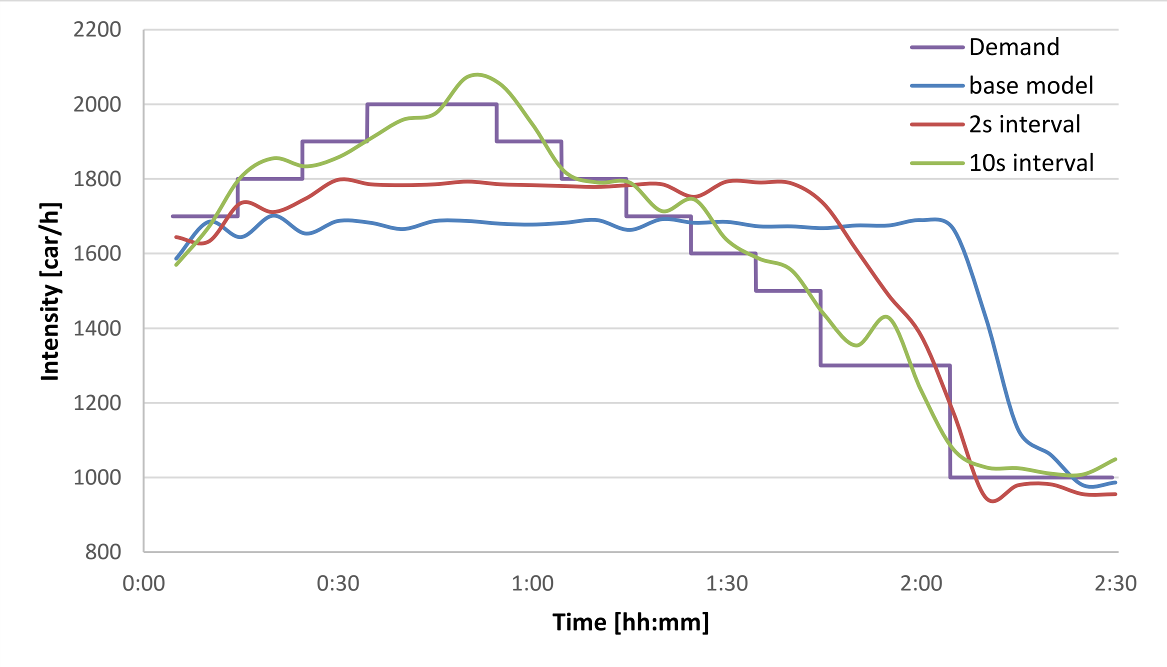


Figure 6.1 Comparison of the course of the average intensities in the three models under identical load

It has to be kept in mind that the metering system is planned to behave differently, i.e. it should only be activated once a queue of a certain length is present, because non-stop operation would be problematic due to the constant changes of red and green signals in the two lanes. Therefore, the queues could not be entirely eliminated but could be kept as short as possible. Adding a red signal of appropriate length into the signal control cycle is a possible method of controlling the downstream intensity. The aim is to keep the queue until the upstream flow decreases below the capacity of the uncontrolled merge. If the queue dispersed before that due to too high efficiency of the metering, the queue would start growing again and the traffic flow would have to be stopped completely until the merging area cleared again. Therefore, the two queue detectors are necessary to control the operation of the system, regardless of which configuration is used. Additional queue detectors may serve to adapt the desired metering capacity according to current conditions and other detectors may be used for further modifications of the metering. One detector may serve several purposes in some cases.

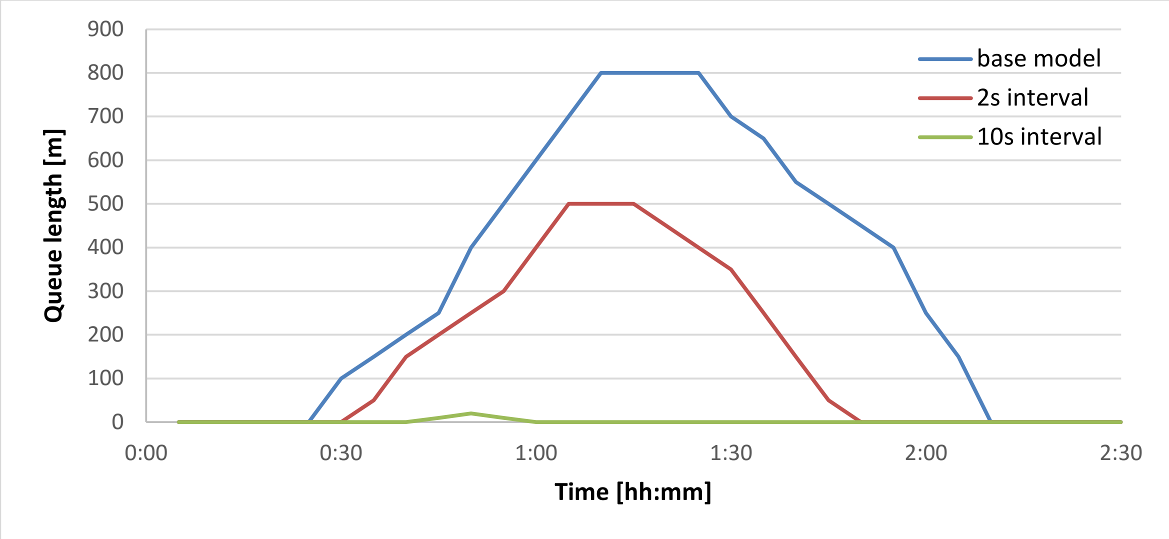


Figure 6.2 Comparison of est. average queue length in the three tested models under identical load

The results of the measurements, review of the existing metering applications and the performed simulations all strongly support the idea of the mainstream metering at lane drops in order to increase the capacity. Even small increase in the capacity may bring significant decrease of delays, travel times and all the other drawbacks of congestions and queues as Figure 6.2 illustrates. Based on the performed simulations, the ten-second interval appears to be a better metering configuration as it always reached higher capacity compared to the two-second interval. However, when the two-second interval was shortened, the capacity increased even above the capacity of the ten-second interval metering. What is the actual shortest feasible interval in field application is however impossible to predict and it may be possible that even the two-second interval would be too short. On the other hand, increasing the length of the ten-second interval might also bring further improvements.

This thesis is a solid prove of concept for the proposed metering system or in fact any kind of metering at lane drops. The advantage of the proposed concept is that it needs only two queue detectors in its minimalistic form and additional detectors are only optional. With proper types of detectors the system is portable and could be used at work zones, too.

Further research and experiments are needed to further develop the concept and find what is the best type of metering for field applications. Field experiments are necessary to fine-tune the configuration and to identify any possible issues due to the limitations of traffic simulation which fail at unstable traffic flow conditions, of which lane merging is a typical example. Some of the possible issues are presented in the thesis but whether they would appear in real-world conditions is unknown and new issues may emerge. Wide public campaign and application of variable message signs is strongly advised when applying any kind of metering in field since the road users are not used to it and may not see the potential benefits of yet another traffic lights blocking their ride. It is expected that the lane merge metering could bring great decrease of delays and travel times for the road users when properly calibrated and utilized.

REFERENCES

AUSTRALIA. 1999. *Act 1999.219 Australian Road rules reg. 148,149*.

BANKS, Jerry, John S. CARSON II, Barry L. NELSON and David M. NICOL. 2001. *Discrete-event system simulation*. 3rd ed. Upper Saddle River, NJ: Prentice Hall.

BARCELÓ, Jaume (ed.). 2010. *Fundamentals of traffic simulation* [online]. New York: Springer, xvii, 440 p. Available at: http://link.springer.com.proxy.findit.dtu.dk/book/10.1007/978-1-4419-6142-6/page/1

CARLSON, Rodrigo Castelan, Ionnais PAPAMICHAIL and Markos PAPAGEORGIOU. 2011. Local Feedback-Based Mainstream Traffic Flow Control on Motorways Using Variable Speed Limits. *IEEE Transactions on Intelligent Transportation Systems* [online]. **12**(4), 1261 - 1276. Available at: http://ieeexplore.ieee.org.proxy.findit.dtu.dk/document/5873154/

Celostátní sčítání dopravy 2010. 2010. *Ředitelství silnic a dálnic* [online]. Praha: Road and Motorway Directorate. Available at: scitani2010.rsd.cz

CHANG, Gang-Len and Zuojin ZHU. 2006. A Macroscopic Traffic Model for Highway Work Zones: Formulations and Numerical Results. *Journal of Advanced Transport* [online]. **40**(3), 265-287. Available at: http://onlinelibrary.wiley.com.proxy.findit.dtu.dk/doi/10.1002/atr.5670400303/abstract

CHUNG, Koohong, Jittachai RUDJANAKANOKNAD and Michael J. CASSIDY. 2007. Realtion between traffic density and capacity drop at three freeway bottlenecks. *Transportation Research Part B* [online]. **41**, 82-95. Available at: http://production.datastore.cvt.dk/filestore?oid=539a7cd712734ed51f04f865&targetid=539a7cd712734ed51f04f867

CZECH REPUBLIC. *Zákon 361/2000 Sb. o provozu na pozemních komunikacích (act on road rules)*.

GARCIA-CASTRO, Alvaro and Andres MONZON. 2013. Variable Speed Limits. Review and development of an aggregate indicator based on floating car data. In: *13th International Conference "Reliability and Statistics in Transportation and Communication"* [online]. Riga, p. 117-127. Available at: http://oa.upm.es/26687/1/INVE_MEM_2013_164713.pdf

GAZIS, Denos C. and Robert S. FOOTE. 1969. Surveillance and Control of Tunnel Traffic by an On-Line Digital Computer. *Transportation Science* [online]. **3**(3), 255-275. Available at: http://dx.doi.org/10.1287/trsc.3.3.255

GIPPS, P. G. 1981. A behavioural car-following model for computer simulation. *Transportation Research Part B: Methodological* [online]. **15**(2), 105-111. Available at: http://dx.doi.org/10.1016/0191-2615(81)90037-0

GIPPS, P. G. 1986. Multsim: a model for simulating vehicular traffic on multi-lane arterial roads. *Mathematics and Computers in Simulation* [online]. **28**(4), 291-295. Available at: http://dx.doi.org/10.1016/0378-4754(86)90050-9

Google Maps. 2016. *Google* [online]. Available at: maps.google.com

GRILLO, Lia, Tapan DATTA and Catherine HARTNER. 2008. Dynamic Late Lane Merge System at Freeway Construction Work Zones. *Transportation Research Record: Journal of the Transportation Research Board* [online]. (2055), 3-10. Available at: http://trrjournalonline.trb.org.proxy.findit.dtu.dk/doi/abs/10.3141/2055-01

GREENSHIELDS, Bruce D., J. R. BIBBINS, W. S. CHANNING and H. H. MILLER. 1935. A study of highway capacity. *Highway research board proceedings*.

HABOIAN, Kevin A. 1995. A Case for Freeway Mainline Metering. *Transportation Research Records* [online]. **1949**, 11-20. Available at: http://onlinepubs.trb.org/Onlinepubs/trr/1995/1494/1494-002.pdf

HALLMARK, Shauna and Nicole ONEYEAR. 2015. *Modelling Merging Behavior at Lane Drops* [online]. Iowa: Smart Work Zone Deployment Initiative. Available at: http://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=1107&context=intrans_reports

HOLCNER, Petr. 2012. *Modelování a simulace dopravního proudu*. Brno. Habilitační práce. Brno University of Technology, Faculty of Civil Engineering.

HRANAC, Robert, Emily STERZIN, Daniel KRECHMER, Hesham RAKHA and Mohamadraza FAR-ZANEH. 2006. *Empirical Studies on Traffic Flow in Inclement Weather* [online]. Washington D.C.: Federal Highway Administration. Available at: http://ops.fhwa.dot.gov/publications/weatherempirical/weatherempirical.pdf

JACOBSON, Eldon L. and Jackie LANDSMAN. 1994. Case Studies of U.S. Freeway-to-Freeway Ramp and Mainline Metering and Suggested Policies for Washington State. *Transportation Research Record* [online]. **1446**, 48-55. Available at: http://onlinepubs.trb.org/Onlinepubs/trr/1994/1446/1446.pdf#page=54

KLEIN, Lawrence A., Milton K. MILLS and David R. P. GIBSON. 2006. *Traffic Detector Handbook: Third Edition - Volume I* [online]. McLean: FHWA. Available at: http://www.fhwa.dot.gov/publications/research/operations/its/06108/06108.pdf

KRISTENSEN, Loke Bager. 2016. *Video Processing for Traffic Data Collection and Analysis*. Kongns Lyngby. Master's Thesis. Technical University of Denmark. Supervisors Thomas Martini Jørgensen and Thomas Kjær Rasmussen.

LARSEN, Rune and Allan LARSEN. 2016. *13400 Simulation in Transport and Logistics: Output Modelling and Analysis (presentation)*. Kongens Lyngby: DTU Transport.

LENTZAKIS, Antonis F., Anastasia D. SPILIOPOULOU, Ioannis PAPAMICHAIL and Yibing WANG. 2007. Real-time work zone management for throughput maximization. In: *2008 Transportation Research Broard 87th Annual Meeting* [online]. Available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.631.7902

MCCALDEN, M. S. Jr. 1984. A traffic management system for the San Francisco-Oakland bay bridge. *ITE Journal* [online]. **54**(5), 46-51. Available at: http://ecommerce.ite.org/iMIS/iCommerce/Sign_In2.aspx?ssoredirect=http://library.ite.org/pub/e23e6e3a-2354-d714-51ef-aaebe92c3994

MIKOLÁŠEK, Igor. 2014. *Analýza dějů při spojování dopravních proudů*. Brno. Bachelor Thesis. Brno University of Technology, Faculty of Civil Engineering, Institute of Road Structures. Supervisor Petr Holcner.

MIZUTA, Arianne, Kim ROBERTS, Les JACOBSEN and Nick THOMPSON. 2014. *Ramp Metering: A Proven, Cost-Effective Operational Strategy* [online]. Washington: FHWA. Available at: http://ops.fhwa.dot.gov/publications/fhwahop14020/fhwahop14020.pdf

OLSTAM, Johan Janson and Andreas TAPANI. 2004. *Comparison of Car-following models* [online]. Linköping: Swedish National Road and Transport Research Institute. ISSN 0347-6049. Available at: http://www.ctr.kth.se/publications/ctr2004_07.pdf

PAPAGEORGIOU, M., I. PAPAMICHAIL, A. D. SPILIOPOULOU and A. F. LENTZAKIS. 2008. Real-time merging traffic control with applications to toll plaza and work zone management. *Transportation Research Part C* [online]. (16), 535-553. Available at: http://www.sciencedirect.com/science/article/pii/S0968090X07000873

PAPAGEORGIOU, Markos, Habib HADJ-SALEM and Jean-Marc BLOSSEVILLE. 1991. ALINEA: A Local Feedback Control Law for On-Ramp Metering. *Transportation Research Record* [online]. **1320**, 58-64. Available at: http://onlinepubs.trb.org/Onlinepubs/trr/1991/1320/1320-008.pdf

PEI, Yulong and Leilei DAI. 2007. Study on Intelligent Lane Merge Control System for Freeway Work Zones. In: *2007 IEEE Intelligent Transportation Systems Conference* [online]. IEEE, p. 586-591. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4357778

RADWAN, Essam, Zaier ZAIDI and Rami HARB. 2011. Operational Evaluation of Dynamic Lane Merging In Work Zones with Variable Speed Limits. *Procedia Social and Behavioral Sciences* [online]. **16**, 460-469. Available at: http://production.datastore.cvt.dk/filestore?oid=546e304d4bc8944e3d046c66&targetid=546e304d4bc8944e3d046c6a

TSS - TRANSPORT SIMULATION SYSTEMS. 2014. *Aimsun 8 Dynamic Simulators User's Manual*.

TSS - TRANSPORT SIMULATION SYSTEMS. 2014. *Aimsun 8 User's Manual*.

TYMPAKIANAKI, A., A. SPILIOPOULOU, A. KOUVELAS, I. PAPAMICHAIL, M. PAPAGEORGIOU and Y. WANG. 2014. Real-time merging traffic control for throughput maximization at motorway work zones. *Transportation Research Part C: Emerging Technologies* [online]. (44), 242-252. Available at: http://www.sciencedirect.com.proxy.findit.dtu.dk/science/article/pii/S0968090X1400103XJ

LIST OF TABLES

LIST OF FIGURES

LIST OF USED ABBREVIATIONS AND SYMBOLS

PT – Public Transport
VSL – Variable Speed Limit
ALINEA – Asservissement linéaire d'entrée autoroutière (a local feedback ramp-metering strategy)
MTFC – Mainstream Traffic Flow Control
ITS – Intelligent Transport System
GHR - Gazis-Herman-Rotery traffic flow model
ŘSD – Ředitelství Silnic a Dálnic (Road and Motorway Directorate)
AADT – Annual Average Daily Traffic
DLM – Dynamic Late Merge
AFT – Adaptive Fine-Tuning
FHWA – Federal Highway Administration
MoE – Measures of Effectiveness
TTT – Total Travel Time
AD – Average Delay

q – intensity
v – speed
ρ – density
v_{max} – maximal speed in Greenshield's model
ρ_{max} – maximal density in Greenshield's model
P – probability
d – tolerable difference between simulation output and collected data
α – level of confidence
n – number of replications
$Z_{1-\alpha/2}$ – Z-score at α-level of confidence for two-tailed estimation
σ – standard deviation of sample
δ – desired standard deviation
k – discrete-time index
q(k) – target entering flow in the next interval *k* in ALINEA or ALINEA-line metering
q(k-1) – intensity in the previous interval
o(k-1) – occupancy in the previous interval
K_R – regulator parameter of ALINEA regulator
ô – target occupancy
N(k) – current number of vehicles in the merging area
N(k-1) – number of vehicles in the merging area prior the last interval
$\widehat{N}$ – desired number of vehicles in the merging area
K_P – proportional regulator parameter of the PI-type regulator
K_I – integral regulator parameter of the PI-type regulator