Cooperative Ramp Metering for Urban Motorways Based on Machine Learning

Gregurić, Martin

Doctoral thesis / Disertacija

2018

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Transport and Traffic Sciences / Sveučilište u Zagrebu, Fakultet prometnih znanosti**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:119:512923

Rights / Prava: In copyright / Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-07-19**

Repository / Repozitorij:

Faculty of Transport and Traffic Sciences - Institutional Repository

DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJI

University of Zagreb

Faculty of Transport and Traffic Sciences

Martin Gregurić

COOPERATIVE RAMP METERING FOR URBAN MOTORWAYS BASED ON MACHINE LEARNING

DOCTORAL DISSERTATION

Supervisors:

Sadko Mandžuka, PhD, Full Professor

Edouard Ivanjko, PhD, Associated Professor

Zagreb, 2018

Sveučilište u Zagrebu

Fakultet prometnih znanosti

Martin Gregurić

KOOPERATIVNO UPRAVLJANJE PRILJEVNIM TOKOVIMA NA URBANIM AUTOCESTAMA ZASNOVNO NA STROJNOM UČENJU

DOKTORSKI RAD

Mentori:

prof. dr. sc. Sadko Mandžuka

izv. prof. dr. sc. Edouard Ivanjko

Zagreb, 2018

Biographies of Supervisors

Sadko Mandžuka PhD is a professor at the Division of Intelligent Transportation System, Faculty of Traffic and Transport Sciences, University of Zagreb. He had the opportunity to work both in academic and industrial environments including Brodarski Institute Ltd, consulting in the Innovation Area for SME's, etc. Prof. Mandžuka is a president of ITS-Croatia, Collaborating member of Croatian Academy of Engineering, Founding member of Croatian Robotic Association, and a member of Technical Committee on Marine Systems (Coordinating Committee on Transportation and Vehicle Systems – IFAC (International Federation of Automatic Control), IEEE (Institute of Electrical and Electronics Engineers), IEEE Intelligent Transportation Systems Society etc. Prof. Mandžuka is author of more than 100 internationally reviewed publications. He is the editor of the Engineering Applications Section of the international scientific journal An International Journal of Control and Optimization: Theories & Applications (IJOCTA). Also, he is a member of the editorial board of international scientific journal International Journal of Intelligent Transportation Systems Research (Springer) and the national journal Ceste i mostovi (Roads and Bridges). Prof. Mandžuka was a project leader and collaborating member of several EU and national Research & Development projects.

Edouard Ivanjko PhD defended his PhD thesis in 2009 on the Faculty of electrical engineering and computing University of Zagreb. Since 2011 he is with the Department of intelligent transportation systems of the Faculty of transport and traffic sciences, University of Zagreb. His research topics of interests include application of artificial intelligence for road traffic control, autonomous vehicles, and estimation and forecasting of road traffic parameters. As a researcher he was or is active on the following projects: FP7 „Intelligent Cooperative Sensing for improved traffic Efficiency" ICSI, IPA IIIc „Computer Vision Innovations for Safe Traffic" VISTA and „System for route optimization in a dynamic transport environment" SORDITO financed from EU structural funds and several MZOS projects. He is or was also a local management committee member for the Republic of Croatia for EU COST actions "Towards autonomic road transport support systems" TU1102 and "High-Performance Modelling and Simulation for Big Data Applications (cHiPSet)" IC1406. He is the local research group leader on the Faculty of transport and traffic sciences for the „Scientific centre of excellence for data science and cooperative systems". Until now he published three chapters in books, 7 scientific papers in international journals and 37 scientific papers on international conferences.

Acknowledgment

I would like to thank my supervisors prof. Sadko Mandžuka and prof. Edouard Ivanjko for guidance and support. This thesis would not have been completed without their support and guidance. Many thanks go to my colleagues at the Department of Intelligent transportation systems and prof. Vanja Ličina. I am deeply grateful to my family, for their remarkable and unconditional support. I am extremely lucky to have them in my life.

Abstract

To cope with today's urban motorway congestions and the inability to increase motorway capacity in urban environments requires the implementation of advanced control methods. These methods are an integral part of Intelligent Transportation Systems (ITS). An ITS essentially integrates information and communication technology to solve the congestion problems. Ramp metering (RM) and Variable Speed Limit Control (VSLC) are some of the most widely used urban motorway traffic control methods. RM provide direct influence over the on-ramp flows by using specialized traffic lights, while the VSLC control speed of mainstream flow by using variable messaging signs. A dedicated algorithm for RM or VSLC uses sensory data form an urban motorway to compute actions that will have a positive impact on both types of traffic flow. This study will focus on the cooperation of an RM and a VSLC systems, and the integration of several different RM algorithms into a single algorithm called INTEGRA. The algorithm is created by using the Adaptive Neuro-fuzzy Inference System (ANFIS) as an instance of machine learning techniques. Furthermore, INTGERA is expanded in order to integrate its original functionality with a recurrent neural network for traffic demand prediction. As the final step, this doctoral thesis will provide evaluation of different criteria for learning dataset functional setup, based on which ANFIS neural network of INTEGRA will be learned. Results of all mentioned approaches will be compared and discussed in relation with other commonly used urban motorway control methods.

Key words

Cooperative control, ramp metering, variable speed limit control, urban motorways, adaptive neuro fuzzy inference systems, recurrent neural network, learning dataset

Prošireni sažetak

Kako bi se ostvario veći stupanj uslužnosti na urbanim autocestama primjenjuju se nove upravljačke metode. Najkorištenije upravljačke metode na urbanim autocestama su upravljanje priljevnim tokovima (engl. ramp metering - RM) i promjenjivo ograničenje brzina vozila. Algoritam za upravljanje priljevnim tokovima ima zadaću računati stupanj propuštanja vozila s ulazne rampe (priljevni tok) u glavni tok u odnosu na ukupan broj vozila koja žele uči u glavni tok, pri tome koristeći ulaznu rampu kao privremeni „spremnik" vozila. Čitava RM metoda upravljanja prometa na ubranoj autocesti zasniva se na prometnim podatcima koji se prikupljaju u stvarnom vremenu posredstvom prometnih senzora (induktivnih petlji, kamera, itd.), te proslijeđeni RM algoritmu. Stupanj propuštanja vozila u glavni tok proračunat od strane RM algoritma prosljeđuje se specijalnoj upravljivoj prometnoj signalizaciji [4].

Glavnina istraživanja u ovom doktorskom radu vezana je upravo za upravljanje priljevnim tokovima s posebnim naglaskom na kooperaciju s drugim sustavima upravljanja prometom, te primjeni strojnog učenja. Također, u kooperaciji s upravljanjem priljevnih tokova razmatrat će se druge upravljačke metode kao što su sustav zabrane prometovanja određenim prometnim trakama, te potpuno ili djelomično upravljanje vozilima opremljenim posebnim računalnim jedinicama. Od strane autora predložen je neuro-neizraziti okvir za učenje koji omogućuje integraciju različitih strategija upravljanja priljevnim tokovima. CTMSIM makro-simulacijski alat koji je izrađen u Matlab programskom okruženju korišten je u simulaciji odabranih metoda upravljanja prometom na urbanim autocestama. Simulator je proširen od strana autora kako bi podržao kooperativno upravljanje priljevnim tokovima, kao i sustav za promjenjivo ograničenje brzina vozila.

Cilj istraživanja: Razviti strategije za upravljanje urbanom autocestom koje će biti zasnovane na konceptu kooperacije te ih evaluirati u relaciji s postojećim samostalnim upravljačkim strategijama. Dodatni cilj je izrada novog okvira za učenje različitih strategija upravljanja priljevnim tokovima.

Hipoteze: Upravljanje priljevnim tokovima zasnovano na strojnom učenju, u slučajevima značajnih promjena prometne potražnje, može ostvariti kraće vrijeme putovanja uz prihvatljivi red reda čekanja na ulaznim rampama te prihvatljivog ukupnog kašnjenja u odnosu na dosadašnje strategije upravljanja priljevnim tokovima. Algoritam koji je zasnovan na kooperaciji između više ulaznih rampi i kooperaciji upravljanja priljevnim tokovima i promjenjivog ograničenja brzina vozila, može ostvariti značajno bolje rezultate u odnosu na samostalne (pojedinačne) aplikacije upravljanja priljevnim tokovima i promjenjivog ograničenja brzina.

Prometne mreže velikih urbanih središta imaju glavnu zadaću opsluživati prometnu potražnju bližih manjih gradova, većih središnjih gradova, te tranzitni promet. Kako bi se rasteretila urbana mreža grada od tranzitnog prometa izgrađene su posebne autoceste pod nazivom - gradske obilaznice, a smještene su na rubnim dijelovima urbanih područja. Urbane obilaznice su ubrzo postale okružene urbanom infrastrukturom s proširenjem urbanih područja. Spomenuto je uzrokovalo njihovu integraciju s urbanom prometnom mrežom. Urbane obilaznice su postale poznate kao urbane autoceste. Glavne značajke urbanih autocesta su:

1. Upitna mogućnost za fizičkim povećanjem postojećih prometnih kapaciteta;
2. Veći broj ulaznih i izlaznih rampi koje su blizu jedna drugoj;
3. Opslužuju tranzitni promet i promet sa svojim izvorom u urbanoj prometnoj mreži.

Konstruktivne značajke prometnica i prometna potražnja urbane autoceste, posljedično uzrokuju njihovo preopterećenje. S obzirom na prostorni i vremenski aspekt, zagušenja su česta u dijelovima urbane autoceste blizu ulaznih i izlaznih rampi tijekom ranih jutarnjih i kasnih popodnevnih sati [1]. Spomenuti tipovi zagušenja su poznati kao vršni sati, te čine vremenski okvir periodičkih zagušenja. Dnevne migracije, prema i od radnog mjesta, mjesta edukacije, itd. su uzrok periodičnih zagušenja tijekom vršnih sati samo ako su ona dovoljno intenzivna i vremenski preklopljena. U prostornom kontekstu, periodička zagušenja su najčešća na dijelovima urbane autoceste s gusto raspoređenim ulaznim i izlaznim rampama. Intenzivniji prometni tok koji se na ulaznim rampama spaja s glavnim tokom može znatno usporiti brzinu i povećati prometnu gustoću glavnog toka. Spomenuta situacija dovodi do značajnog zagušenja niz glavni tok. Nadalje, ako je veći broj vozila dovoljno dugo zaustavljen na ulaznoj rampi posljedično će se stvorit dugi red čekanja na ulaznoj rampi [2]. Red na ulaznoj rampi može narasti do takvih razmjera da u konačnici blokira prometni tok na dijelu urbane mreže za koji je vezana ulazna rampa [3]. Ne-periodična zagušenja uzrokovana su naglim smanjenjem prometne propusnosti na određenom dijelu urbane autoceste kao posljedica incidenata ili događaja od velike važnosti za javnost. Prometni scenariji koji sadrže prethodno identificirane kritične tipove zagušenja biti će implementirani u model urbane autoceste. Temeljne značajke modela autoceste biti će napravljene prema parametrima Zagrebačke obilaznice, kao primjera urbane autoceste. Spomenuta obilaznica biti će korištena za evaluaciju odabranih sustava upravljanja prometom na urbanim autocestama.

Izradit će se analiza koja će identificirati postojeće upravljačke metode unutar upravljačkih sustava autoceste sa mogućnošću smanjenja utjecaja kritičnog zagušenja na protočnost autoceste. Naglasak će biti postavljen na algoritme za upravljanje priljevnim tokovim. Analiza će također uključivati upravljačke strategije kao što su promjenjivo ograničenje brzina vozila (engl. Variable speed limit control – VSLC), sustav zabrane prometovanja određenim prometnim trakama, te posebni upravljački uređaji u vozilima. Temeljem ove analize bit će identificirani temeljni nedostaci u slučajevima samostalne implementacije. Istraživanje će uključivati analizu prednosti kooperativnih sustava u odnosu na samostalne upravljačke implementacije. Motivacija za omogućavanje kooperativnog ponašanja između upravljačkih entiteta na autocesti je temeljena potrebom za ostvarivanjem specifičnih ciljeva koji su važniji od svrha individualnih entiteta. Spomenuto implicira kako kooperacija može podrazumijevati hijerarhijsku strukturu odlučivanja [5].

Prvo će se uspostaviti i analizirati kooperacija između rampi kao izvorištu priljevnih tokova, a zatim i mogućnosti kooperacije između različitih metoda upravljanja na urbanoj autocesti. Kooperacije između RM-a i VSLC-a biti će okosnica razmatranja kooperacije između različitih metoda upravljanja prometom na urbanoj autocesti. Razlog spomenutom odabiru je utjecaj spomenute kooperacije na sve tipove prometnih tokova na urbanoj autocesti. Kooperacija između VSLC-a i RM-a omogućuje smanjenje brzine vozila koja prilaze prometno opterećenoj ulaznoj rampi kako bi se reducirao efekt šok valova i omogućio veći priljev vozila sa zagušene rampe [6] [7]. U svrhu provedbe kooperacije RM-a i VSLC-a izvest će se nadogradnja CTMSIM makro-simulacijskog alata, kako bi se omogućila primjena kooperativnog RM-a i VSLC-a

Također, doktorski rad će na konceptualnoj razini razmatrati druge oblike kooperacije kao što je sustav zabrane prometovanja određenim prometnim trakama i VSLC-a, te kooperacije između VSLC-a ili RM-a s vozilima opremljenim upravljačkom jedinicom (engl. On-Board-Unit – OBU). Kooperacija s vozilima opremljenim OBU-om pruža dodatne mogućnosti kao što su: potpuno automatsko vođenje vozila, polu-automatsko vođenje vozila ili pružanje informacija vozaču.

Ključni znanstveni doprinos ostvarit će se razvojem RM algoritma s mogućnošću integracije različitih RM upravljačkih strategija. Spomenuti RM algoritam biti će zasnovan na naprednom okviru za strojno učenje različitih strategija upravljanja priljevnim tokovima. Spomenuti okvir zasnivat će se na Adaptive Neuro-Fuzzy Inference System (ANFIS) algoritmu, te će imati mogućnost pružanja odgovarajućih odgovora na značajne promjene u prometnoj potražnji na urbanoj autocesti. Skup podataka za učenje biti će stvoren pomoću znanja o upravljanju koje će biti prikupljeno od strane RM algoritma s različitim upravljačkim strategijama, budući da se kreće od pretpostavke da svaka RM upravljačka strategija daje bolja rješenja za odgovarajući prometni scenarij. Koristiti će se sljedeća tri RM algoritma s različitim upravljačkim strategijama: lokalnom (ALINEA), prediktivno-nadmetajučom (SWARM), te kooperativnom (HELPER). Spomenuti RM algoritmi nazivati će se RM algoritmi učitelji, a njihovo upravljačko znanje biti će prikupljeno provođenjem niza simulacija u istom simulacijskom okruženju. Korištenjem kriterijske funkcije odabiru se odgovarajuća rješenja (dobivene od strane RM algoritama učitelja) koja će biti će uvrštene u konačni skup podataka za učenje. Ovisno o strukturi rješenja uvrštenih u skup podatka za učenje ovisit će ostvareni rezultati INTEGRA algoritma prema pojedinim mjerama uslužnosti autoceste. S obzirom na važnost kriterijske funkcije u radu INTEGRA algoritma provest će se analiza postavki kriterijske funkcije kako bi se ostvario optimalan odnos njenih parametara u svrhu dobivanja boljih vrijednosti mjera uslužnosti. Istraživanje postavki kriterijske funkcije pokazalo je da se mogu postići bolja rješenja s kombinacijom parametra koji daju veću težinu ukupnom kašnjenju na autocesti u odnosu na vrijeme putovanja glavnim tokom.

INTGERA algoritam bit će proširen cikličnom neuronskom mrežom koja će vršiti predviđanje prometne potražnje na prilazima autocesti. Rezultati predviđanja prometne potražnje utjecat će na konačni izračun stupnja propuštanja priljevnog toka u glavni tok. Stupanj utjecaja predviđanja prometne potražnje na računanje konačnog stupnja propuštanja vozila u glavni tok biti će reguliran sa četiri jednostavna pravila. Integracija predviđanja prometne potražnje s izvornim INTEGRA algoritmom omogućilo je kraće kašnjenje na autocesti budući se stvaraju virtualni redovi čekanja na prilazima prije nego zagušenje počne. Spomenute metode upravljanja prometom uspoređene su u usporednoj analizi s drugim uobičajeno korištenim metodama upravljanja prometom na urbanoj autocesti.

Na osnovi rezultata i ograničenja ovog istraživanja, buduće istraživanje moguće je nastaviti u nekoliko pravaca. Korištenje makro-simulacijskog okvira za testiranje spomenutih metoda upravljanja prometom na autocesti. Moguće je povećati skup parametara kriterijske funkcije za analizu. Također, moguće je omogućiti da kriterijska funkcija uključuje više parametra prilikom izračuna odgovarajućeg prometnog rješenja koje će biti uključeno u skup podatka za učenje. Stupanj utjecaja predviđanja prometne potražnje na računanje konačnog stupnja propuštanja vozila u glavni tok treba biti regulirano naprednijom optimizacijskom metodologijom.

Table of Content

List of Abbreviations

AASHTO - American Association of State Highway and Transportation Officials

ACTM - Asymmetric Cell Transmission Model

AI - Artificial Intelligence

ALINEA - Asservissement lineaire d'entree autoroutiere

ANFIS - Adaptive Neuro-Fuzzy Inference System

ANN - Artificial Neural Networks

API - Application Programming Interface

C-ITS - Cooperative Intelligent Transportation Systems

COM - Microsoft Component Object Model

CTMSIM - Cell Transmission Model SIMulator

D - Delay

DIS - Driver Information Systems

DYNASMART - Dynamic Network Assignment-Simulation Model for Advanced Roadway Telematics

FIS - Fuzzy Inference System

GPS - Global Positioning System

GUI - Graphical User Interface

HOT - High-Occupancy Toll

HOV - High-Occupancy Vehicle

IA - Intelligent Agent

ILC - Iterative Learning Control

ITS - Intelligent Transportation Systems

ISA - Intelligent Speed Adaptation

LLM - Local Linear Model

LoS - Level of Service

LR - Reinforcement Learning

LWR - Lighthill-Whitham-Richards Model

MAE - Mean Absolute Error

microSDK - Microscopic Simulator Software Development Kit

MLN - Multileg Network

MoM - Middle of Maximum

MoS - Measure of Service

MRE - Mean Relative Error

NARX - AutoRegressive with eXogenous Inputs

NMSS - Non-minimal state space

OBU - On-board-unit

PeMS - Performance Measurement System

PIP - Proportional-Integral-Plus

PLUS - Prohibiting Lane Use System

RMSE - Root Mean Square Error

RMS-r2v - On-ramp control computer dedicated for communication with vehicles

RNN - Recurrent Neural Network

STN - Spatiotemporal Pattern Recognition Network

SWARM - System-Wide Adaptive Ramp Metering

TDNN - Time-Delay Neural Network

TSS - Transport Simulation Systems

TT - Travel Time

TTS - Total Time Spent

VMS - Variable Messaging Systems

VSLC - Variable Speed Limit Control

VSLCDR - Density Reactive VSLC

VSLCTR - Temporal Reactive VSLC

List of Figures

List of Tables

1. Introduction

The invention of the assembly line by Henry Ford introduced in 1913, revolutionized the automobile industry and the concept of manufacturing worldwide. The invention led to the Ford *Model T* automobile, which was the first sturdy and cheap automobile intended for mass consumption. Over the years, vehicle performance in terms of safety and driving characteristics went up while prices went down. Industrialization based on cheaper production led to the growth of urban areas which consequently led to urban population growth. The population in urban areas became characterized by greater purchasing power and greater demand for goods and mobility. Cities became increasingly populated by residents that owned and used vehicles in their daily routines. This reduced mobility in simple urban traffic networks, which are part of larger urban areas. The result of this effect was the development of more complex urban road networks. With the increase of vehicle speed, safety and more efficient fuel consumption, urban road networks of neighboring cities quickly became connected with a special type of road. This road type enabled faster vehicle speeds and is today known as the motorway (in the United Kingdom, Germany and many other EU country's), the freeway (in Australia and some parts of the USA and Canada), or the expressway (some parts of Canada, parts of the USA, and many Asian countries).

In this thesis, the term motorway will be used since it is commonly used in the EU traffic system. The term and its synonyms are often confused with the term highway, which is general term for denoting a public way for purposes of vehicular travel, including the entire area within the right-of-way [8]. This thesis will make a clear distinction between the terms motorway and highway despite the fact that those terms are often used as synonyms in the literature. According to [9] and the AASHTO "Green Book" motorways are defined as a highway with full control of access and two or more lanes for the exclusive use of traffic flow in each direction. Motorways provide uninterrupted[1] traffic flow. Opposing directions of flow are continuously separated by a raised barrier, an at-grade median, or a continuous raised median [9].

In the past, road traffic networks of larger urban areas had to cope with traffic demand originating in smaller nearby cities, larger central cities and with transit traffic. Transit traffic

[1] Uninterrupted is used to describe the type of facility, not the quality of the traffic flow at any given time. A freeway experiencing extreme congestion, for example, is still an uninterrupted-flow facility because the causes of congestion are internal [1].

relied on several motorways connecting different urban areas' road traffic networks in order to reach its final destination. In order to rid urban traffic networks from transit traffic, special types of motorways known as urban bypasses were constructed in urban area outskirts. With the expansion of the urban areas, urban bypasses became surrounded by urban infrastructure and consequentially integrated with the urban traffic network. The aforementioned type of urban bypasses is known as the urban motorway.

In larger urban areas of the world, there is often no more room for constructional expansion of urban motorways, since they have become surrounded by urban infrastructure and urban traffic networks. Simultaneously, residents of urban areas became aware that it was possible to reach their destinations within the same city more quickly by using the urban motorway - They were convinced that they were avoiding congestions and traffic lights in the urban road network, and consequently saving time. Such commutes with similar driver mind-sets are still happening today on a daily basis. If these commutes are occurring in larger quantities at a specific time of a day then they are known as recurrent daily migrations. Recurrent daily migrations combined with transit traffic and traffic that is originating or has a final destination in a particular urban traffic network can induce congestions or slowdowns in urban motorways and consequently reduce their originally planned higher Level of Service (LoS). Generally, LoS is defined as a group of qualitative measures that characterizes operational conditions within traffic flow and their perception by drivers. In most cases it is defined as the rate of traffic service defined by the maximum hourly rate at which vehicles can cross a point or a road section depending on road, traffic and control conditions. In this case, traffic infrastructure can be evaluated by five traffic rates of service marked form letters A to F (letter A denotes roads with the best throughput, while the F level is not used because it is unstable). In order to describe LoSs of motorways more accurately, several qualitative measures such as travel time, delay, etc. are introduced in the form of Measures of Services (MoS) [9], [10].

An important distinction between the classic motorway and urban motorway is the the fact that an urban motorway has a larger number of entrance/exit ramps, which are fairly close to each other. These are built to achieve better integration with the related urban traffic network. Since on- and off-ramps on urban motorways are close, there is significant traffic dependency between them and they are therefore places where incidents can occur, which can lead to non-recurrent congestions.

Entrance and exit ramps or on- and off-ramps are also places where it is possible to make a significant impact on urban motorway mainstream traffic and on on-ramp queues by using appropriate motorway control methods. Space-wise, on- and off-ramps on urban motorways often give rise to congestion because they merge mainstream and on-ramp traffic flow. Time-wise congestions are common during peak hours [11].

1.1. Motivation and aims

Urban motorways, due to all the previously mentioned reasons, are affected by traffic congestions or at least slowdowns on an almost daily basis. The main objective of this thesis is the reduction or mitigation of congestions and slowdowns on urban motorways by taking into account the characteristic constructional parameters of urban motorways and their role in urban traffic system. The aim is to use the latest solution for mitigation of traffic congestion by applying new traffic control approaches from the domain of Intelligent Transportation Systems (ITS). This is very important since in most cases there is no more space for a constructional build-up of existing urban motorways, as it has already been mentioned. Congestions and slowdowns on high-speed roads such as urban motorways can cause serious problems regarding safety and consequently increase the risk of incidents [11]. The aim of this thesis is not to affect urban motorways safety directly, but only to reduce congestions and slowdowns. Indirectly, by reducing urban motorways slowdowns and congestions, it is possible to reduce the risk of incidents and increase safety since congestion is one of the factors which can cause certain types of incidents [10].

1.2. Methods

One of the most used traffic control method on the urban motorway for mitigation of motorway congestions is ramp metering. The main goal of ramp metering is to increase the throughput of urban motorways by restricting access to on-ramp traffic by using special traffic lights [2]. Significance of ramp metering for its future development in European Union is highlighted by the EURAMP project funded by the 6th RTD Framework Programme. Publication [12] is the capital deliverable of this project and can be seen as a framework for the future implementation of ramp metering on European urban motorways.

The most important part of ramp metering is an appropriate control algorithm that makes decisions concerning the quantity of on-ramp vehicles/traffic flow [2] allowed to merge with mainstream traffic flow. The usefulness and the effectiveness of a ramp metering algorithm significantly depends on its ability to react to unforeseen situations such as incidents, vehicle breakdowns and rapid changes in traffic demand within a short time interval. In this thesis, an integration of existing ramp metering algorithms into a new ramp metering algorithm (INTEGRA) is propounded. INTEGRA, the new ramp metering algorithm is made to cope with the mentioned challenges in the design of ramp metering algorithms. Integration is conducted by using an advanced learning framework based on the Adaptive Neuro-Fuzzy Inference System (ANFIS) architecture, [13]. The evaluation of various existing standalone ramp metering algorithms and the newly developed INTEGRA algorithm is performed by using the CTMSIM simulator. The CTMSIM is a macro-simulation tool for simulating traffic flows on a motorway system and will be used during development of the INTEGRA algorithm as well. A use case scenario will be created based on the Zagreb bypass which can be considered an urban motorway due to its operational characteristics. To cover the wide range of traffic scenarios on the Zagreb bypass, a case model is simulated using three distinctly different ramp metering algorithms (ALINEA, SWARM, and HELPER). All three algorithms are chosen as the teaching algorithms for the proposed INTEGRA ramp metering algorithm and will be described in more details later.

Ramp metering has recently been used in cooperation with additional motorway control methods like Variable Speed Limit Control (VSLC), Prohibiting Lane Use System (PLUS), Driver Information Systems (DIS), etc. This thesis will also describe the development and the evaluation of a control method based on cooperation between ramp metering and density reactive VSLC. Cooperation between the VSLC and ramp metering decreases the speed of incoming mainstream vehicles to congested on-ramp areas. Under certain conditions, this can reduce the shock wave back propagation in the mainstream flow between the congested on-ramp area and upstream VSLC regions [14] [6] [15]. In this thesis, the original CTMSIM macroscopic motorway traffic simulator is augmented in order to enable the simulation of cooperative ramp metering and VSLC. Cooperation between several ramp metering systems and VSLC will be evaluated in line with the previously mentioned motorway control methods. Furthermore, it is important to emphasize that ramp metering and the VSLC can be modified in order to reduce the risk of incidents and increase safety on urban motorways. However, this thesis will not be dealing with that.

1.3. Research objective and hypothesis

Based on the stated motivation and aims, the research objectives and hypothesis will now be defined. The objective is to develop several types of motorway control methods based on the concept of cooperation and to evaluate them in relation to existing standalone applications. An additional objective is to introduce a new ramp metering learning framework with the ability to learn different ramp metering control behaviour. In order to reach that goal, the following hypotheses are proposed:

1) A ramp metering approach based on machine learning can learn control behavior from the most widely used teaching ramp metering algorithms, and produce lower travel time compared to mentioned ramp metering algorithms, with an acceptable increase of on-ramp queues or overall delay. These results are especially noticeable in conditions of significant changes in traffic demand on urban motorways.
2) The algorithm that is based on cooperation between different on-ramps and cooperation between ramp metering and the VSLC can produce significantly better results than standalone ramp metering and VSLC algorithms.

1.4. Expected scientific contribution

Based on the research objectives and hypotheses, and taking into account the employed methods and their limitations, this research is expected to expand knowledge in the field of Traffic and Transport Technology and Control with the following research outputs:

- An advanced learning framework for ramp metering that will be able to cope with significant changes in traffic demand on urban motorways.
- A model of cooperative motorway management strategies applicable to ramp metering and variable speed limit control.

1.5. Outline

In the introductory chapter, the motivation for the research, the hypotheses, and the research objectives were presented. Additionally, the overview of the methods used and the expected scientific contribution was given.

The second chapter, titled **Current problems on urban motorways**, contains the definition of the urban motorway and the concept of its evolution in the context of urban region development.

Furthermore, this chapter describes the recurrent and non-recurrent congestions on urban motorways and defines them in spatiotemporal terms.

In the third chapter, titled **Methodology for the design of urban motorway control methods**, a general overview of the current urban motorway control methods is given. The methods are divided into three categories: ramp metering, VSLC, and PLUS. The most prominent algorithms used by the aforementioned control methods are described, with special emphasis on algorithms from the domain of machine learning. This chapter also tackles problems related to cooperative control implemented in urban motorway control systems. Simulation environments for simulating various urban motorway control methods are categorized in this chapter, described in detail, and compared.

In the fourth chapter titled **Ramp metering based on machine learning,** a general overview of current approaches in the application of machine learning methodologies in ramp metering algorithm development is given along with the proposition of the new INTEGRA algorithm based on the same methodology. The INTEGRA algorithm is augmented in order to take into account prediction based data, and the search for the best setup of the INTEGRA criteria function is also described. Furthermore, this chapter includes a description of the cooperation between cooperative ramp metering and the VSLC.

In the fifth chapter, titled **Results and discussion,** results of several comparative analyses based on several Measures of Services (MoS) are presented. Comparative analyses include all the mentioned algorithms which are related to urban motorway control. This chapter also includes results regarding the different setup of criteria functions for the INTEGRA algorithm.

The final chapter contain conclusion and proposals for future research. In this chapter, all relevant research objectives are reviewed and the effect of all the used and newly developed urban motorway control measures is elaborated. In conclusion, the overall best modification of the INTEGRA algorithm is selected based on the comparative analysis of the MoS.

2. Current problems on urban motorways

Traffic congestion and slowdowns are the main problems of almost every traffic network and therefore the main focus of traffic engineering. An urban motorway can be considered a unique sub-system of the overall urban traffic network due to its complex integration with the urban traffic network via on- and off-ramps and its originally planned large capacity. In many cases, they represent a bypass between suburban areas and urban centres. In these cases, urban motorways can serve as the urban traffic network's backbone. Congestion related problems on the urban motorways are a consequence of their constructional characteristics, their position and role in the urban traffic network. This chapter will describe the concept of urban motorways with an emphasis on their historical genesis and their current role in the urban traffic network. It will then deal with the temporal and spatial aspects of the sources and impact of traffic congestion on urban motorways.

2.1. Concept of the urban motorway

Larger urban area road networks and the traffic demand originating within a particular urban area have to cope with the traffic demand of nearby smaller cities, central larger cities, and transit traffic. Smaller cities can be interconnected via inter-city roads. In most cases, standard inter-city roads with a small projected capacity could not deal with the gradual increase of traffic demand brought on by the expansion of neighbouring cities. The solution was the construction of motorways. Their main goal was to connect larger urban areas that have evolved from neighbouring smaller cities directly. Originally, they were designed to provide larger maximum traffic capacity by enabling a higher LoS as opposed to urban road networks and inter-city roads. The secondary role of such motorways was to serve as a bypass around larger urban areas.

In Croatia, motorways are called *autoceste* (Croatian pronunciation: [ˈaʊtotsesta]), and they are defined as roads with at least three lanes in each direction (including hard shoulder). They are marked by a special road sign, similar to the road sign depicting a motorway/autoroute/autobahn in other parts of Europe [16]. In Croatia, this sign has a green background. The national speed limit on the motorway, in case no other speed limits are in effect, is 130 km/h (81 mph), with a legal tolerance of 10% on speeds over 100 km/h (as of 2009).

Motorways, which had originally served as urban bypasses, quickly became surrounded by urban infrastructure and its corresponding road traffic network as larger urban areas expanded. In most cases, the arterial or adjacent roads of the urban traffic network became directly connected with on- and –off ramps of the urban bypass. That process enabled a strong integration of urban bypasses with the urban traffic network. Furthermore, they also became a critical part of the urban traffic infrastructure for connecting suburbs with urban centers. The aforementioned role of the urban bypass led to its frequent use for daily migrations by residents of particular urban areas. Taking into the account all the stated constructional characteristics and their role in the urban traffic network, urban bypasses can be considered urban motorways. Urban motorways are characterized by a larger number of on- and off-ramps, which are fairly close to each other as compared to standard motorways. The constructional characteristics of urban motorways consequentially induce greater dependency between traffic flows generated by neighbour on-ramps. Greater dependency between traffic flows generated by neighbour on-ramps regularly induces lower average mean speeds in the urban motorway mainstream. This effect is most prominent during peak hours when traffic demand is rapidly increasing.

Overall traffic demand, which affects urban motorways, whether originating in adjacent urban traffic networks or in transit traffic, causes their overload. The main approach to this problem in the context of urban traffic networks would be a classical constructional expansion of the existing capacity. Urban motorways are usually surrounded by buildings and other infrastructure, so in most cases physical expansion of their capacity is not possible. Even when physical expansion of urban motorways is possible, it is usually economically untenable. In Figure 1, it is possible to see the urban motorway in Berlin which is fully integrated with the urban traffic network. It can be seen that the mentioned urban motorway connects the suburban area with the urban centre.

Figure 1: Example of the Berlin urban motorway network

2.2. Traffic congestions on urban motorways

Almost every day one can witness traffic congestion or at least slowdowns in various traffic flows. This is especially the case on urban motorways designed to provide a higher LoS in the context of fast and safe traffic flow, but similar congestions can often occur on other types of roads. Congestions or jams are caused by bottlenecks. A bottleneck is defined as a local reduction of road capacity [17]. With respect to the genesis of bottlenecks, they can be divided in permanent (static) or moving (temporary) bottlenecks. Moving bottlenecks are usually induced by slower vehicles or motorway segments affected by speed limits. This induces a moving jam, which can be defined as a localized structure that moves upstream of mainstream flow [18]. The moving jam or stop-and-go wave has an upstream moving downstream front (jam head) and an upstream moving front (jam tail). Within the jam fronts vehicle speeds, flow rate, and density vary sharply.

On the other hand, permanent bottlenecks have a relatively static downstream front. Permanent bottlenecks can be a result of roadworks, increased traffic demand originating at on- and off-ramps (characteristic of urban motorways), a decrease in the number of traffic lanes, traffic incidents, road curves and road gradients, etc. Congestion shockwaves propagation length or propagation velocity of the jam tail induced by the aforementioned types of bottlenecks will vary in time depending on the traffic situation of the upstream jam and the jam itself. In [17] and [18] shockwave propagation against the driving direction with a general characteristic velocity in the order of 10 to 20 [km/h] can be seen [17], [3].

The onset of traffic congestion is accompanied by a sharp and sudden drop in average vehicle speed. This effect is known as the traffic breakdown phenomenon [3], [18]. The downstream front of a congestion is a place where vehicles usually accelerate towards the space of free flow (downstream of the fixed bottleneck head). It has been found that the capacity of a congested bottleneck, i.e. after the breakdown phenomenon at the bottleneck has occurred, is often lower than the capacity in free flow state before [17]. This phenomenon is called "capacity drop". The capacity drop in a homogeneous traffic flow is illustrated in the fundamental diagram (density-flow relation) in Figure 2.

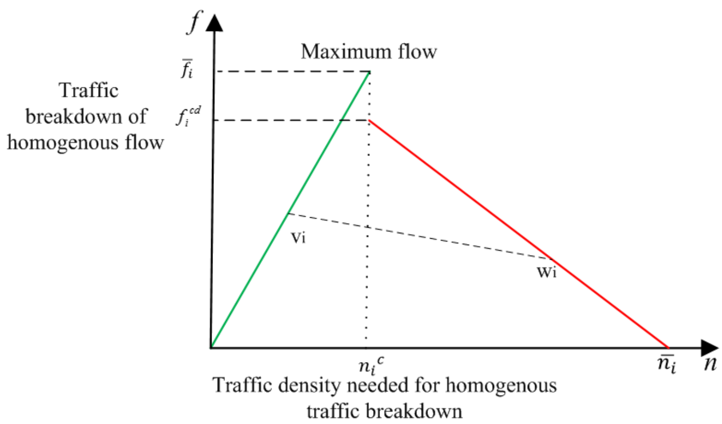

Figure 2: Capacity drop in homogeneous traffic flow illustrated in the fundamental diagram

In Figure 2 $\bar{f}_i$ represents the maximum traffic flow in motorway segment i, f_i^{cd} is the traffic flow in the motorway segment after the traffic breakdown, n_i^c is the critical density of the motorway segment i, $\bar{n}_i$ is the maximum possible density that can be achieved in motorway segment i, v_i is the free flow speed in motorway segment i, and w_i is the jam speed in motorway segment i. Space-wise and time-wise, congestion occurs in parts of an urban motorway near large urban areas during the early morning or late afternoon (known as peak hours). The source of the congestion is attributed to daily migrations, to and from one's place of employment, education, etc. These migrations are characterized by intense traffic demand that occurs in short time intervals. Consequently, the aforementioned effect can induce a traffic breakdown phenomenon in bottleneck areas. This is usually refer to as an effectual bottleneck. An effectual bottleneck is one where transition from the free flow to congestion flow frequently occurs in clear temporal patterns [17]. In other words, if daily migrations are intense and synchronous, they can produce an effectual bottleneck that will induce so called recurrent congestion.

Recurrent congestions are easy to predict and therefore easier to handle because it is stable in space and time. On the other hand, congestion that usually causes a sudden drop in the traffic throughput of a particular motorway is known as a non-recurrent congestion. The frequency of this congestion type cannot be described by any clear temporal pattern. The main causes of non-recurrent congestion are various traffic accidents or events of great public interest (sports events, concerts, sales in malls, etc.). Unlike recurrent congestion, which originates from individual daily routines of citizens, non-recurrent congestion (if there are no announced public events) are very hard to predict and therefore harder to handle.

As it is mentioned earlier, problems with congestions on urban motorways are most noticeable near urban areas. Consequences of urban motorway congestions manifest themselves with the following indicators: traffic demand exceeds road capacity, increased number of accidents and incidents, queues at on-ramps spill over into urban traffic arterials (spillback effect) and induced peaks in traffic demand which are the result of platooned vehicle entry from on-ramps [19]. For urban motorways, on-ramps are the crucial places because they are directly connected to adjacent urban traffic networks. The connection is usually made by urban arterial roads. Another issue related to the on-ramps is related to significant dependency between them in a traffic context. Traffic flows originating from on-ramps depend greatly on each other as they merge with mainstream flows due to the short average distance between on-ramps. Drivers on the on-ramps that are merging with the mainstream flow very slowly can produce a *spill back effect* since the on-ramps are heavily burdened by traffic demand. On the other hand, there are problems with aggressive driver behaviour as well. Even when the mainstream is near maximum capacity, it can adopt one or two merging vehicle from an on-ramp. However, in the case when platoons of vehicles attempt to do an aggressive breakthrough into the mainstream flow, turbulence appears. This turbulence usually causes a mainstream traffic breakdown and consequently, an effectual bottleneck [8]. Turbulence in merging zones can also cause various types of accidents in heavy traffic conditions. It can be concluded that a common source of periodic congestions on urban motorways is in heavy on-ramp flows, which are originating from on-ramps and merging into the mainstream traffic flow. The place on the urban motorway where these two flows actually come into contact is known as a downstream bottleneck. This bottleneck can become an effectual downstream bottleneck if it is frequently affected by the traffic breakdown phenomenon.

Generally speaking, any location on the urban motorway where the downstream traffic front (jam had) of the congested pattern is spatially fixed and frequently affected by the traffic breakdown phenomenon will be an effective location of the effectual bottleneck [17]. Besides locations near on-ramps, such locations can be also be found near road curves, induced by a decrease in a number of traffic lanes, etc. In Figure 3 the position of an effective location of the downstream bottleneck and the spillback effect on the adjacent local urban road network is illustrated.

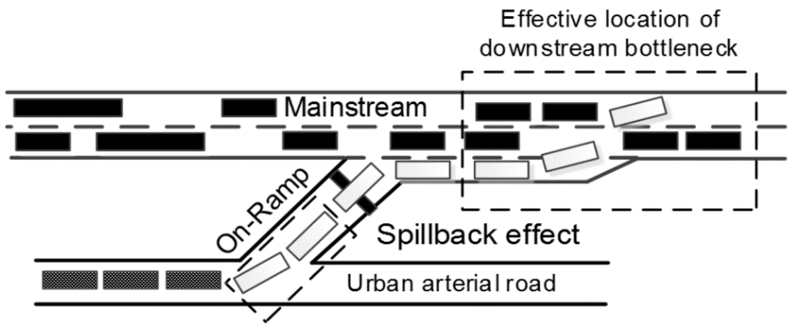

Figure 3: Illustration of a downstream bottleneck location and spillback effect

Conventional traffic engineering methods (e.g. speed limits) are applied for all traffic situations on a particular urban motorway without taking into account the current state of traffic flow and human factors. As was mentioned earlier, the classical build-up approach in the context of urban motorways' capacity expansion is usually technically very difficult or/and economically untenable due to its enormous overall cost. The answer to the problem is better control of traffic flows that are using road capacities. In general, better control of traffic flows in any segment of traffic infrastructure can be achieved by the application of specific traffic control methods [20]. These methods are specifically designed for the particularities of certain types of traffic infrastructure, the traffic legislation concerning specific types of traffic infrastructure (which has to be observed during construction and use), and the specific behaviour of traffic flows. Today, all traffic control methods are usually observed as part of a broader concept known as an ITS. In this context, an ITS can be defined as a holistic, managerial, and *Information and Communications Technology* (ICT) upgrade of classic transport systems which allows significant improvements in the performance of traffic flows (reduction of congestions and incident situations and an increase of a motorway LoS) and generally improves the safety of traffic users [21].

3. Methodology for the design of urban motorway control methods

In general, a higher projected LoS for urban motorways can be significantly reduced due to traffic congestions and slowdowns. In order to reduce the impact of congestions on urban motorway LoS, control over the traffic flows at on-ramps and mainstream flows with respect to motorway legislations constraints should be established. For example, mainstream flow on motorways cannot be controlled by traffic lights, since that type of traffic flow cannot be completely stopped by any traffic control method. This is one of the constraints related to the control of traffic flows on urban motorways.

In order to enable control over the traffic flows on the urban motorway, three types of traffic control methods have been widely implemented: ramp metering, the VSLC and PLUS [3]. The main purpose of ramp metering is to regulate the traffic flow rate from on-ramps by using specialized traffic lights. The VSLC conducts the homogenization and reduction of vehicle speeds which consequently decreases the time needed to create a capacity drop and backpropagation of a traffic shockwave [22]. Computed speed limits are posted for drivers via Variable Messaging Systems (VMS). PLUS controls the number of active traffic lanes on a motorway mainstream. Each of the mentioned traffic control methods uses control logic, which is developed on the basis of specific control requirements. In most cases, the control logic is written/coded in the form of an algorithm – a self-contained step-by-step set of operations that have to be performed. Furthermore, it is important to mention that each traffic control algorithm is designed based on a specific target function as the final goal which has to be met in the course of its operation.

Early control logic solutions relied on a manual (operator) based control approach without the need for an algorithm design. The first traffic control algorithms used a control logic which was created on the basis of previously collected traffic data (historic data) analysis. These algorithms had short repeatable control cycles with fixed signal plans, since they did not take ongoing traffic scenarios into account. Therefore, the mentioned traffic control structure was completely unaware of sudden traffic flow fluctuations (referred to as noise in control theory) which are very common in a traffic flow due to its stochastic nature. A time reactive control structure of algorithms can be considered the first significant step towards improving fixed control structures. The aforementioned approach brings about a more robust control output of traffic flow fluctuations. This type of traffic control algorithms provides traffic control action based on the time of day and/or specific days (e.g. opening more traffic lanes during peak hours, different traffic light signal plans during the weekend, etc.). This control structure requires current time as its input, but it still performs poorly when dealing with traffic fluctuations unrelated to peak hours or a specific day of the week. Most contemporary traffic control algorithms are traffic responsive control algorithms. They have a logic structure which enables adaptation to current traffic flow fluctuations and, compared to the other mentioned control structures, they provide more comprehensive and robust control results because of this ability. These algorithms require the acquisition of traffic data in real time and are based on data processing which occurs within the traffic control algorithm when the control output is computed.

All this being said, from a control theory perspective, traffic responsive control logic can be divided into open and closed loop (feedback) control structures [23]. The control action in an open loop control structure is computed by a control algorithm which is not aware of the "system output" (or value of the "controlled process variable"). On the other hand, control algorithms based on a closed-loop control structure compute the control action in accordance with a system (motorway) output by using a feedback loop. It can be concluded that the closed loop structure is more robust in the face of frequent changes in traffic systems due to its awareness of how its previous decisions affected the system in each control step. In order to attain traffic responsive control, it is imperative to acquire the traffic parameters of an adequate motorway section. The traffic parameters are acquired by using different traffic sensors (e.g. inductive loops, traffic cameras, ultrasonic sensors, etc.). All mentioned control logic structures affect traffic flows on motorway system by using traffic actuators, e.g. traffic lights, VMS, etc. Figure 4 presents a generic structure of traffic responsive control based on a closed control loop.

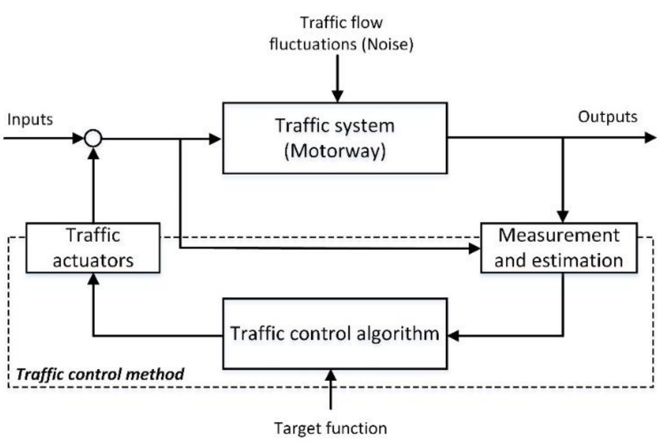

Figure 4: Generic structure of traffic responsive control based on closed control loop

It is important to emphasize that all the mentioned control methods can be observed in the context of the ITS domain, so integration, coordination, and cooperation with other traffic control and information systems in the domain is a possibility. This chapter will provide insight into the currently most frequently used algorithms which form the basis of VSLC and ramp metering control methods. In most cases PLUS is used as the supporting technology for ramp metering and the VSLC but it can be applied on its own in incident situations or in the case of roadworks.

In most cases, it is difficult, and/or in collision with the local legislature, to test complex algorithms for urban motorway control methods on actual motorways. Therefore it is necessary to use traffic simulation tools in order to model a particular motorway system utilising the desired traffic control method (which is planned to be implemented on particular motorway system). It is then necessary to conduct a simulation of the created motorway system model utilising the developed traffic control method. After one or several simulation runs it is possible to evaluate the simulation results and compare them to other potential traffic control methods for the same motorway or to a no control scenario. Furthermore, if the simulations show promising results, they can be presented to the authorities or operational personel and the traffic control method can be applied to the actual motorway system.

3.1. Simulation of motorway traffic flows

A motorway system can be considered as a traffic system that contains numerous on- and off-ramps (used by on- and off-ramp flows) connected directly to motorway main-lanes (consisting of two mainstream flows oriented in opposite directions). On a larger scale, a motorway system is usually presented as a set of linked nodes. Motorway nodes contain several on- and off-ramps fairly close to each other, so there is a strong interaction of traffic flows between them. The links represent straight elements of motorway which connecting motorway nodes. It is possible to simplify the modeling process, and model a motorway section by dividing it into numerous cells. Each cell represent a part of the node or a whole link between nodes. Furthermore, it is possible to model a motorway system as a collection of links interconnected by connectors, etc. Every traffic simulator (a simulation program) contains a set of tools that are used to model motorway systems and/or other types of traffic networks with respect to its traffic simulation model. A particular traffic simulator’s approach to motorway modeling methodology will significantly depend on its traffic simulation model.

Traffic simulation models are an integral part of traffic simulators and can be classified according to discrete or continuous values of time, traffic state, and space. Additionally, traffic models can be microscopic, mesoscopic or macroscopic according to their representation of a traffic flow or vehicle movement [24], [1].

The microscopic traffic model calculates the parameters of every individual vehicle continuously or discretely (e.g. position, speed, acceleration, etc.) during the entire simulation run, [1] . Simulators based on microscopic traffic models take into account relatively small changes in the physical environment, such as switching a mainstream lane or the speed differential between merging on-ramp flows and mainstream flows, but requires more computational and modeling time for larger motorway systems. This type of traffic model can provide very accurate data on a traffic situation at a motorway node (e.g. 2 on-ramps, 2 off-ramps), but should more nodes be present, it can be unpractical. Some of the commonly known traffic micro simulation programs are the PARAMICS, the MITSIM, the CORSIM, the SUMO, the VISSIM, the AIMSUN2, the TRANSIM, etc.

The macroscopic model calculates cumulative traffic flow characteristics (e.g. speed, flow, and density) and their inter-relationships on the basis of traffic flow equations. Traffic flow equations describe traffic disturbances, which are broadcast through the motorway system in the form of shockwaves. A macroscopic model simulation is computationally less demanding and its motorway modelling approach is oriented towards the segmentation of the entire motorway system. Macroscopic models were originally developed for motorway systems because of their ability to predict/simulate the spatial and sequential extent of congestion caused by exceeded traffic demand or incidents in a motorway network, [1]. Congestion and incident prediction in a spatial and temporal context is the most desirable feature of a ramp metering simulation. This is the reason why macroscopic models are widely used for simulating ramp metering. The disadvantage of this traffic model is the inability to model the interactions of individual vehicles between the on-ramp and mainstream traffic flow. Traffic simulation programs that use macroscopic models are the CTMSIM, MASTER, the EMME, SATURN, TransCAD, the VISUM, etc.

Mesoscopic models combine the properties of microscopic and macroscopic simulation models. This type of model defines and monitors the states of each individual vehicle in a similar way as a microscopic model does, but the activities and interactions between vehicles are based on aggregated (macroscopic) relationships. Mesoscopic models are usually applied when the simulation of in-vehicle and real-time travel information systems [1] is required. This type of model is usually tailor-made, so the model can be adjusted for a ramp metering model simulation if need be. The drawback of such a simulation model is that it can be computationally intensive, and consequently demand a large amount of computation power. Traffic simulation programs which utilise mesoscopic models are the Cube Avenue, DYNASMART, INTEGRATION, METROPOLIS, the VISSIM (optional), etc. In Figure 5 it is possible to see an illustration of a motorway simulation in three different representation levels regarding used traffic model.

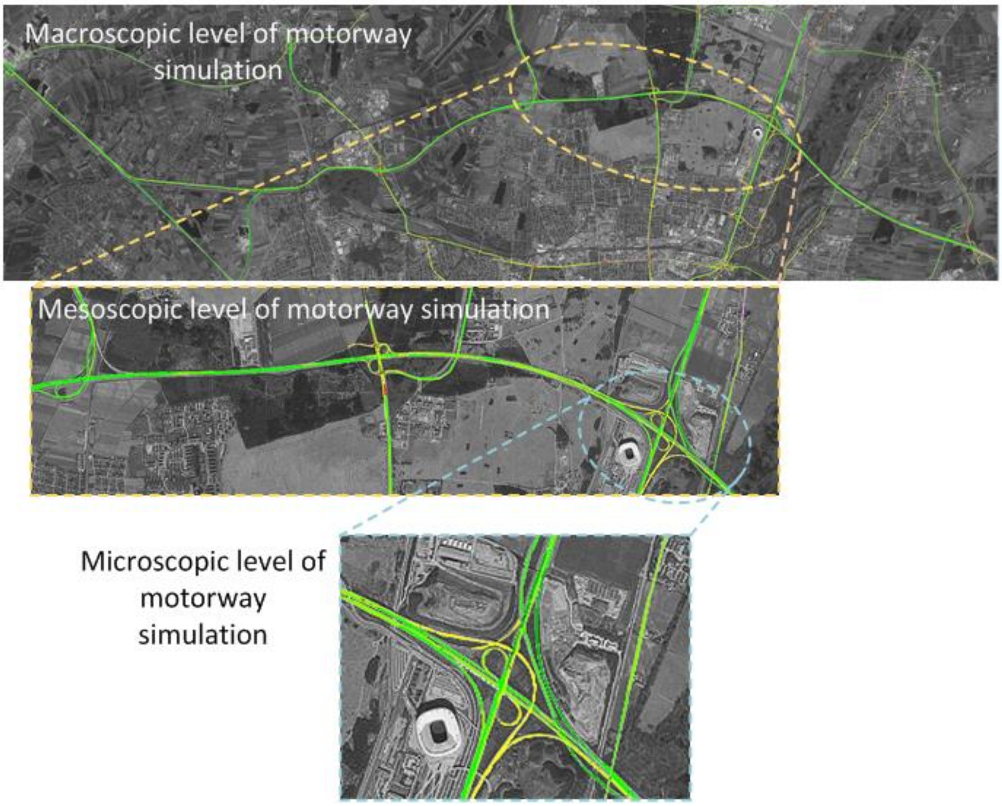

Figure 5: Illustration of a motorway simulation in three different representation levels with respect to the utilised traffic model

Each of the mentioned traffic models has their own strengths and weaknesses which will be covered in this chapter along with the most widely used simulators which use them. At this point, one can assume that a motorway model is created. The next step is to integrate designed control algorithms with simulator frameworks. Some simulator frameworks conduct discretization of simulation times into the time steps or do not support the direct design of signal plans. This is the case with macroscopic based simulators so it is necessary to adequately adapt the structure of the aforementioned algorithms and their control actions to these constraints.

Generally, all the mentioned traffic models can be used for the simulation of ramp metering and the VSLC. Microscopic simulators do not depend on theoretical traffic flow models but on vehicle to vehicle interactions. They are appropriate for evaluating local ramp metering algorithms. Macroscopic oriented traffic simulators are faster and better for evaluating complex control methods such as cooperative motorway control on larger motorway systems since such control approaches are computationally more expensive.

In simulating ramp metering and the VSLC it is imperative to achieve adequate simulation accuracy and simulation speed with respect to the size of motorway system and the complexity level of the control algorithm. The simulation speed is an especially important issue in the process of simulating ramp metering and VSLC control methods on larger motorway systems with numerous on- and off-ramps. This is even more important when advanced traffic control approaches which include different optimization methods, estimation, forecasting and machine learning related computations are used. Simulating such complex control systems can be time-consuming.

Advanced detailed visualization and high realism of driver behaviour are common in advanced commercial simulators based on microscopic models such as the VISSIM, the AIMSUN, and PARAMICS can additionally reduce simulation speeds. At the same time, their advantages are the potential for visual inspection of traffic flows during a simulation as well as high simulation accuracy. The CTMSIM has a relatively simple simulation visualization interface and conducts macroscopic traffic modelling which enables a general evaluation of traffic flows as opposed to focusing on every car in a flow separately. These features of the CTMSIM enable much faster simulations of the same traffic processes when compared to the other described simulators. This is the case even when more complex algorithms for traffic control are simulated.

It is important to emphasize that the CTMSIM is developed exclusively for simulating motorway traffic flows and the influence of ramp metering on them. This gives the CTMSIM a certain advantage over other commercial simulators, which cover a much wider range of traffic related environments and control methods. At the same time, the CTMSIM is free and the entire source code is open for user customization. If need be, users can also add a more sophisticated visualization interface.

There are a lot of simple motorway simulators designed for relatively specific purposes such as assessment of tooling, incident management, congestion analysis on a specific type of traffic network, etc. One of the most important requirements for a ramp metering and a VSLC simulation is the production of results useful for the evaluation of implemented traffic control methods. Modern commercial simulators offer a wide selection of simulation result representations, but it is still very difficult to find a simulator with an adequate output representation. The best type of result representation for ramp metering and VSLC design is one with the traffic parameters presented independently for each motorway segment, link, node, etc. Most modern simulators do not support this kind of result representation. If a developer wants to add an advanced ramp metering, a VSLC traffic control approach to a motorway simulation or a desirable representation of output results, he must often add this feature in the simulator source code or use COM interfaces between the application (which contains traffic control algorithm and simulator output data processing) and the simulator. Modern traffic simulators such as the VISSIM, PARAMICS Quadstone, the MITSIM, the CTMSIM, etc. allow adding a specific type of traffic control algorithm and produce an adequate result representation for their evaluation. The most important simulators applicable to ramp metering and VSLC simulations will now be described briefly.

3.1.1. Microscopic traffic simulators

Microscopic traffic simulators are among the most widely used in traffic engineering. Traffic flows and their interactions are modelled based on the description of the motion of each individual vehicle composing these traffic flows [25]. The motion of each vehicle is described in terms of its acceleration, deceleration, lane changes, etc. in response to the surrounding vehicles in simulated traffic flows. The core mathematic elements of microscopic traffic models are car-following models as a form of stimulus-response equations, where the response is the driver reaction to the motion of a vehicle which is immediately preceding him in a simulated traffic flow [26]. The *General Motors Group* developed series of models for simulating car-following behaviour whose basic equation is:

$$Response(t+T) = Sensitivity \times Stimulus(t). \tag{1}$$

The response is always to accelerate or decelerate in a proportion to the magnitude of the stimulus in time t and begins after a time lag T, which represents the reaction time of the follower [1]. The simple model assumes that sensitivity denoted by ϑ is constant. If x_n(t) and x_{n-1}(t) are the positions of the leader and the follower, respectively, in time t, then the linear car-following model is:

$$\ddot{x}_{n+1}(t+T) = \vartheta(\dot{x}_n(t) - \dot{x}_{n-1}(t)), \tag{2}$$

where the response is acceleration, while deceleration depends on the sign of stimulus: (1) positive if the relative speed is positive, (2) negative if the relative speed is negative, (3) or no action if speeds on the left side of an equation (2) are equal [25]. Cellular automaton (CA) models are also popular in microscopic modelling. Each road section can either be occupied by a vehicle or be empty and the dynamics are given by the update rules. Depending on used CA model update rule is given (e.g. Rule 184, Biham–Middleton–Levine traffic model, Nagel–Schreckenberg model) [27].

Their application is common when dealing with smaller parts of traffic networks (those consisting of one or several signalized intersections in an urban traffic network), or a motorway section with one or several on- and off-ramps grouped in separate motorway nodes, etc. As was mentioned earlier, the strength of microscopic models lies in higher simulation accuracy, since each vehicle is modelled separately. In the case of more complex traffic networks burdened with heavy traffic loads, higher simulation accuracy can cause problems with computational time.

The VISSIM is an illustrative example of simulators based on microscopic traffic models. The simulator uses the so-called psycho-physical driver behaviour model originally developed by Wiedemann (1974) [25]. Along with the car following model, which is based on the psycho-physical Widemann model, VISSIM uses models for lateral vehicle movements, which includes lane selections, lane changing, and continuous lateral movement. Classical linear car-following and lane changing models are additionally extended in a form of a tactical driving behaviour, which is oriented on planning ahead of vehicle movement in a temporal and spatial dimension. Vehicle movement in traffic network and the corresponding traffic demand can be modelled based on the, (1) fixed routes and (2) dynamic assignment where route search, route assessment, and final route choice for each individual vehicle are governed by a specialized algorithm. VISSIM also supports pedestrian movement modelling. In Figure 6 it is possible to see an example of modelling of an on-ramp with priority rules in the VISSIM.

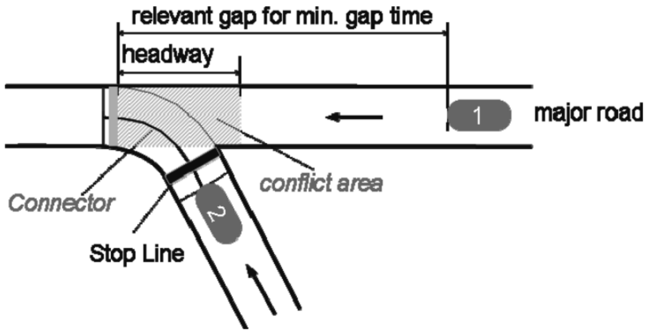

Figure 6: Example of modelling an on-ramp with priority rules in the VISSIM [25]

The VISSIM is widely used for simulating and solving various problems by traffic engineers in practice as well as by researchers for developments related to road traffic. Traffic network modelling is based on links and connectors. Connectors connect links in order to form traffic networks. VISSIM contains an additional interface based on the Microsoft Component Object Model (COM), which is a technology tasked mainly with enabling inter-process communication between an external application and the VISSIM. The external application can contain various data processing and traffic control algorithms. For example, it can be MATLAB or a standalone application written in a high-level programming language. The VISSIM COM interface defines a hierarchical model in which the functions and parameters of the simulator originally provided by the GUI can be manipulated through programming. Using the VISSIM COM the user is able to manipulate the numerous attributes of internal objects dynamically [28].

Aimsun was developed at the University of Catalonia but was commercialized and distributed under Transport Simulation Systems (TSS) [25]. Originally, the Aimsun acronym stands for an advanced interactive microscopic simulator for urban and non-urban networks. The main application of the Aimsun simulator is the improvement of construction and planning of road infrastructure, application of methods for pollution emissions and congestion reduction (ITS services), and design of urban environments for vehicles and pedestrians.

One of the most advanced features of Aimsun is its multithreaded architecture that enables high speeds in running simulation processes. Therefore, the modelling and simulation of a major city traffic network or large and complex motorway systems can be done much faster in comparison with other microscopic simulators. The car-following model is based on the model proposed in [29]. It presents an extension of the traditional empirical model, in which the model parameters are not global but determined by the influence of local parameters depending on the type of driver (e.g. speed limit acceptance of vehicle), road characteristics (speed limit on the section, speed limits on turnings, etc,), the influence on vehicles in adjacent lanes, etc. [25]. These features are especially interesting for ramp metering and VSLC simulations . Along with the mentioned car-following model, Aimsun contains Lane-change, Look-ahead and gap-acceptance models. Furthermore, Aimsun also applies Dynamic Assignment in stochastic/discrete route choice. Dynamic Assignment in routing forms all routes that will be simulated by using Origin-Destination (OD) matrices. These matrices governs traffic flow between defined origin and destination pairs for each planned route.

The Application Programming Interface (API) of Aimsun contains a collection of functions in Python and C++ programming languages. These functions allow the implementation of ITS related elements into a simulation and enable the design of non-standard adaptive traffic control, advanced traffic management, vehicle guidance, etc. The potential of non-standard adaptive traffic control is especially significant in ramp metering because it enables the development of advanced ramp metering algorithms. The Aimsun Microscopic Simulator Software Development Kit (microSDK) enables users to override Aimsun's behavioural models (car-following, lane-changing, etc.) and create customized behavioural models, which can be programmed in C++.

3.1.2. Macroscopic traffic simulators

The method of modelling traffic flow at a macroscopic level originated from the assumption that traffic streams as a whole are comparable to fluid streams [30]. The first major step in the macroscopic modelling of traffic was taken by *Lighthill and Whitham* in 1955, when they indexed the comparability of "traffic flow on long crowded roads" with "flood movements in long rivers" [31], [32], [33]. A year later, *Richards* (1956) complemented the idea with the introduction of "shock-waves on the highway", completing the so-called LWR model [33]. Macroscopic modelling may be classified primarily with respect to the type of traffic as homogeneous and heterogeneous, and then with respect to the order of the fundamental mathematical model.

This section will provide a brief overview of macroscopic modelling requirements since the macroscopic traffic models are selected for simulating ramp metering in this thesis. Macroscopic simulation tools usually model traffic flows based on the continuum traffic flow theory. The main objective of this kind of modelling is to describe the time-space (*x*-*t*) evolution of characteristic traffic parameters used for macroscopic flow definition: volume *f(x,t)*, speed *v(x, t)* and density *n(x, t)*. These parameters are defined at every instant in time *t* and every point in space *x*. Today several equations exist with the main task to formally represent this theory. The most widely used is the conservation or continuity equation (3):

$$\frac{\partial f(x,t)}{\partial x} + \frac{\partial n(x,t)}{\partial t} = 0. \tag{3}$$

To solve the equation (3) it is imperative to provide the hypothesis that flow *f(x,t)* is a function of density *f* = *f(n)* or equivalently, that speed is also function of density *v* = *v(n)*. This assumption only holds if there are no on- or off-ramps or, in other words, if the motorway system is in a state of equilibrium. The equation (3) can be enhanced with the function *g(x,t)* which represents vehicles entering and leaving mainstream traffic flow [25]:

$$\frac{\partial f(x,t)}{\partial x} + \frac{\partial n(x,t)}{\partial t} = g(x,t). \tag{4}$$

Also, it is imperative to include the speed-density equation of the state or some of the theoretical speed-density relationship *v* = *v(n)*. Often the *May-Keller* empirical equation is used for this purpose:

$$v = v_f \left[1 - \left(\frac{n}{n_{jam}}\right)^{\alpha}\right]^{\beta}, \tag{5}$$

where v_f is the free flow speed, n_{jam} is the jam density and, α and β are the calibration parameters. Furthermore, *Payne* replaced equation (5) with a second order partial differential equation corresponding to the momentum equation in fluid dynamics. This was done because equation (5) could not accurately describe non-equilibrium traffic flow dynamics. The *Payne* equation was a great breakthrough in simulating vehicle merging processes between on-ramp and mainstream traffic flows. Nevertheless, the *Payne* model generally shows good performance in low density traffic, but under dense traffic near on-ramps and/or in case of lane drops its accuracy decreases. Numerous extensions of the *Payne* model are proposed in order to improve its accuracy in these traffic situations. Most of these extensions are developed in the direction of relaxation which represents the traffic flow tendency to adjust speeds due to influence of on- and off-ramp flows. The latest *Payne* extended models use an anticipation term, which represents driver reactions to downstream traffic conditions [25]. It is possible to conclude that macroscopic models are not computationally demanding due to a relatively simple mathematical model at their core, but on the other hand, their simulation accuracy is inferior compared to the other traffic models.

3.1.2.1. CTMSIM

The CTMSIM is a macroscopic road traffic simulator for the MATLAB environment. It is primarily used for analysing traffic flows characteristic for motorways. That means that the CTMSIM contains a collection of ".m", ".fig" and ".mat" MATLAB files that can be altered in order to fulfil different simulation demands regarding traffic flows on the motorway. Each segment of a motorway is represented as a cell, which can have one or more on- and/or off-ramps. The simulator performs a traffic simulation using a cellular approach, which is based on the traffic parameters such as traffic demand, capacity, critical and jam density, etc. [34], [35]. The CTMSIM is based on the Asymmetric Cell Transmission Model (ACTM) and it allows user-pluggable on-ramp flow rate and on-ramp queue length controllers. The ACTM model can be seen as a first order approximation of traffic flows on a motorway system. On-ramp flow controllers are based on a collection of standard ramp metering algorithms, which are already implemented in the CTMSIM, [35]. Furthermore, the CTMSIM environment allows users to build their own ramp metering algorithms. Simulation results can be directly compared with Caltrans - Performance Measurement System (PeMS) data, [36]. The CTMSIM interface can operate in a graphical (interactive) mode and command line (batch) mode [37].

The CTMSIM is an open source simulator and is utilized by using the MATLAB script programming language. It also contains certain toolboxes embedded into the MATLAB framework. Because of the mentioned the CTMSIM design structure, it is possible to build and simulate new ramp metering algorithms by using different MATLAB toolboxes, which can be easily integrated with the CTMSIM. The various MATLAB toolboxes that support machine learning, fuzzy logic, neural networks, ontology and evolutionary computing make this simulator suitable for the development of advanced ramp metering algorithms [37].

According to its developers, the CTMSIM simulator has two major components. Both components can be called by the user in the form of a MATLAB function in the MATLAB command environment [34]. The first component is the "Freeway configuration editor". It is used in order to build a motorway configuration from scratch or to edit an existing one. Thereby an existing motorway configuration is stored as a MATLAB variables ".mat" file. The graphical interface for the "Freeway configuration editor" is presented in Figure 7.

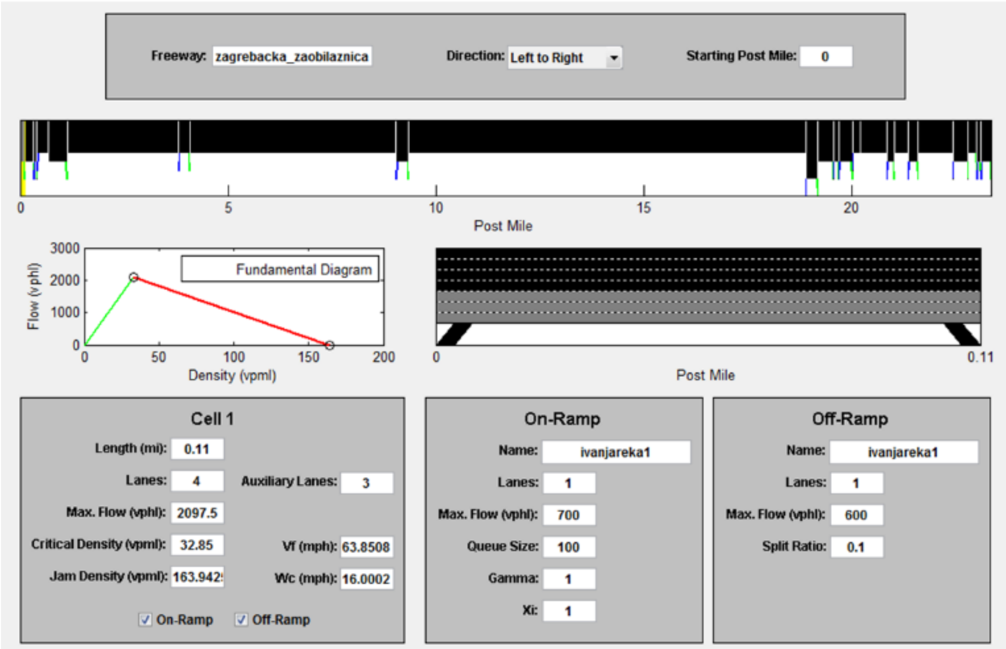

Figure 7: Graphical interface of the “Freeway configuration editor” [37]

The second major component of the CTMSIM is its simulation interface itself - the graphical interface for the CTMSIM. The simulator is presented in Figure 8.

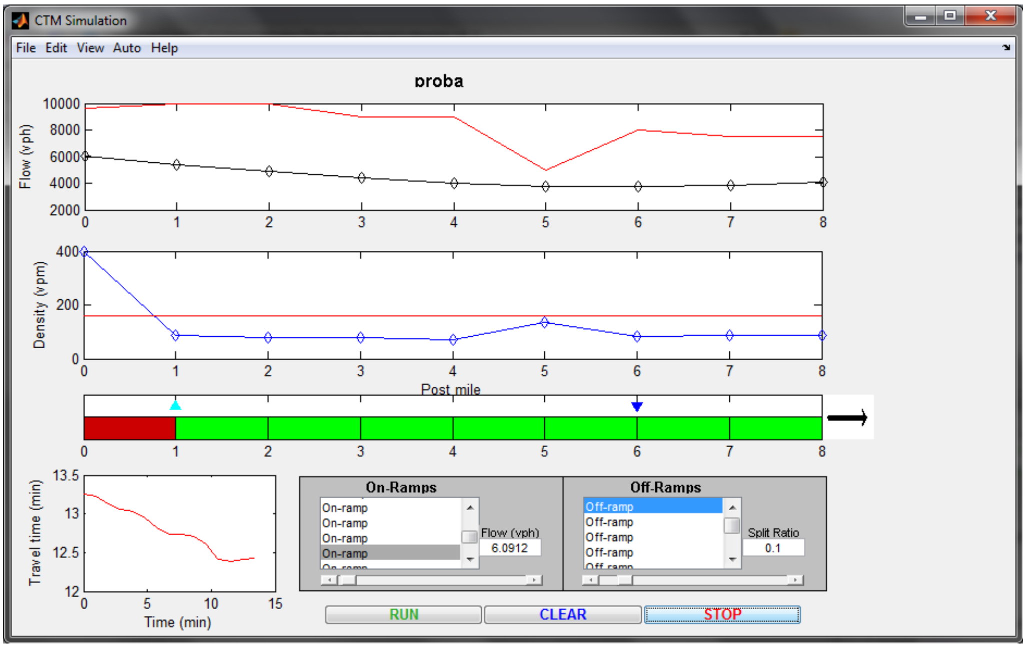

Figure 8: Example of the graphical interface for the CTMSIM Simulator [37]

The ACTM used in CTMSIM is explained in details in [34], [38] and further on basic equations are given. Simulation time is divided into K intervals with length Δt. In Figure 9, it is possible to see an example of a basic ACTM architecture applied on three cells.

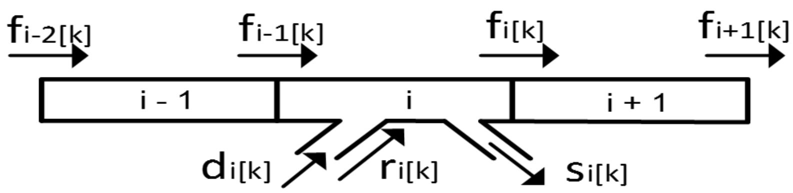

Figure 9: Basic ACTM architecture [38]

The variable $f_{i[k]}$ is the number of vehicles moving from cell *i* to cell *i+1* (or mainstream flow) during the time interval *k* and it can be obtained according to Eqs. (6) and (7), $r_{i[k]}$ is the number of vehicles entering the cell *i*, from its associated on-ramp at time step $k\Delta t$ which is computed according to Eq. (11), $d_{i[k]}$ represents the demand for on-ramp in cell *i*, $s_{i[k]}$ is the off-ramp flow in cell *i* during the time interval *k* which is described with the Eq. (12). The following two equations explain how value $f_{i[k]}$ is computed [38], [39], [40]:

$$f_{i[k]} = \left\{ \begin{matrix} min\, v_i \times \left(1-\beta_{i[k]}\right)\left(n_{i[k]} + \gamma \times r_{i[k]}\right), \\ w_{i+1}\left(\bar{n}_{i+1} - n_{i+1} - \gamma \times r_{i+1[k]}\right), F_{i[k]} \end{matrix} \right\}, \tag{6}$$

$$F_{i[k]} \triangleq min \left\{ \begin{matrix} \overline{f}_i, \\ \dfrac{1-\beta_{i[k]}}{\beta_{i[k]}} \overline{S}_i \end{matrix} \right\}, \tag{7}$$

where $\beta_{i[k]}$ is a split ratio for the off-ramp flow of a particular off-ramp, and γ is the on-ramp flow blending coefficient, both are from the interval [0, 1]. The blending coefficients define the amount of traffic flow, which is added or separated from mainstream traffic flow right before its value is computed [38], [39], [40], [41]. Furthermore, v_i is normalized free flow speed, $n_{i[k]}$ is a number of vehicles (or mainstream density) in the cell *i* at time step $k\Delta t$, while w_{i+1} is the normalized congestion speed in cell *i+1*. $F_{i[k]}$ is the congested flow which leaves cell *i*, $\overline{f_i}$ is the mainline capacity of cell *i* while $\bar{s}_i$ is off-ramp capacity in cell *i*.

On-ramp flow values are now determined with constraints given by Eqs. (8), (9) and (10), where $l_{i[k]}$ is the number of vehicles queuing at the on-ramp in cell *i* at time $k\Delta t$, $c_{i[k]}$ is the value of the metering rate computed by the chosen ramp metering algorithm, while ϑ_i is the on-ramp allocator parameter.

$$r_{i[k]} \leq l_{i[k]} + d_{i[k]} \quad (8)$$

$$r_{i[k]} \leq \vartheta_i(\bar{n}_i - n_{i[k]}) \quad (9)$$

$$r_{i[k]} \leq c_{i[k]} \quad (10)$$

A number of the vehicles that can be merged with a mainstream from an on-ramp in the cell *i*, during the time interval *k* (on-ramp flow if the form of a metering rate), is obtained by Eq. (11).

$$r_{i[k]} = \min\{l_{i[k]} + d_{i[k]}, \vartheta_i\left(\bar{n}_i - n_{i[k]}\right), c_{i[k]}\} \quad (11)$$

Computation of the number of vehicles leaving cell *I* by using an off-ramp during the time interval *k* (off-ramp flow) is described with Eq. (12).

$$s_{i[k]} = \frac{\beta_{i[k]}}{1 - \beta_{i[k]}} f_{i[k]} \quad (12)$$

The number of vehicles in the cell *i* during the interval *k+1* (mainstream density) can be computed from the *mainstream conservation law* given with Eq. (13).

$$n_{i[k+1]} = n_{i[k]} + f_{i-1[k]} + r_{i[k]} - f_{i[k]} + s_{i[k]} \quad (13)$$

Mainstream speed in the cell *i* is obtained according to the Eq. (14), where v_i^{ff} is the free flow speed value for cell *i*, and L_i is the length of cell *i*.

$$v_i^c = \min\left(\frac{f_{i[k]}/(1 - \beta_{i[k]})}{n_{i[k]} + \gamma \times r_{i[k]}} \left(\frac{L_i}{\Delta t}\right), v_i^{ff}\right) \quad (14)$$

3.1.2.1.1. Augmentation for VSLC

One of the CTMSIM augmentations done in this thesis involves the implementation of VSLC for every cell in the simulation model. VSLC is implemented through the modification of a cell mean speed equation given in Eqs. (15) and (16), where v_i^{SLC} is the current VSLC value for the i^{th} cell. The VSLC value must be lower than the free flow speed value of the current cell [37].

The original CTMSIM GUI traffic fundamental diagram is modified to include the option of defining VSLC parameters as presented in Figure 10.

$$v_i^{fd} = \frac{f_{i[k]}/(1-\beta_{i[k]})}{n_{i[k]}+\gamma r_{i[k]}} \left(\frac{L_i}{\Delta t}\right) \tag{15}$$

$$v_i^c = min\left(v_i^{SLC},\quad v_i^{fd},\quad v_i^{ff}\right) \tag{16}$$

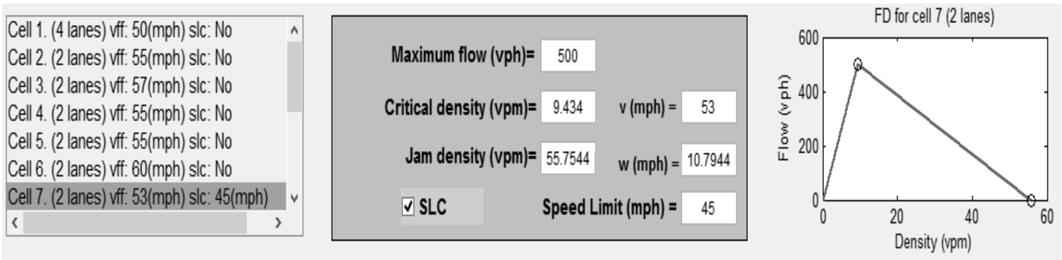

Figure 10: Modification to the fundamental diagram GUI to include the VSLC option [38]

It is important to emphasize that all drivers do not comply with the speed limit imposed by the VSLC. One of the solutions for the mentioned problem is the application of Intelligent Speed Adaptation (ISA). This is a system which uses an on-board unit in the vehicle in order to inform the driver to reduce the vehicle's speed or it can automatically reduce vehicle speed if it is higher than the one imposed by enabling automated driving control. The additional modification has to be included in the CTMSIM VSLC module in order to simulate various penetration levels of the vehicles equipped with the ISA. Such an analysis exceeds the scope of this thesis, but preliminary results regarding this issue can be found in [37].

3.1.2.1.2. Augmentation for cooperative control

In its original version the CTMSIM does not support cooperative control in the form of its technical definition. Cooperative system is defined as a system, which involves multiple dynamic (control) entities that share information or tasks in order to accomplish a common, though perhaps not singular, objective [5]. Detailed insight in cooperative approaches applied in general technical systems and specifically in urban motorway systems will be provided in Chapter 4.

The original version of CTMSIM is augmented in order to provide the effect of cooperation between on-ramps and direct cooperation between the VSLC and ramp metering. The effect of cooperation between on-ramps is enabled by adding an additional (augmented) simulation sequence. As it can be seen in the bottom part of Figure 11 the original CTMSIM simulation sequence runs only through defined cells in a particular time step. The proposed augmentation adds an additional simulation step is at end of each time step. It is computed after all traffic parameters for every cell of the motorway model are computed. This additional simulation step provides an access to traffic data from all cells and stores them in a single data storage variable. At this point, all cells with on-ramps have an access to this data storage. This action enable data exchange between all on-ramps by accessing the same data storage. Based on this data it is possible to design a cooperative control method that will adjust on-ramp rates of all on-ramps based on the overall traffic situation on the motorway model. The cooperative control method is placed into this one location, but it effects each on-ramp since it has access to all the available data, just as it would in the case that this cooperative logic were executed in each individual on-ramp control entity.

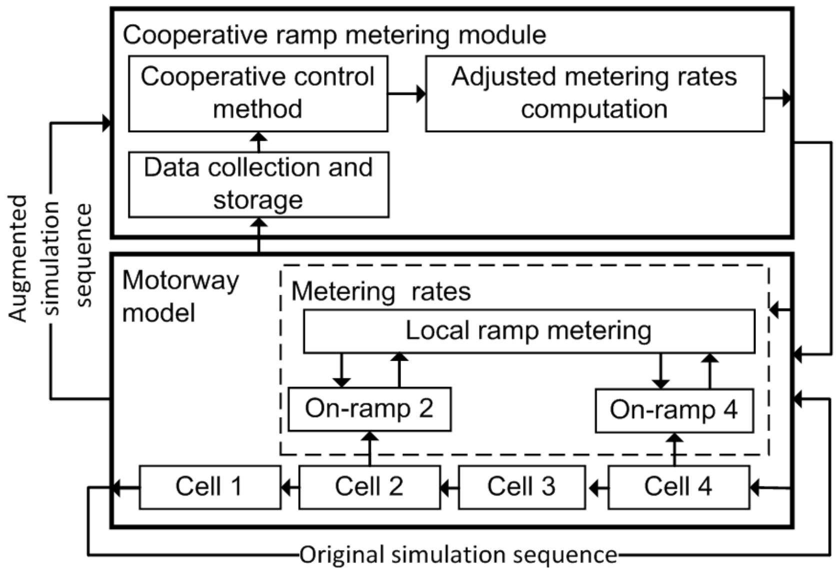

Figure 11: Augmented CTMSIM simulation structure for effect of cooperation between on-ramps [38]

Ramp metering algorithms based on cooperation operate in two phases. In the first phase, the metering rate for each on-ramp is computed by local ramp metering algorithms. Furthermore, in the second phase, additional adjustment of each local on-ramp metering rate is done based on system-wide information about the traffic situation on the whole motorway segment. The HELPER algorithm is one among the first algorithms, which has used the mentioned cooperative ramp metering working principle [37]. It creates virtual queues in upstream on-ramps in order to reduce queue length on congested ones. Communication (data exchange) between on-ramps at a particular motorway segment is the crucial property for the implementation of the HELPER algorithm [37].

In order to enable cooperation between two different motorway control methods, it is necessary to use a direct communication or data exchange between them in order to enable cooperation. The first step is to compute local control variables (speed limits, metering rates) based on the traffic data from a particular cell during the original simulation sequence. The second step involves the exchange of data between traffic control entities. In this case, VSLC algorithm and ramp metering algorithm are considered as the traffic control entities with the ability to exchange information. Furthermore, if it is possible to compute speed limit and metering rate for a particular cell, they will be firstly computed by local logic and then they will be adjusted by the cooperative control module in the VSLC and ramp metering algorithm. After the whole process is done, execution of the control logic will be repeated for all cells in the motorway model with enabled VSLC and if on-ramp exists in those cells. In Figure 12, it is possible to see an illustration of the direct cooperation between the VSLC and ramp metering.

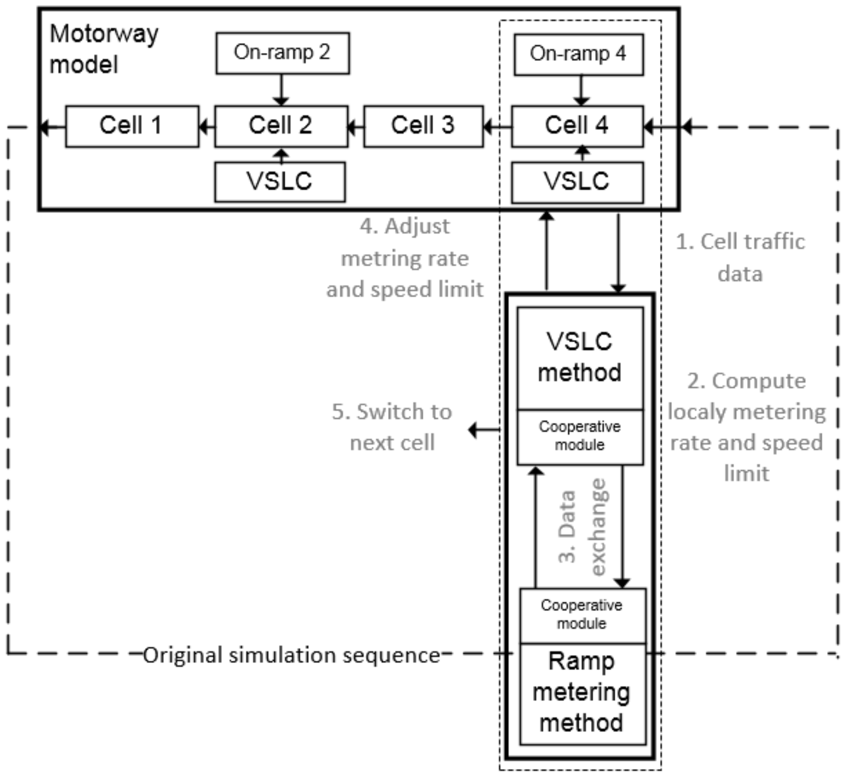

Figure 12: Illustration of the direct cooperation between VSLC and ramp metering

In order to implement cooperation between ramp metering and the VSLC, two specialized variables are added. The first contains a set of data generated by the ramp metering algorithm and is passed to the VSLC algorithm. The second variable contains a set of data generated by the VSLC. This set of data will be delivered to the ramp metering algorithm.

In order to explain the position of these two variables, it is necessary to describe briefly the basic structure of the CTMSIM relevant for these variables. The function which enables automatic control over the on-ramps and the VSLC algorithm is called by a higher-level function responsible for the control of the main simulator window GUI. It is important to mention that the VSLC algorithm is incorporated in the simulation step function responsible for the computation of speed, density, and flow for each simulation step. Thus the VSLC impacts computed speeds and their effect on other traffic parameters directly. This structure of the CTMSIM is the reason why the two variables, which enable direct cooperation, are implemented in data storage generated by the CTMSIM GUI. The positions of the variables that enable cooperation in the CTMSIM software structure are graphically presented in Figure 13.

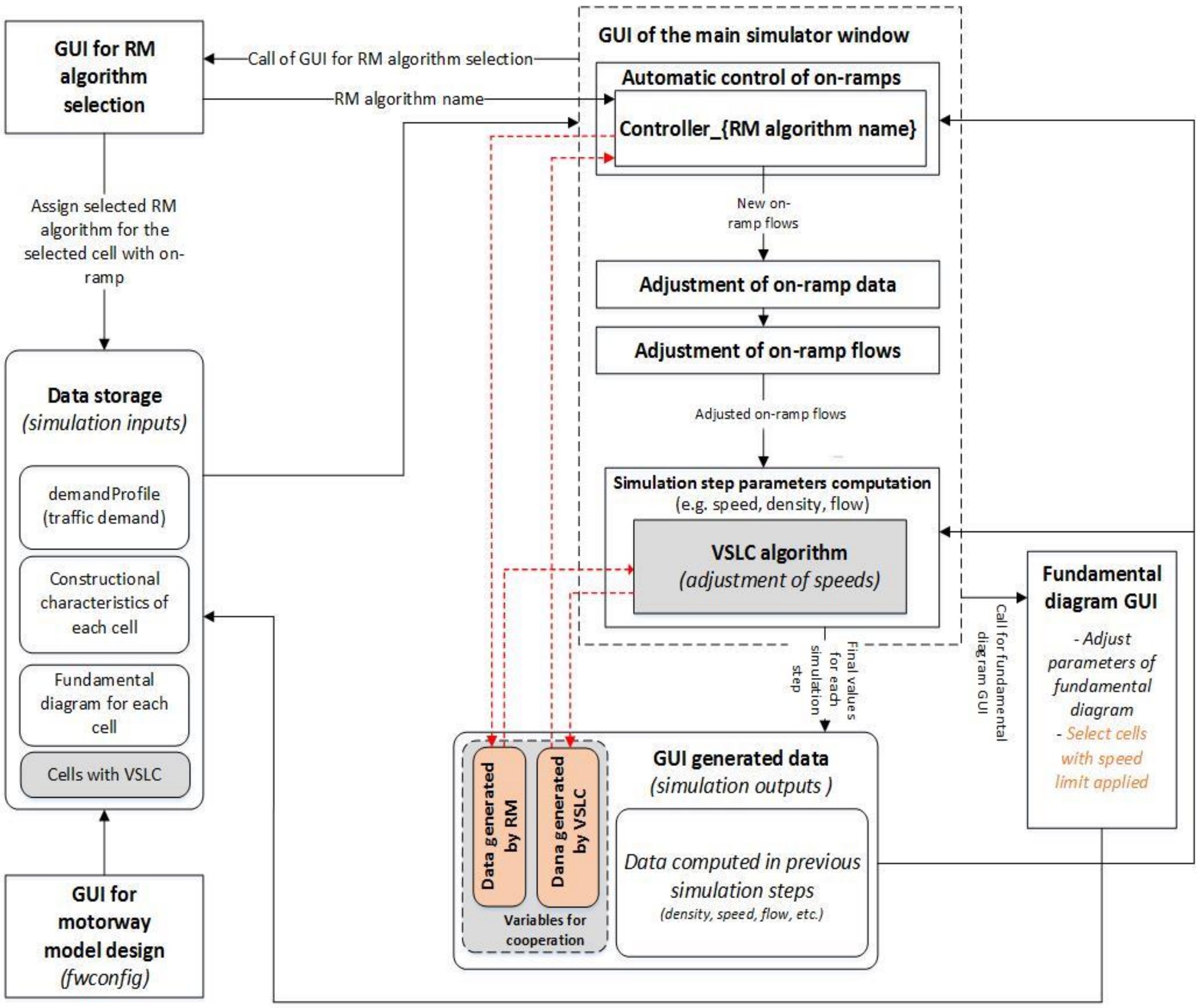

Figure 13: Position of variables that enable cooperation in CTMSIM software structure

3.1.2.2. Other macroscopic simulators

METANET is macroscopic simulator with a similar model structure as the CTMSIM simulator. Motorway network is represented by a directed graph consisting of links and nodes [1]. METANET contains six different types of motorway links what makes this simulator more accurate and consequently more computationally demanding in comparison with the CTMSIM. A normal motorway link provides a second-order discretization of traffic flow (mainstream) without the influence of on- and off-ramp traffic flows. This model is suitable for free flow, critical and congested traffic conditions. Origin links are used for receiving traffic demand and forwarding it into motorway mainstream. It is primarily used for motorway sections with one on-ramp since it contains a simple queue model. The Store-and-Forward link is used for a number of reasons, such as motorway-to-motorway control, simplified consideration of non-motorway routes with limited capacity, modelling the impact of queue spillback on the traffic flow on upstream links, etc. It is possible to implement the same simple queuing model as mentioned in the previous type of link. Traffic conditions in a destination link are influenced by the downstream traffic condition, which may be provided as a boundary condition for the entire simulation horizon [1]. Dummy links are auxiliary links modelled with zero length and they do not effect traffic dynamics.

It can be concluded that each link has uniform characteristics, i.e. no on-ramps or off-ramps and no major changes in geometry. Where a major change occurs in the characteristics of a motorway stretch or in road geometry (e.g., on-ramp or off-ramp), a node is placed, [42]. Traffic enters a node *n*, through a number of input links and is distributed to the output links according to the following equations:

$$Q_n(k) = \sum_{\mu \in I_n} q_\mu, N_\mu(k), \tag{15}$$

$$q_{m,0}(k) = \beta_n^m(k) Q_n(k) \; \forall \, m \in O_n, \tag{16}$$

where I_n is the set of links entering node *n*, O_n is the set of links leaving node *n*, $Q_n(k)$ is the total traffic volume entering node *n* at period *k*, $q_{m,0}(k)$ is the traffic volume that leaves *n* via outlink *m*, and $\beta_n^m(k)$ is the portion of $Q_n(k)$ that leaves node *n* through link *m*. Thus, $\beta_n^m(k), m \in O_n$ are the turning rates of node *n*.

MASTER is a macroscopic simulator based on the gas-kinetic (Boltzmann-like) model that was systematically derived from a “microscopic” description of driver vehicle behaviour and non-local traffic model. Gas-kinetic traffic equation for the so-called phase-space density is the following:

$$f(x, v, t) = \rho(x, t) d(v; x, t), \tag{16}$$

which describes the spatial vehicle density *ρ(x,t)* at location *x* and time *t*, multiplied by the spatiotemporal distribution *d(v;x,t)* of individual vehicle velocities *v.* The Gas-kinetic Boltzmann-like model (arising from his approach to acceleration) obeys the so-called continuity equation when dealing with the spatiotemporal evolution of the phase-space density:

$$\frac{\partial f}{\partial t} = -\frac{\partial}{\partial x}\left(f\frac{dx}{dt}\right) - \frac{\partial}{\partial v}\left(f\frac{dv}{dt}\right), \tag{17}$$

and describes the conservation of the number of vehicles in the absence of on- and off-ramps. The mentioned gas-kinetic equation (17) allows us to systematically derive the related macroscopic traffic equations. The corresponding partial differential equations for vehicle density and average velocity are directly related to the quantities which are characterizing individual driver-vehicle behaviour [43]. The simulator enables fast and robust numerical integration so that several thousand motorway kilometers can be simulated in real-time. It turns out that the model does not conflict with the experimentally observed properties of motorway traffic flow. It actually reproduces the characteristic outflow and dissolution velocity of traffic jams, as well as the phase transition to "synchronized" congested traffic. MASTER also generalizes macroscopic equations for multi-lane and multi-user class traffic [43].

3.1.3. Mesoscopic traffic simulators

Mesoscopic models represent a compromise between the accuracy of the microscopic model and the computational efficiency of the macroscopic model. These models are often used in the case when a real-time simulation with a high level of detail is needed. Most of these models are based on the extended Gas-Kinetic models [27].

Some simulators such as the VISSIM and Aimsun contain an additional mesoscopic modelling level. That additional level enables to dynamically assign mid-sized traffic networks with simulation settings accelerated by a factor of +/-50 compared to microscopic simulation (in the case of VISSIM). In the same time, it is possible to study the effects of traffic light signals on travel times. Shorter computing times are one of the major benefits of mesoscopic models. Simulation models created by using VISSIM on a mesoscopic level have the ability to simulate larger networks at higher speeds. Compared to microscopic models, mesoscopic models have a lower level of detail. This reduction in depth of detail significantly decreases the effort involved in modelling and makes it more efficient to work with [44].

Since the mesoscopic based simulation is a kind of bridge between the microscopic and macroscopic simulations, there are a lot of examples of simulators which use hybrid models based on the microscopic and mesoscopic models. The VISSIM has the ability to select the depth of detail, which means that it is possible to combine mesoscopic and microscopic simulations to produce a hybrid simulation. For example, if users need to get highly detailed traffic parameters at specific corridors or nodes, they can define sections of the mesoscopic simulation in which all modes of transport and their interactions will be simulated at a microscopic level. This gives VISSIM users a tool that allows them to select the level of detail they need for their specific application [44].

DYNASMART (Dynamic Network Assignment-Simulation Model for Advanced Roadway Telematics) is based on a discrete time mesoscopic simulation model. It is designed to model traffic patterns and evaluate the overall network performance in real-time information systems. This simulator combines (1) dynamic network assignment models, used primarily in conjunction with demand forecasting procedures for planning applications, and (2) traffic simulation models, used primarily for traffic operational studies [45], [46].

DYNASMART was specifically developed for the study of the effectiveness of alternative information-supplying strategies, as well as alternative information/control system configurations. It is effective in the macroscopic modelling of traffic flow dynamics such as congestion formation and shock wave propagation. This simulator uses macroscopic parameters such as traffic speed and several traffic flow equations which enable modelling of link travel times on a network level of detail (e.g. effective path of vehicle platoon travel times) [46]. Individual drivers' location tracking is based on microscopic models [47].

The input data for each simulator depends on the type of the traffic network that is analysed and the level of detail required by the user. The complexity of the network can range from a linear motorway system to an integrated urban network with High-Occupancy Vehicle (HOV) lanes, High-Occupancy Toll (HOT) lanes, ramp metering, transit services, incident scenarios and signal controlled intersections on adjacent streets. Application to date has includes metropolitan and regional networks with up to 35,000 nodes and 100,000 links, with nearly one million vehicles simulated over simulation horizons of several hours [47].

3.2. Control methods for urban motorways

The most common control methods for urban motorways are ramp metering, the VSLC, and PLUS. All these control methods are described in the current European ITS deployment strategy as apart of traffic management services [48], hence Croatia's ITS development strategy [49] relies on the implementation of new motorway control methods and services. This is one of the key motivating factors for the writing of this thesis. Each of the mentioned control methods is applied for specific purposes. Ramp metering is primarily used for controlling on-ramp flow rates, while the VSLC and PLUS control methods affect mainstream flows.

PLUS is more restrictive compared to the VSLC, since it allows or denies the use of entire mainstream traffic lanes in the case of, e.g. ongoing traffic incidents, roadworks, etc. The VSLC posts the maximum allowed speed at a given time at the VMS's for the mainstream vehicles. This can be problematic due to a high percentage of drivers who do not comply with the posted speed limits. In this thesis, it is assumed that all drivers obey posted speed limits since the focus of the thesis is ramp metering. On the other hand, ramp metering and PLUS do not encounter these problems with driver compliance with posted control actions. The reason for this is that drivers expect higher driving speeds on motorways and from the experience in the urban traffic network they are used to obeying traffic lights which are used by ramp metering and PLUS (uses specialized VMS traffic lights). In most cases PLUS is managed by motorway operational personnel and is used in extreme situations, while ramp metering and VSLC motorway control methods are managed by various traffic responsive control algorithms and they change control actions more frequently. Additionally, problems with the mentioned motorway control methods are that the VSLC can underperform in the case of low and high traffic demand, while ramp metering and PLUS can create huge on-ramp and mainstream queues if they are inadequately used in the mentioned traffic scenarios. The focus of this section will be describing the key impact of the mentioned traffic control methods on motorway traffic flows. Furthermore, the most widely used ramp metering algorithms and their effects according to their categorization regarding control strategy will be explained.

3.2.1. Ramp metering traffic control approach

Ramp metering as one of the mentioned motorway control methods computes the restriction rate on the total traffic flow which intends to enter a motorway mainstream from a particular on-ramp. This action is conducted by temporarily storing the mentioned traffic flow at on-ramps. This process is known as "access rate reduction." Ramp metering uses road traffic lights and other signals at on-ramps primarily to control the rate or platoon size of entering vehicles. The entire system is based on traffic data collected in real time by road sensors (inductive loops, cameras, etc.) and controllable traffic lights. Sensors are usually placed on the ramps and on the mainstream road. They measure and estimate traffic parameters of the mainstream flow and length of the queue at its on-ramps. A basic ramp metering installation on the motorway is illustrated in Figure 14. Only one local ramp is presented in Figure 14, so it has only one local control unit (server), which can be connected to a higher-level control unit (control server for the whole motorway segment).

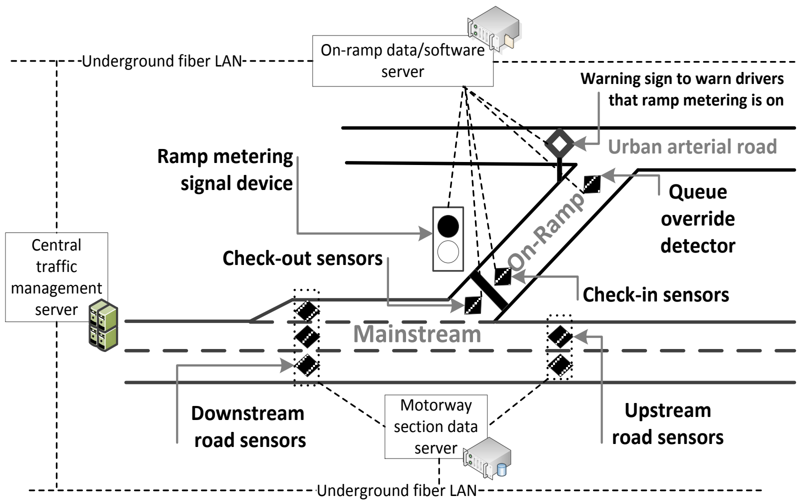

Figure 14: A basic ramp metering installation on a motorway section

Ramp metering can be used for many different purposes. Originally, ramp metering was used as a countermeasure for an increased number of drivers using urban bypasses in order to avoid congestion on urban traffic networks. Ramp metering can increase travel times due to traffic lights at the on-ramps and discourage the use of urban bypasses which serves primarily as a city bypass. In the future, most urban bypasses will evolve into urban motorways so ramp metering will change its role too. This effect of ramp metering is still being taken into account during mobility planning in urban areas. Furthermore, when traffic is dense on a motorway mainstream, ramp metering can prevent a traffic breakdown by adjusting the metering rate so that the density remains below critical values. Besides the aforementioned uses, ramp metering can be used for accomplishing the following effects:

- Reduction of travel time on urban motorways and increased reliability in planning the time required to travel across an urban motorway;
- Prevention of accident and incidents on a motorway;
- Improving environmental protection as a result of reduced noise and rational fuel consumption.

Several field and simulation studies have shown the effectiveness of ramp metering in the mentioned roles [50]. Ramp metering can be based on local (or isolated) and area-wide (or coordinated) control strategies depending on their algorithm working principles [51].

Local strategies consider only the local traffic situation, while area-wide strategies (sometimes refered to as coordinated) consider the overall traffic situation on an entire controlled motorway segment [52]. Some of the literature considers cooperative control strategies as a subcategory of coordinated strategies [53], [4], [50], [51]. The explanation for this categorization can be found in the fact that cooperative control strategies are based on a lower control level compared to the coordinated strategies. Cooperative control strategies are based on information exchange between control entities only. A control entity does not have to establish communication (receive and send data) between all other entities (it can only communicate with control entities near them), which is not the case with coordinated strategies which have one coordination unit governing the behaviour of all control entities or at least receive data from them. This is the reason why coordinated control strategies are in some cases considered as system-wide control strategies and cooperative control strategies their subcategory. Ramp metering algorithms based on cooperative control strategies will be the main focus of this thesis. In this thesis, coordinated, cooperative and integrated strategies will be considered as a subcategory of area-wide control strategies.

Coordinated strategies enable selection between different local control activities to ensure that their global objectives are met by modifications to their original plans [52]. The selection can be conducted using a higher level control module. The module governs the behaviour of all local control activities under particular circumstances [54].

As it was mentioned earlier, cooperative strategies use direct communication between local control entities. In the case of these control strategies, a local control entity also contains the entire logic structure needed for processing of the data exchanged between control entities. The previously computed local decisions are adjusted with respect to the findings of traffic data processing. By cooperating, the local controllers compute an action that can be suboptimal locally, but better for the overall system. A cooperative strategy can also be seen as a subcategory of coordinated strategies which resolves a specific situation with conflicting interests between local control activities. This type of strategy selects a dominant control activity, and all other activities support the dominant one in order to achieve a common goal [54].

Furthermore, there are two other categorizations of ramp metering algorithms: competitive and integrated algorithms. Competitive algorithms execute local and area-wide control ramp metering logic and both of compute an appropriate solution for a current traffic situation. The results of both algorithms types are compared and a final solution is chosen using a specific criteria function or a solution with minimal metering rates. Integrated algorithms are based on the optimization of a specific LoS value while considering constraints such as maximum allowable on-ramp queues, bottleneck capacity, etc. on the entire controlled motorway section [35].

3.2.1.1. Local ramp metering

Local strategies include ramp metering algorithms which take into account only the traffic condition on a particular on-ramp and its nearby motorway segment. The traffic conditions on other on-ramps are not taken into account. The most important local strategies are ALINEA, Demand-Capacity, and Percent-Occupancy.

ALINEA is the French acronym for “Asservissement Lineaire d’entree Autoroutiere” (engl. linear ramp metering control) and is the most widely used local ramp metering algorithm. This ramp metering algorithm offers an optimal ratio of simplicity and efficiency. The main task of ALINEA is to keep the downstream occupancy of the on-ramp at a specified level by adjusting the metering rate. The specified level of downstream occupancy is called the occupancy set-point O_i. Its value is slightly lower or equal to the occupancy at the maximum downstream capacity [18]. The resulting metering rate can be obtained by the following equation:

$$r_i(k) = r_i(k-1) + K_R[O_i - O_i^{out}(k)], \tag{18}$$

where $r_i(k)$ is the current metering rate in cell i, $r_i(k-1)$ is the metering rate from the previous iteration in cell i, K_R is the regulating parameter, and $O_i^{out}(k)$ is the measured downstream occupancy from the previous iteration. The recommended value for K_R is 70 [veh/h] [18]. ALINEA has numerous enhanced versions and is used as part of many other local and coordinated ramp metering approaches. The basic working principle of ALINEA is shown in Figure 15.

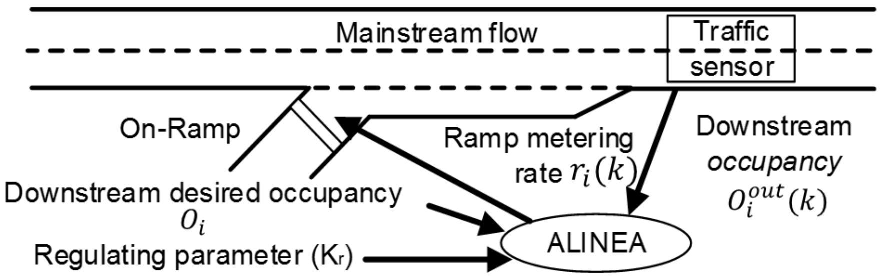

Figure 15: Basic ALINEA working principle scheme [54]

The Demand-Capacity algorithm uses downstream occupancy measurement data. If the downstream occupancy is above a specified critical occupancy ($O_{i\,(crit)}$) in time step k, it is assumed that congestion exists and the metering rate is set to the predefined minimum value. Otherwise, the metering rate is set according to the difference between downstream capacity ($f_{i\,(max)}$) and the measured upstream traffic volume ($f_{i-1}(k)$) in time step k. Mathematical computation of metering rates can be seen in the following equation:

$$r_i(k) = \begin{cases} \max\left(f_{i\,(max)} - f_{i-1}(k), r_{i(min)}\right) & if\ O_i^{out}(k) \leq O_{i\,(crit)} \\ r_{r_{i(\min)}}, & otherwise \end{cases}. \tag{19}$$

The basic working principle of the Demand-Capacity algorithm is shown in Figure 16.

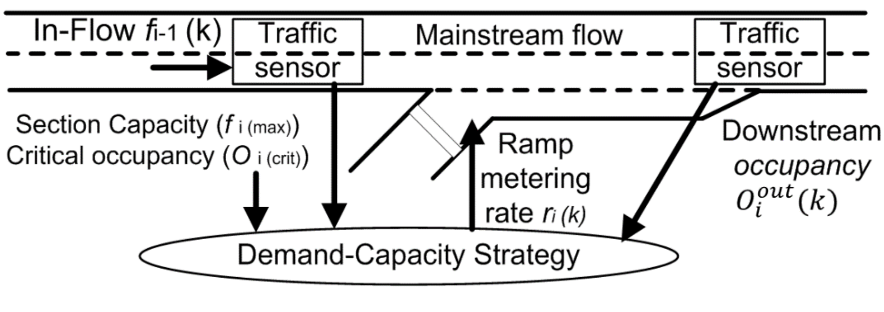

Figure 16: Basic Demand-Capacity algorithm working principle scheme

The Percent-Occupancy algorithm uses two types of constants. The first constant (K_1) is the value of traffic flow at critical density [veh/h] and the second constant (K_2) represent constant based on the slope of a straight line approximation of the uncongested part of the fundamental diagram [veh/h]. Parameter $O_i^{in}(k-1)$ represents the measured upstream occupancy. The metering rate is computed using the following equation:

$$r_i(k) = K_1 - K_2\, O_i^{in}(k-1). \tag{20}$$

The basic working principle of the Percent-Occupancy algorithm is shown in Figure 17.

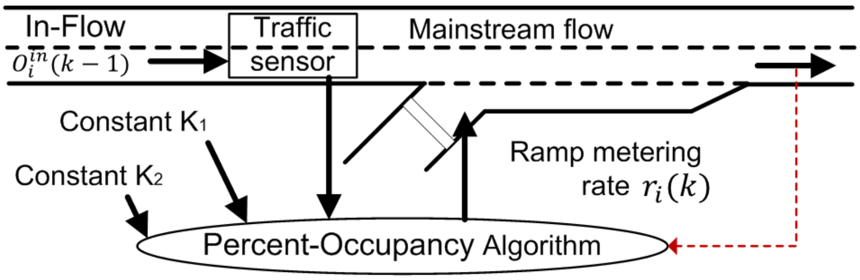

Figure 17: A basic Percent-Occupancy algorithm working principle scheme [54]

The main disadvantage of the Percent-Occupancy ramp metering algorithm when compared to the ALINEA and Demand-Capacity algorithms is its inability to provide a proper reaction at the moment congestion build-up starts. ALINEA can provide a reaction at the start of the congestion build-up due to its closed-loop control structure and appropriate traffic sensor position.

Other local ramp metering algorithms used today are based on artificial neural networks (ANN) and fuzzy logic reasoning. Local algorithms based on the ANN use the learning capabilities of the ANN to produce metering rates for all on-ramps. Learning data sets which will be presented to the ANN are generated using a traffic simulation model of motorway or measured traffic data from a real motorway system. The ANN-based control algorithms provide better results when they are used as part of area-wide ramp metering strategies [55].

3.2.1.2. Area-Wide (or coordinated) ramp metering

Area-wide (or coordinated) control strategies in general involve all algorithms which take into account the overall traffic situation on the controlled motorway segment. Ramp metering algorithms developed on the basis of the Area-Wide control strategy can be further divided into cooperative, competitive and integrated algorithms [53].

3.2.1.2.1. Cooperative algorithms

Cooperative algorithms are applied to several local control entities – on-ramp controllers. Each on-ramp controller can use its local and cooperative control logic. Cooperative control logic has the ability to override locally computed metering rates. Local control actions can be overridden if they do not accomplish the Area-Wide common goal. In order to initiate their cooperative control strategy, the local on-ramp controllers have to exchange information with each other. This information usually includes current locally computed metering rates and locally sensed data. Based on the data received from other local on-ramp controllers, cooperative control logic of each on-ramp controller computes metering rates that are in line with the Area-Wide cooperative strategy [52]. The cooperative algorithm can be considered as an intermediary for resolving specific situations with conflicting interests between local control activities on an Area-Wide level. Firstly, it is imperative to detect the location of a bottleneck and enrol several upstream on-ramps in order to create virtual on-ramp queues. Virtual queues have the primary role to stop forwarding additional traffic flow into the mainstream in order to mitigate upstream congestion [52]. The on-ramp closest to the location of the bottleneck will have a different regime for computing metering rates compared to the upstream on-ramps that are tasked with inducing virtual queues. The typical examples of such algorithms are: HELPER and LINKED [53], [4]. The HELPER algorithm was the first algorithm developed based on cooperation.

The **HELPER** algorithm was one of the first algorithms developed for ramp metering based on the cooperative control strategy. It includes several local traffic responsive metering algorithms that store their inputs (local traffic situation) and outputs (metering rates) data in one place. The data collected is used to create a "big picture" of the traffic situation on the entire controlled motorway. Override control, which can be operated as a centralized unit, adjusts locally computed metering rates based on previously collected data. A particular on-ramp on a motorway section is categorized as "master" if the queue detector at on-ramps or mainstream detector exceeds a pre-determined threshold value. The override module enables the effect of cooperation by increasing the metering rate at the "master" on-ramp by one level and reduces the rate for several upstream on-ramps by one level. These upstream on-ramps are categorized as “slave” on-ramps. The main idea is to exploit the “slave” on-ramps queue capacity in order to mitigate downstream congestion in the mainstream. The working principle of cooperative ramp metering algorithms can be very complex and sensitive considering the fluctuations in traffic parameters. Because of this reasons, it is imperative to conduct simulations in order to analyse the impact of cooperative ramp metering on traffic flows with respect to a specific urban motorway segment [53], [52].

The **LINKED** algorithm is a much more complex algorithm than the HELPER algorithm. It is based on the Proportional-Integral-Plus (PIP) controller type, which is dedicated to various control tasks in general. A detailed insight into the mentioned controller type is presented in [56]. The LINKED algorithm uses the Non-Minimal State Space (NMSS) description of the system which is to be controlled. The NMSS is formulated using states, past value of outputs, past value of inputs, and additional integral-of-error states. A special form of the NMSS description based on the Local Linear Model (LLM) for each point in the motorway system is formulated so as to be applied to a motorway system. Each point of the motorway system where measurement of traffic data can be done is modelled by a special form of the NMSS using the previously sampled measurements at a certain point as well as the upstream and downstream locations compared to that point. This allows the model to handle both congested and uncongested traffic conditions in a spatial and temporal context. The on-ramp flow is used as an additional variable [4]. The LINKED algorithm is based on the demand-capacity concept, and the local metering rate is determined based on upstream flow measurements at each location. Its wide-area functionally is similar to that of the HELPER algorithm. Whenever a ramp’s metering rate is among the lowest among three metering rates, the upstream ramp is required to meter at the same rate or lower, and, if necessary, the ramps further upstream are also required to do so [4].

3.2.1.2.2. Competitive algorithms

Competitive algorithms contain two different control logics: a local and an Area-Wide control logic. During the execution of the ramp metering algorithm, each local control logic provides an appropriate solution for the current traffic situation. The more restrictive metering rate or the one more in line with the criteria function will be chosen as the final one. The typical examples of these algorithms are Bottleneck and SWARM [53], [4], [57].

The **Bottleneck** algorithm has two components that provide two different metering rates. The first component calculates a local metering rate based on occupancy control that selects a ramp metering value from a finite set of discrete predetermined metering rates with respect to the upstream occupancy. The second component calculates a so-called bottleneck metering rate. A particular section is identified as a bottleneck if it satisfies two conditions: the capacity condition and the vehicle storage condition. The bottleneck metering rate is calculated to keep the flow of traffic at a defined bottleneck below capacity [4].

The **System-Wide Adaptive Ramp Metering (SWARM)** algorithm consists of two independent algorithms – SWARM 1 and SWARM 2. SWARM 1 algorithm in its control logic involves short-term predictive component, which uses system-wide information's. SWARM 2 includes two local traffic responsive ramp metering algorithms and has two versions: SWARM 2A and SWARM 2B. SWARM 2A is not implemented in the CTMSIM. In the CTMSIM, SWARM is a combination of SWARM1 and SWARM2B controllers, [35], [57]. The on-ramp flow is chosen as a minimum of the two values – one produced by SWARM 1 and the other one produced by SWARM 2B. SWARM 1 forecasts the traffic state at predetermined bottleneck locations and adjusts metering rates based on the obtained forecasts. It divides the motorway into zones whose boundaries are determined by the bottlenecks in such a way that a bottleneck cell is the last cell of the zone. These zones are enumerated. The controller's zone parameter of the controller determines to which zone the cell with this on-ramp belongs [4].

SWARM 2B is a local traffic responsive ramp metering algorithm. It introduces the concept of a storage zone, stretches from the mainstream upstream vehicle detection station (VDS) to the next downstream mainline VDS. Using the parameters obtained from the VDS, the number of vehicles present in this zone is computed. The VDS collects real-time data on a motorway traffic flow. Each vehicle detection station is typically configured to collect standard traffic flow parameters on a lane-by-lane basis (volume, occupancy, speed), [4]. In the case of the CTMSIM, the data available to the SWARM algorithm obtained from the simulated VDS is: on-ramp demands, flows and queues. The metering rate is set in this manner to keep the number of vehicles below the defined critical value.

3.2.1.2.3. Integrated algorithms

Integrated algorithms contain a control module based on an optimization engine with defined boundaries and a goal that has to be achieved during the control period. The typical examples of these algorithms are METALINE, the FHWA/BALL Space, DYNAMIC, and *fuzzy* logic based algorithms [53]. *Fuzzy* logic based algorithms are the most sophisticated in this group. They can be described as a type of expert ramp metering systems. *Fuzzy* logic based algorithms make decisions by using converted empirical knowledge about traffic flow parameters and ramp control to fashion so-called fuzzy *rules* [54]. The rules contain inputs in logical relations and their impact on a particular traffic parameter is defined as a rule output. *Fuzzy* logic based algorithms are suitable for ramp metering because *fuzzy* logic is ideal for making decisions which are based on inaccurate input data e.g. inexact traffic models and noisy sensor measurements [4].

3.2.2. Variable speed limit control

In the development of the VSLC, two approaches are used in order to avoid congestion. The first increases homogeneity of traffic flow by removing the sources of disturbances, e.g. eliminating the larger speed differences of vehicles in a platoon by lowering the speeds of faster vehicles. A higher difference between vehicle speeds in a platoon can result in braking and other actions such as line changing, etc. From the macroscopic viewpoint it is possible to say that if the speed limits used are above critical speed, i.e. the speed at critical density with the maximum flow, the speed limitation is considered to have a homogenizing effect [58].

The second approach is to increase the stability of traffic flow, which can be achieved by reducing headway between vehicles [58]. In [22] these principles are embodied into two analytically different approaches towards speed limitation: one utilizing the homogenizing effects from decreased speeds, and the other using flow reduction for preventing traffic breakdown or resolving a prevailing jam [58]. The second approach is more suitable for macroscopic simulations and is therefore considered by this thesis.

Application of the VSLC directly changes the corresponding fundamental diagram of a motorway section [59], [60], [61]. The changes of the slopes in the fundamental diagram with Variable Speed Limits (VSL) and without the application of the VSL can be seen in Figure 18.

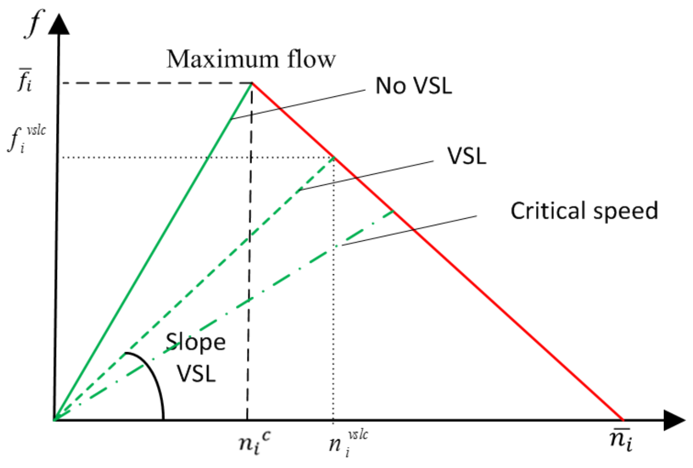

Figure 18: The slope of a traffic flow diagram with and without the VSL [59]

Although this thesis is focused on the improvement of traffic flow performance by using the VSLC motorway control methods, it is necessary to mention its other important impacts such as the minimization of the number of traffic accidents, the reduction of air pollution and road noise. The spatially determined frequency of registered accidents is one of the main parameters which can be used for the selection of the motorway segments where the VSLC can potentially be installed. The positive impact of the VSLC on traffic safety is the result of speed homogenization which decreases need for frequent braking and other manoeuvres which can lead to traffic accidents. Thus, the probability of accidents is reduced. Evaluations of motorways with the VSLC installed show a reduction in the accident number reaching up to 30 % [60].

Studies made in Finland show that a VSLC installed with a primary role of improving traffic flow performance also increases the level of traffic safety in cases of bad weather conditions [60]. These positive effects of the VSLC are primarily brought on by conducting efficient recognition of hazardous weather and road conditions. This recognition is supported by road signs informing drivers about variable slipperiness and moderate use of the highest speed limit. In Finland, speed limits are lowered during winter time on most two-lane roads in this manner. Application of the VSLC enables the option of showing higher speed limits in good conditions what is impossible with fixed signs which permanently reduce speeds in a certain time period [61]. Vehicle emissions are also reduced by using VSLC by 0.40 to 2.85 % depending on the algorithm used and the emission type measured [62].

In order to implement the VSLC for increasing the throughput of a motorway, it is necessary to use adequate traffic response control logic which will compute speed limits posted on the

VMSs. The main problem with the VSLC is driver obedience with the posted speed limits. In [63] a study regarding drivers' obedience with the posted speed limits was conducted. Measurements for this study were performed at Croatian motorway A1 (section Jastrebarsko - D. Zdenčina). Vehicle speeds were measured before and after changing the speed limit value posted on the VMS. Furthermore, vehicle speeds were also measured when the VMS was not used. On the A1 motorway, the current speed limit is set to 130 $[km/h]$. For purposes of the experiment, the speed limit was changed to 100 and 80 $[km/h]$. The resulting speed distribution obtained is given in Table 1.

Table 1: Vehicle speeds distribution on A1 motorway section Jastrebarsko - D. Zdenčina [63]

	Without VMS 130 $[km/h]$	**Speed limit on VMS 100 $[km/h]$**	**Speed limit on VMS 80 $[km/h]$**
Arithmetic mean	147.299	141.867	136.586
Median	148.056	141.618	136.791
Mod	152.884	146.446	140.009
Standard dev.	18.609	18.991	20.032

The VSLC efficiency depends highly on driver's obedience to the posted speed limits. The implementation of systems such as the ISA can increase the efficiency of the VSLC significantly. The ISA can be defined as a supporting system which notifies drivers about the maximum permitted speed. Current ISA systems are largely comprised of three parts: a GPS receiver, a small on-board computer, and a support unit with a display, which shows the posted speed limit and gives a warning signal to the driver if his speed value exceeds the speed limit. Using GPS technology, the ISA system registers a vehicle's speed and compares it to the posted speed limit at the vehicle's current location. The display on the instrument panel continuously shows the posted speed limit for the particular road segment. This speed limit related data is obtained from a road database. The ISA can initiate several warning measures if the speed limit is exceeded. The most basic warning measure is a sound alert; other systems use the accelerator pedal to indicate that the speed limit has been exceeded through counter-pressure or vibration. The concept of the ISA covers a wide range of systems which can be divided into two major classes. The first class is the advisory ISA in which in-vehicle information of the current speed limit is provided for the driver, but the speed is still controlled by the driver as in the situation without the ISA. The second class is a fully automated ISA in which in-vehicle information of the current speed limit is provided for the driver but this information is passed to the vehicle in order to automatically adjust its current speed to the posted speed limit for a current road segment [64].

3.2.3. Prohibiting lane use system

The process of initiating the redirecting of mainstream traffic flow from the right to the middle and left lanes (depending on the number of urban motorway lanes) which are dedicated to faster vehicles can be very important in specific traffic scenarios. This process can be managed by a motorway control method known as PLUS. One of PLUS's goals is to clear the right mainstream lane (the lane closest to the merging lane) from mainstream traffic flow vehicles and enable a quick and safe merging process. Furthermore, in the case of incidents PLUS can close the lane affected by an incident or close a left (faster) lane up to the location of the incident for all types of vehicles except for emergency vehicles. This control method must be supported with appropriate traffic signalization and a driver information system. It can work on its own or in cooperation with a ramp metering system or the VSLC.

If PLUS cooperates with ramp metering it uses the VMS to inform drivers when they need to perform lane changing from the right to the middle or left lanes. The periods when the VMS informs the drivers to initiate lane changing depend on the motorway segment traffic situation, current controlled on-ramp parameters, part of the motorway segment positioned closely to controlled on-ramp and the applied metering rate. For example, if an on-ramp queue is long and traffic density on the motorway segment close to the on-ramp is low, drivers in the mainstream should move into a faster lane. The lane closest to the merging lane will be free of dense traffic and drivers from the on-ramp will be able to merge with the mainstream traffic flow without hesitation. This additional control process increases the safety of the merging process and reduces the possibility for the creation of bottlenecks.

The concept of cooperation between PLUS and ramp metering can be seen in Figure 19. An interconnection between the two systems (ramp metering and lane changes prohibiting systems) can be noticed. The interconnection exchanges crucial traffic information and can be applied to the local or high control level. When applied to the high control level, the corresponding cooperative unit can further improve the LoS of a motorway segment.

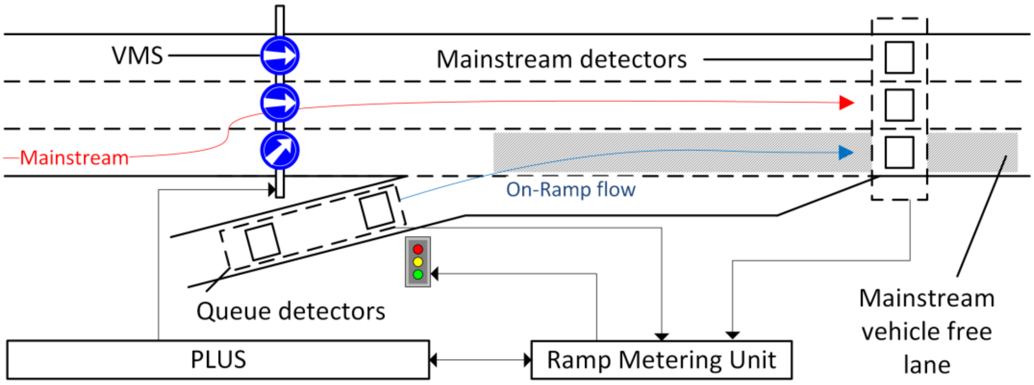

Figure 19: Concept of cooperation between PLUS and ramp metering

The cooperation between PLUS and the VSLC can produce information which will be forwarded to the vehicles equipped with an On-Board-Unit (OBU). Vehicle OBUs can be adjusted to work as a personal advisory information system. The OBU can display information about recommended speeds and lane change to the driver. Information for every vehicle equipped with the OBU can be calculated based on vehicle location, current speed, route selection and desired destination. This way it is possible to establish integration of this information system with vehicle guidance systems. This type of integrated system can provide a comprehensive set of travel information to the driver. The mentioned set of information can contain optimal/mandatory lane use and optimal/mandatory speed which is calculated based on current stats of a vehicle and the regulation for a selected route. Drivers are more likely to follow personal advisory information provided by their own in-vehicle equipment than general VMS messages that are provided for all vehicles on the same road segment. Additionally, with this kind of personal advisory systems it is possible to monitor drivers more scrupulously and consequently control them to a higher degree.

PLUS and the VSLC are usually implemented in motorway systems which are used by a large number of heavy vehicles – mainly trucks [65]. Manoeuvres of trucks such as acceleration, deceleration and lane changes may easily disturb traffic flow and increase overall travel time. Authors in [65] suggest that the lack of improvement on travel time via the use of standalone VSLC is due to numerous lane changes that take place close to a bottleneck. This sort of situation can lead to a severe capacity drop. The combination of a lane change system and the VSLC enables lane changes in advance in order to alleviate the capacity drop. The VSLC additionally decreases the speed of traffic flow heading in the direction of the bottleneck. The concept of combining VSLC and PLUS can be seen in Figure 20.

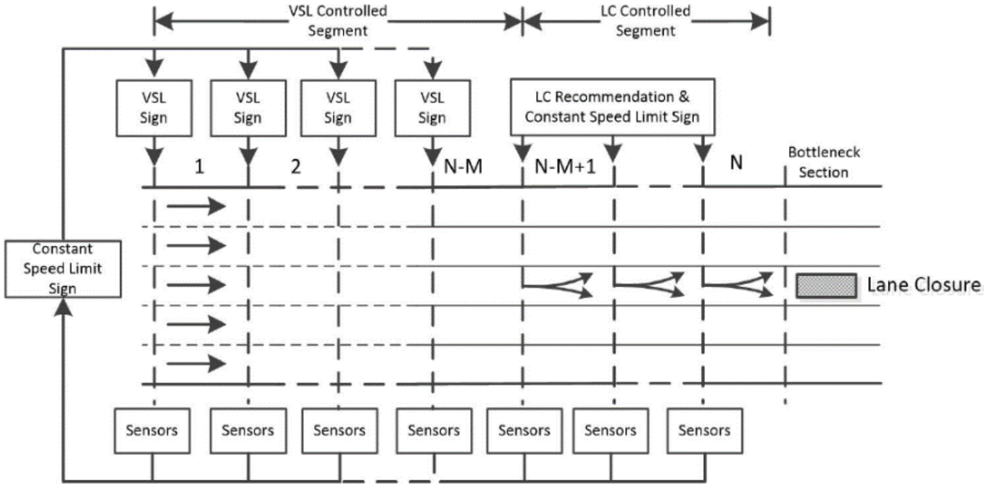

Figure 20: Concept of combined PLUS and VLSC [65]

3.3. Applied methodology for machine learning

The usability and the effectiveness of ramp metering and other motorway control methods significantly depends on their ability to react to unforeseen traffic scenarios such as incidents, vehicle breakdown and rapid changes in traffic demand within a short time interval [66]. These traffic scenarios can be recognized if recurrent patterns from traffic flow data are isolated and their behaviour formalized in some form of a knowledge base. On the basis of recurrent traffic flow patterns it is possible to spot deviations and even identify their patterns. Contemporary mathematical models cannot cope with these challenges efficiently. To deal with these challenges, new approaches based on machine learning are used.

Machine learning is conducted through different types of algorithms that can adjust parameters according to a given dataset and produce predictions, classification, etc. Such algorithms overcome the need for explicit decision making programming since they decide or predict on the basis of data-driven model building from sample inputs [67]. Approaches based on machine learning conduct generalization of data in a given dataset – the learning dataset. It is possible to conclude that this learning dataset can be considered a representative of the space of occurrences in the mentioned system. Based on these representatives it is possible to create a knowledge base of the system model by the end of the learning process and eventually develop an adequate control method. After the learning process, it is possible to perform accurately on a new set of examples/tasks from the same system.

The machine learning approach is usually applied to systems with complex stochastic behaviour which can be described by a generally unknown probability distribution. The mentioned approach is feasible for implementation in the domain of urban motorways since behaviour and interactions of a traffic flow on the urban motorway exibit a stochastic nature. If all the data provided in the learning dataset can cover a sufficient number of different recurrent and non-recurrent traffic scenarios it is possible to create a sufficiently accurate model of urban motorway traffic behaviour. Based on the knowledge of model patterns it is possible to derive a motorway control method.

The most widely used approaches in the domain of machine learning are based on fuzzy logic, the ANN, genetic algorithms and hybrids of these approaches. In this chapter, the *Fuzzy* Inference System (FIS) based on *fuzzy* logic, the ANN and their combination will be described in more detail.

3.3.1. Fuzzy inference system

Rule-based algorithms are widely used in motorway control methods (ramp metering and VSLC) because they are easily understood and applied [68]. Control actions in classic rule-based algorithms are determined based on pre-specified rules. The main problem is that most algorithms which are proposed in the literature under this category are too crude for controling motorway traffic. They are crude because they use crisp sets based on which control rules are created. Crisp sets cannot adequately represent traffic situations on the motorway system due to the stochastic non-linear and non-stationary nature of traffic flows [61], [69].

Fuzzy sets used by *fuzzy* logic-based motorway control methods enable separation of attribute domains into several overlapping intervals [69]. The discretization using *fuzzy* sets can help overcome the sensitivity problem caused by crisp discretization used in the existing motorway control algorithms [68]. The *Fuzzy* Inference System (FIS) is the most widespread *fuzzy* logic-based control approach. It can be considered as an expert system based on a reasoning system with the rule-base designed using *fuzzy* sets.

Input–output relations are defined by a set of the mentioned *fuzzy* control rules, e.g., IF–THEN rules that represent knowledge and experience of an expert that controls processes in a particular system [70]. The *fuzzy* logic-reasoning system contains two distinctive data sets. The first type of data set contains labels and parameters of membership functions assigned to input and output variables. The accurate selection of these represents is one of the most critical stages in the design of the *fuzzy* logic-reasoning system [71]. The other set of data is related to the rule-base that processes *fuzzy* values of the inputs to fuzzy values of the outputs [72].

A generic FIS architecture is composed of four parts. The first, the *fuzzification* part is tasked with converting crisp input values into linguistic variables according to adequate membership functions. The parameters of all membership functions are previously stored in the database as one part of the FIS knowledge base. The FIS knowledge base can be considered as a specific part containing a rule base and a database according to which the remaining three parts are provided with data. The inference engine is the core part of the FIS with the main task of evaluating the input's degree of membership to *fuzzy* output sets. It uses *fuzzy* rules which are stored in the rule base [71]. Finally, the de*fuzzification* block transforms the *fuzzy* input in this block into a crisp value which represents the final control action or decision. The generic architecture of the FIS is shown in Figure 21 [73].

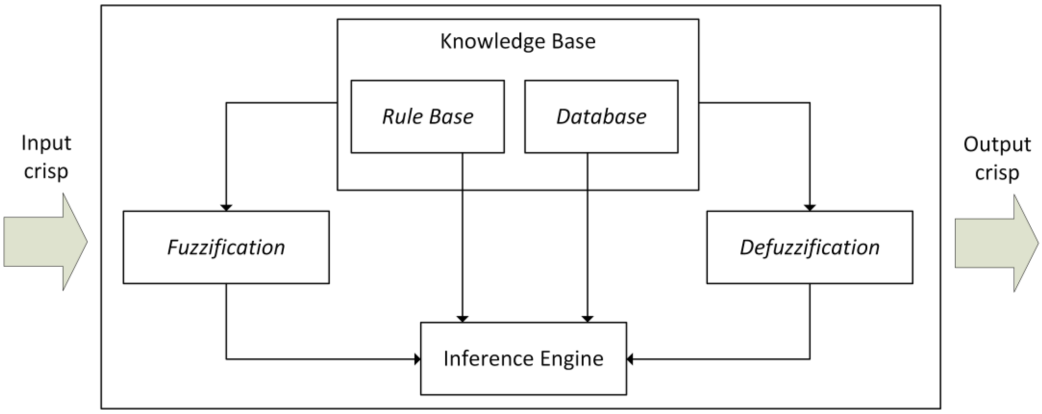

Figure 21: Generic architecture of FIS [71]

The FIS makes decisions based on the following five steps:

- The first step is to take the inputs and determine the degree to which they belong to each of the appropriate *fuzzy* sets via membership functions. The inputs are usually a crisp numerical value limited to the universe of discourse of the input variable and the output is a fuzzy degree of membership in the qualifying linguistic set. The process is known as ***fuzzification;***
- The second step is initiated after the inputs are *fuzzified* and involves the **application of *fuzzy* operators**. If the premise of a given rule has more than one part, the *fuzzy* operator is applied to obtain one value that represents the result of the antecedent for that rule

[74]. The output is a single *bool* value. The commonly used AND methods are the minimum and product, while the OR methods are the maximum and the probabilistic OR method (also known as the algebraic sum).

- The third step involves the application of the **implication method**. This method can be implemented after proper weighting is assigned to each rule. The consequent is a *fuzzy* set represented by a membership function, whose weights appropriately describe the linguistic characteristics that are attributed to it. The consequent is reshaped using a function associated with the premise (a single value). The input for the implication process is a single value given by the premise, and the output is a *fuzzy* set. The implication is implemented for each rule [74]. Two implication methods are common, and they are the same functions that are used by the AND method: minimum, which truncates the output *fuzzy* set, and product, which scales the output *fuzzy* set.
- The fourth step is to **aggregate all outputs.** Aggregation is the process by which the *fuzzy* sets that represent the outputs of each rule are combined into a single *fuzzy* set. Aggregation only occurs once for each output variable [74]. Since the aggregation method should always be commutative it is possible to use the following aggregation method: maximum, the sum of the each rule's output set and probabilistic OR.
- The fifth step requires ***defuzzification*** which transforms *fuzzy* sets into crisp outputs. This is done by deriving one crisp value of the aggregated *fuzzy* set by applying the centroid method, the bisector method, the middle of maximum method, etc.

It is important to mention that there are three types of the FIS. The illustration of the three FIS types is presented in Figure 22.

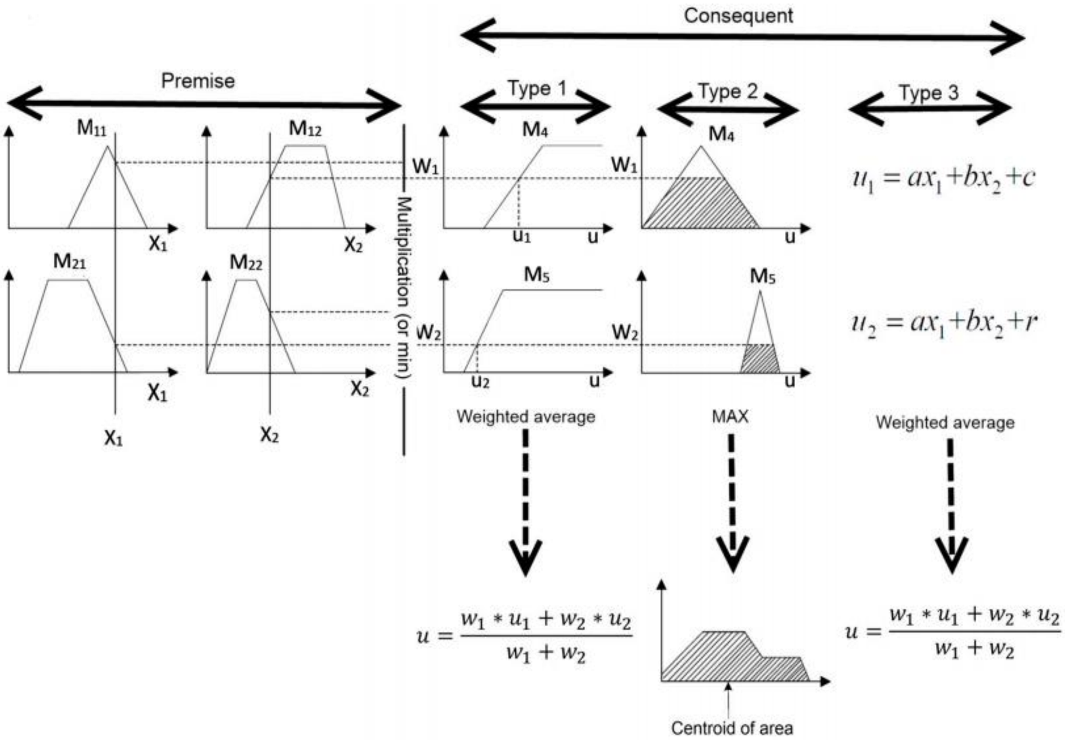

Figure 22: Illustration of three FIS types [71]

In Figure 22, an example of two *fuzzy* rules with the same premise but a different type of consequent parts is presented. M_{11}, M_{12} (first rule), M_{21} and M_{22} (second rule) represent linguistic labels of membership functions in the premise of the *fuzzy* rules. The, implication method of minimum is applied to both rules.

Type one FIS in Figure 22 is a classical *Mamandi* FIS with a weighted average of consequence part of the rule. This type of FIS in the consequent part contains two membership functions, M_4 for the first rule, and M_5 for the second rule. W_1 and W_2 represent degrees of the activation function for the consequent part of the *fuzzy* rule after the implication method is initiated. The aggregation of the consequent part of both rules will be conducted by computing the weighted average according to the equation presented in Figure 22.

Type two FIS is a Mamdani FIS with the output function based on the overall *fuzzy* output (explained in the previously mentioned 5 steps). Aggregation of both rules is conducted by computing the centroid of an area after the maximum function is applied on the membership function in the consequence part of the both *fuzzy* rules.

Type three is the Takagi-Sugeno FIS. The Takagi-Sugeno FIS has *fuzzy* inputs in the premise and a crisp output (linear combination of the inputs) in the consequent part of the *fuzzy* rule. It is computationally efficient and suitable for application in optimization and adaptive techniques. It is possible to conclude that it is very suitable for control problems in dynamic nonlinear systems [68] such as traffic flows on urban motorways. In Figure 22, x_1 and x_2 represent the inputs for both rules, while the parameters *a, b, c,* and *r* are the tuning parameters. Each rule is weighted by the firing strength w_i of the rule. For example, if an AND operator is applied in the fuzzy rule, the firing strength of both *fuzzy* rules will be computed as following $w_i = ANDmethod(M_{i1}(x_1), M_{i2}(x_2)),\ i = 1, 2.$ The final output will be computed as the weighted average similar to the type one FIS.

3.3.2. Artificial neural networks

Artificial Neural Networks (ANN) can be considered as statistical learning models. They are inspired by biological neural networks that are used as one form of blueprints for this machine learning approach. Biological neural networks can be found in nature in the form of central nervous systems, such as the brain. The ANN is represented as a system of interconnected “neurons”, which send messages to each other. The connections within the ANN can be systematically adjusted based on inputs and outputs, making them ideal for supervised learning. In Figure 23, it is possible to see a simple neuron model used in a perceptron ANN.

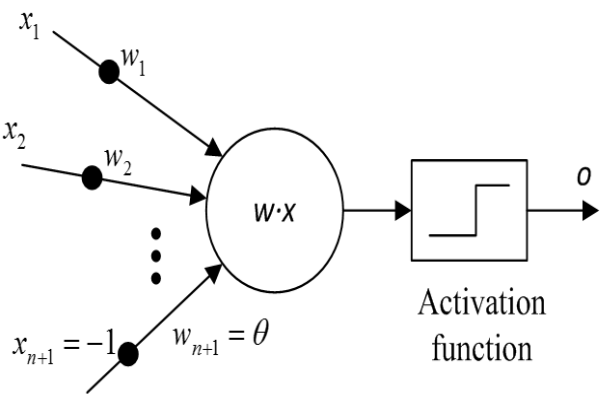

Figure 23: Illustration of a simple neuron model used in a Perceptron ANN

The perceptron is a simple neuron based model that takes input signals coded as input vectors $\bar{x} = (x_1, x_2, \dots, x_{n+1})$ through the associated vector of synaptic weights $\overline{w} = (w_1, w_2, \dots, w_{n+1})$. The output *o* is determined by:

$$o = f(net) = f(\overline{w} \cdot \bar{x}) = f\left(\sum_{j=1}^{n+1} w_j\, x_j\right), \tag{21}$$

where *net* denotes the weighted sum of inputs, and *f* is the activation function. By convention, if there are *n* inputs into the Perceptron, the input *(n+1)* will be fixed to *-1* and the associated weight to θ, which is the value of the excitation threshold.

In order to approximate complex non-linear functions or to learn a variety of association tasks, feed-forward ANN models are used. In feed-forward ANNs, neurons are organized in layers. There are no connections among neurons within the same layer; connections only exist between successive layers. It is important to mention that each neuron from layer *l* has connections to each neuron in layer *l +1*. As has already been mentioned, the activation functions need to be differentiable and are usually of the sigmoid shape [75]. In Figure 24, it is possible to see a one-layer feed-forward ANN.

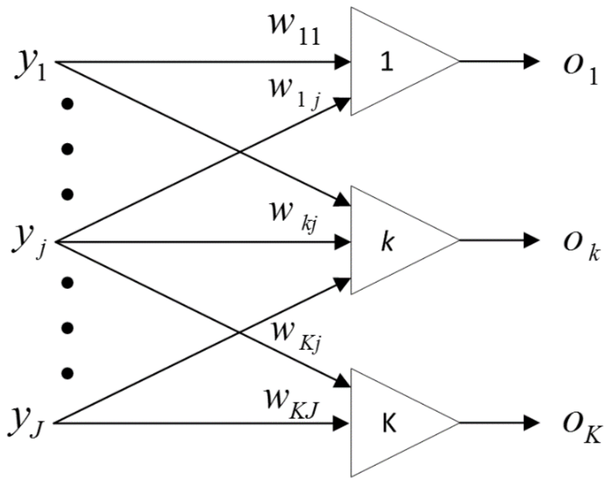

Figure 24: A one-layer feed-forward ANN [75]

Consider the single-layer of the ANN in Figure 24. The input vector is presented as $\bar{y} = (y_1, \dots, y_{j,} \dots, y_J)$, and the output vector is presented as $\bar{o} = (o_1, \dots, o_{k,} \dots, o_K)$. The output value $o_k = f(net_k)$ and net_k can be computed by following equation:

$$net_k = \sum_{j=1}^{J} w_{kj}\, y_j. \tag{22}$$

The desired output for each of the corresponding outputs is the following vector $\bar{d} = (d_1, \dots, d_{k,} \dots, d_K)$. This vector becomes important after the learning process is finished and it is necessary to evaluate already learned ANN. In literature this vector is commonly known as the target vector which is compared against all learning patterns $p = 1, \dots, P$ from the learning data set A_{learn}. At this point, the learning problem can be transformed to an optimization one by defining the following error function [75]:

$$E_p = \frac{1}{2}\sum_{k=1}^{K}(d_{pk} - o_{pk})^2, \tag{23}$$

where p is the learning point index, E_p denotes the error rate of the ANN and it is computed as the squares errors sum of the output neurons. The learning process can be considered as the search for the weight settings that minimizes E_p. This can be done by using the gradient-based steepest descent on E_p [75]:

$$\Delta w_{kj} = -\alpha \frac{\partial E_p}{\partial w_{kj}} = -\alpha \frac{\partial E_p}{\partial (net_k)} \frac{\partial (net_k)}{\partial w_{kj}} = \alpha \delta_{ok} y_j, \tag{24}$$

where α is a positive learning rate which governs the speed of learning, δ_{ok} is the generalized learning signal in the k-th output neuron so it is possible to notice that $\delta_{ok} = -\partial E_p / \partial (net_k)$. The final rule for updating the *j*-th weight of the *k*-output neuron is defined by the following equation:

$$\Delta w_{kj} = \alpha (d_{pk} - o_{pk}) f_k' y_j, \tag{25}$$

where f_k' is the derivative of the activation function with respect to net_k, expression $(d_{pk} - o_{pk}) f_k'$, is the generalized error signal flowing back through all connections ending in the *k*-th output neuron.

Another step in the process is adding one output layer or one or several “hidden” layers between the input and output layer in the ANN in order to create a multilayer feed-forward ANN. An illustration of a two layer feed-forward ANN can be seen in Figure 25.

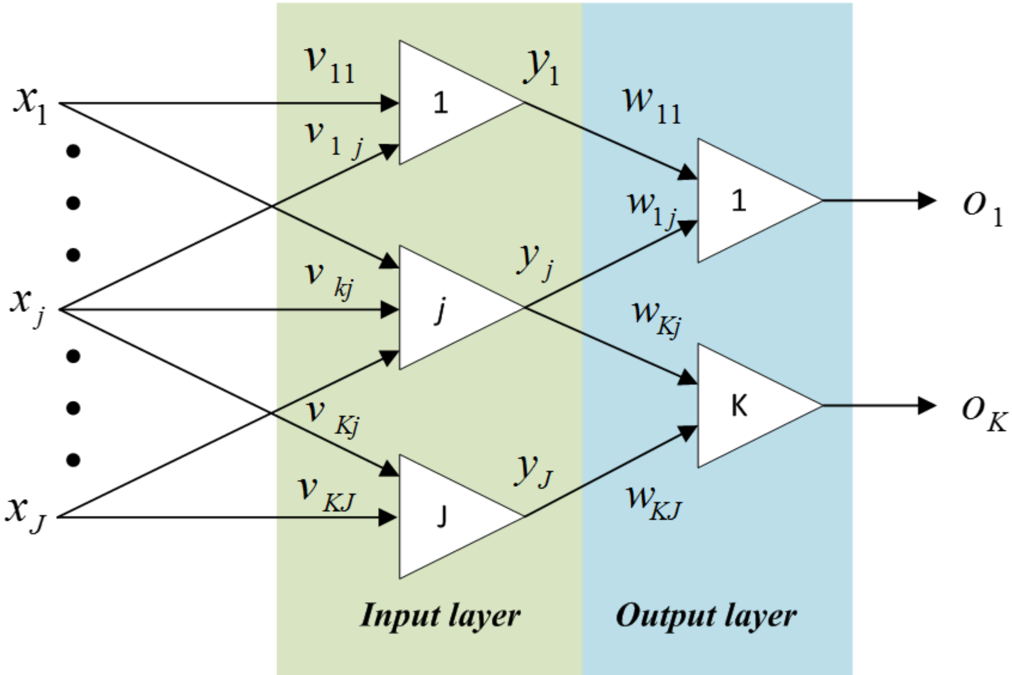

Figure 25: A two-layer feed-forward ANN [75]

The ANN in Figure 25 has two layers. The first layer is known as the input layer since its takes the input vector $\bar{x} = (x_1, \ldots, x_{k,} \ldots, x_K)$ and it has its own weights denoted by v_{ji}. The output vector of the input layer in Figure 24 is the input vector for the next layer which is presented in Figure 25, and represents the output layer in a multilayer network. The real breakthrough in learning multilayer feed-forward ANNs occurred when the error backpropagation method was introduced. This learning method is based on making the transfer functions differentiable [75]. The error backpropagation learning method applied on multilayer ANN consists of six steps:

- **Step 1.** Set the learning rate. Randomly initialize weights in the ANN to small values. Initialize counters and cumulative error (*k=1*, *p=1*, *E=0*);
- **Step 2.** Apply input $\bar{x}^p$ and compute the corresponding $\bar{y}^p$ and $\bar{o}^p$;
- **Step 3.** For every output neuron compute δ_{ok} (as the generalized error signal flowing back through all connections ending in the k-th output neuron), and for input neuron determine $\delta_{yj} = \left(\sum_{k=1}^{K} w_{kj}\right) f_k'$;
- **Step 4.** Modify the weights of the input layer $v_{ji} \leftarrow v_{ji} + \alpha\delta_{yj}x_i$ and output layer $w_{ji} \leftarrow w_{ji} + \alpha\delta_{ok}y_i$;
- **Step 5.** If $p < P$, increase the value of p by one and go to step 2. Otherwise, go to step 6;

- **Step 6**. Fixing the weights and computing the error E. If the E is the lower that the predefined value learning process will be stopped, otherwise, it is possible to permute elements of A_{learn}, set $E = 0$, p = 1 and increase k by one, and go to step 2.

In the presence of temporal dependencies, e.g., when learning to predict future elements of time series (with a certain prediction horizon), the feed-forward ANN needs to be extended with a memory mechanism to be able to take into account the temporal structure in the data. The first version of the ANN with a fixed time delay was the so-called Time-Delay Neural Network (TDNN). The input window into the past has a finite length D. If the output is an estimate of the next step of the input time series, such a network can be considered as a nonlinear autoregressive model of order D [75]. In some cases, its simple architecture with fixed D cannot capture the temporal characteristic of a data generation source.

This problem has created a demand for Recurrent Neural Network (RNN) models which contain feedback loops to preserve information about the past in the form of the information processing state as well as feed-forward connections. [75]. The RNN enables flexibility of the input window length. In contrast to the feed-forwarded ANN, the RNN can contain connections between neurons of the same layer and/or between a higher and a lower layer. These connections are made possible by the use of time delays. Furthermore, it is possible to represent the RNN as a feed-forwarded ANN with some fixed one-to-one delayed connections. This can be done by introducing an additional context layer with delayed activations of neurons from a selected layer or several layers. All RNN models are very convenient for making all kinds of predictions, therefore, in this thesis, one of them will be selected and used in order to predict the stochastic nature of traffic flows on motorways.

According to [76] RNN models can be divided into two classes: 1) fully connected networks, and 2) Nonlinear AutoRegressive with eXogenous Inputs (NARX) models. The NARX network architecture comes only with one feedback connection from the output neuron rather than from hidden states, which is in contrast with fully connected recurrent networks that are computationally rich due to a lot of feedbacks.

This thesis will use a class of NARX models as the discrete-time nonlinear systems, in order to predict traffic demand. The NARX model uses the following generic equation for computation of its output:

$$y(t) = f(u(t-n_u), \dots, u(t-1), u(t), y(t-n_y), \dots, y(t-1), \qquad (26)$$

where $u(t)$ is input in the network at time t, and n_u and n_y are the input and output order, and the function f is a non-linear function. If the mentioned function is presented by multilayer perceptron, the overall model can be considered a NARX network [76]. In Figure 26, it is possible to see an example of a NARX model architecture.

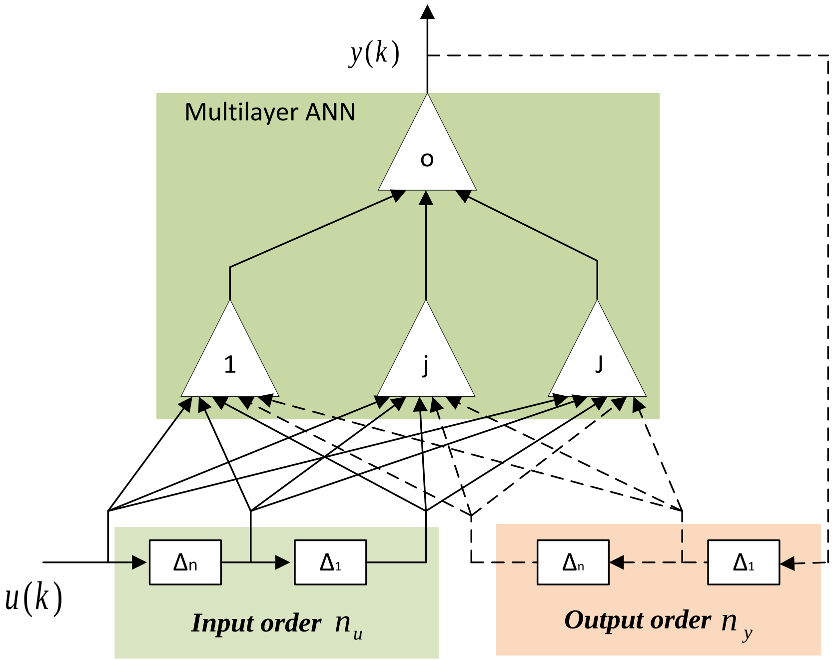

Figure 26: Example of a NARX ANN architecture [76]

3.3.3. Adaptive neuro-fuzzy inference system

The knowledge base (a set of IF-THEN rules) and parameters of FIS membership functions are usually tuned by a human expert. The human expert gathers data through the working experience in his domain of expertise and creates a set of rules by using his own central nervous system or the natural neural network of his brain. These rules created by the expert's brain represent his expert knowledge base (IF-THEN rules) by virtue of which he can perform control actions in an expertise domain or make predictions. Some systems such as urban motorways are extremely complex due to their stochastic nature and therefore generate a huge traffic database to which the FIS must be tuned. Such a huge amount of data can represent a processing problem for a human expert and undermine his goal to adequately tune FIS parameters. For example, a human expert will easily define the most prominent linguistic based IF-THEN rules for the FIS whose main function is to compute metering rates. On the other hand, it will be very hard for him to define membership function parameters for the linguistic variables based on a huge set of traffic data or derive a set of IF-THEN rules from the same dataset. It is possible to conclude that it is nearly impossible to find such an expert or a group of such experts.

The ANFIS is introduced in order to bypass the need for a human expert and his tendency to make poor decision when presented with a huge amount of data. The ANFIS's main function is to use a set of learning data which will be presented to the adaptive ANN, and the output of the ANN will be the modelled and tuned FIS. The ANFIS architecture is based on the adaptive ANN whose main objective is to model *Takagi–Sugeno* FIS as the final control structure.

The adaptive ANN is one type of the aforementioned multilayer feed-forward ANN. During the learning process, these ANNs usually use learning algorithms based on the supervised learning method. Furthermore, the architecture of the adaptive ANNs enable characteristics that consist of a number of adaptive nodes interconnected directly without any weight value between them [77]. It is important to emphasize that these nodes are not strictly connected with the definition of a neuron although the nodes can be neurons themselves. Each node in the adaptive ANN has different functions and tasks, and the output depends on incoming signals and parameters that are available in the node [77]. A learning rule must be designed in such a way that it can affect the parameters in the node. Furthermore, the mentioned learning rule must be designed in order to reduce the occurrence of errors at the output of the adaptive ANN.

The simple ANFIS structure with two inputs and one output will be used in order to explain the ANFIS architecture. Furthermore, the mentioned ANFIS structure will require few rules, and therefore its structure will be easier to explain. In the following example, two rules will be used to design the IF-THEN structure for the *Takagi–Sugeno* FIS model:

Rule 1. If x is A_1 and y is B_1 Then $f_1 = p_1x + q_1x + r_1$,

Rule 2. If x is A_2 and y is B_2 Then $f_2 = p_2y + q_2y + r_2$,

where A_1, A_2, B_1, and B_2 are parameters which represent the membership functions assigned to inputs x and y. Since the *Takagi–Sugeno* FIS model will be designed, the mentioned parameters belong to the premise part of the rules, while the parameters p_1, p_1, r_1 and p_2, p_2, r_2 are linear parameters of the consequent part of the rules.

In line with reference [77] there are five layers in the ANFIS architecture and each layer has its unique role in FIS modelling. An illustration of the ANFIS architecture can be seen in Figure 27. A brief description of the architecture of ANFIS will be now provided.

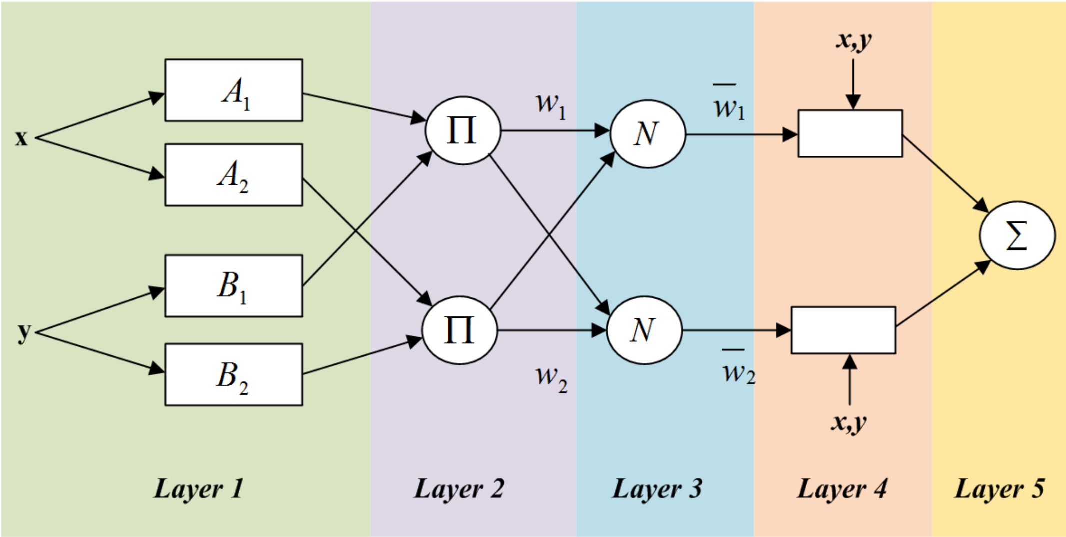

Figure 27: Illustration of the ANFIS architecture

Layer 1. Each node represents one membership function and assigns a degree of membership value that is given by the input to the particular membership functions. It is possible to conclude that the output of this layer represents the degree of membership for each input value.

Layer 2. Each node in this layer is fixed (non-adaptive). The output of each node in this layer is computed by multiplying all its input signals. The input signals are degrees of membership values that were computed based on the membership functions in the previous layer. Those input signals form a single rule. For example, input signal for the first rule will be a sum of membership degree value of input *x* according to the membership function A_1 and input *y* according to membership function B_1. Each node in this layer represents the firing strength w_i of each rule. In the second layer, the T-norm operator with a general performance, such as AND, is applied to obtain the output [77]. The generic equation for the computation of one rule firing strength from two possible rules in this example is presented in the following equation:

$$w_i = \mu_{Ai}(x) \times \mu_{Bi}(y), \quad i = 1, 2 \tag{27}$$

where $\mu_{Ai}(x)$ and $\mu_{B1}(y)$ are the degree of membership functions for the fuzzy sets A_i, and B_i, respectively.

Layer 3. Each node in this layer is fixed. The main goal of this layer is to assess the implications and consequences of particular rules. Namely, each *i*-th node calculates the activation value of *i*-th rule in the sum of all activation rules' values that are available within the ANFIS ANN. This layer is providing the so-called - *normalized activation value*. Normalized activation value can be computed by the following equation:

$$\overline{w}_i = \frac{w_i}{\sum_i w_i}. \tag{28}$$

Each node corresponds to one *fuzzy* rule, which means that weights between the third and fourth layer correspond to normalized factors of confidence in the veracity of each *fuzzy* rule. They are established in the learning phase by tuning weights (w_1, w_2) and by analysing activation functions' results in each node of the system.

Layer 4. contains the procedure for realizing disjunction of a consequent part in the *fuzzy* rules. Each node in this layer adapts its firing strength according to an output of a previous layer. They produce rule outputs based on consequent parameters of this layer. Outputs of this layer are computed by following generic functions for both of the example rules [77].

$$\overline{w}_i f_i = \overline{w}_i (p_i x + q_i x + r_i), \quad i = 1, 2 \tag{29}$$

where $(p_i x + q_i x + r_i)$ is the parameter of the node. The parameters in this layer are referred to as consequent parameters. It is possible to conclude that the fourth layer provides the product of a normalized firing strength from layer three $\overline{w}_i$ and its corresponding consequent parameter set.

Layer 5. contains only one node. It is a fixed or non-adaptive node and its main goal is to compute the overall output as the summation of all incoming signals from the previous nodes [77].

The ANFIS is commonly trained by a hybrid learning algorithm. In the forward pass, the learning algorithm uses the least-squares method to identify consequent parameters associated with layer four. In the backward pass, errors are propagated backward and current parameter values are updated using a gradient descent method [78], [79]. The corresponding *fuzzy* logic rules are established, and the relation generating method and inference synthesis algorithm are developed. The membership function types and parameters are determined by using an adaptive neuro training method. At the end of the training process, a newly calibrated *fuzzy* interference system is attained. It is possible to conclude that the ANN part of the ANFIS replaces the natural neural network in the brain of a human expert.

3.4. Cooperative control approach for urban motorways

Cooperative control approach applied in a general control system is based on sharing information's or tasks between dynamic control entities in order to achieve a global or multiple goals. The ITS Action Plan for the application of cooperation in traffic systems uses the terms Cooperative, connected and automated mobility in the ITS (C-ITS). The importance of data sharing and cooperation was emphasized on August 5th 2008 when the European Commission adopted the Decision 2008/671/EC to reserve the 5.9 GHz band for safety-related ITS applications. The Decision will adjust the terms of use with availability and efficient use of this frequency band on a non-exclusive basis in mind [80].

Data sharing and communication is especially important in traffic systems such as urban motorways. In urban motorways, there are a lot of different static infrastructural control entities and, as of late, vehicles utilizing the OBU which can be considered dynamic control entities. The concept of cooperative systems in traffic was first introduced between vehicles since control is highly distributed between them. In this case, near vehicles exchange information between each other which is, in fact, the strict definition of cooperation.

On-board driver assistance systems coupled with two-way communication between vehicles and with the road (motorway infrastructure) can help drivers have better control over their vehicle. This can have positive effects in terms of safety and traffic efficiency. Vehicles can also function as moving sensors and provide information regarding weather and road conditions including information about incidents. In this case, they can be used as high-quality information services [80]. These benefits of information exchange between vehicles and road infrastructure can be very useful on the urban motorway due to a need for achieving high data accuracy and even direct control over the vehicle in order to meet a higher LoS and higher safety standards. Furthermore, urban motorways are very often affected by bottlenecks, incidents or bad weather conditions at one of their segments while other segments remain unaffected. The information about the traffic situation for a particular motorway segment is especially important for vehicles and motorway control systems located upstream.

Cooperative systems between vehicles on in terms of usage on the urban motorway or in any other traffic system are still in the experimental phase. The main reason for this is the low penetration rate of vehicles equipped with the OBU and cooperation capabilities in real traffic systems in general. Currently, there are numerous concepts which involve information or assistance to vehicles in a ramp metering and VSLC region. A presentation of one of the mentioned concepts will be given in this chapter.

The latest approach in the application of cooperative systems on urban motorways includes the application of cooperation between different motorway control methods, such as ramp metering, the VSLC, and PLUS. Action 4.2 of the ITS Action Plan aims specifically for the "Development and evaluation of cooperative systems in view of the definition of a harmonized approach; assessment of deployment strategies, including investments in intelligent infrastructure" [80]. It is possible to conclude that the development of cooperative systems in traffic systems will rely heavily on traffic infrastructure. EU-funded cooperative systems research projects, e.g. Coopers, the CVIS and Safespot have delivered promising results that will contribute to the further development of cooperative systems in traffic systems, including urban motorways. It is interesting that all these projects are part of the COM eSafety project which has provided a definition of a communication architecture for cooperative systems [80].

3.4.1. Concept of cooperation

In a cooperative approach, each controlled entity tries to act in line with global performance goals [73]. The process of information and task sharing during cooperation between control entities is conducted in order to accomplish one or several global performance goals that are greater than the local goals of each individual control entity. In most cases, each individual entity can have their locally oriented goals as well. Some of these goals could be more important than the goals of other control entities in terms of their global goal. This implies that cooperation may assume hierarchical forms too [81]. In a cooperative system, decision-making processes are typically thought to be distributed or decentralized to some degree [81].

The potential benefits and the core logic of cooperation are illustrated by the example of a non-zero-sum game (game theory). This example involves two prisoners locked in separate cells. The prisoners want to spend a minimum amount of time in prison because they both need to be free in order to conduct their “businesses”. Each of the prisoners can choose between three choices given by the authorities:

1. If both confess to the charges, both will be jailed for five years;
2. If only one confesses, he will be freed but the non-confessor will be jailed for ten years;
3. If neither confesses, both will be charged for a minor offense and will be jailed for one year.

If both know that the other will act selfishly or if they communicate in some way, they will take the collective interest into consideration, so neither will confess and both will serve only one year in jail. In this case, they will continue to conduct their "businesses" in one year, sooner than in the case of the other two scenarios. This is a scenario where cooperation and common interest ("you help me, and in return, I will help you") wins and the pursuit of self-interest loses [82].

In Figure 28, it is possible to see an illustration of the cooperation concept between three control entities. The main elements of each controlled entity are actually located in the logic core. The logic core of a single entity contains the local control logic which is integrated with the cooperative control logic. The cooperative control logic is tasked mainly with making adjustments of control actions computed by the local logic. These adjustments are made according to the received data from other local entities. The data processing shell process input data from the environment and the data received from other control entities. Additionally, the mentioned data processing shell prepares the data for transmission to other control entities.

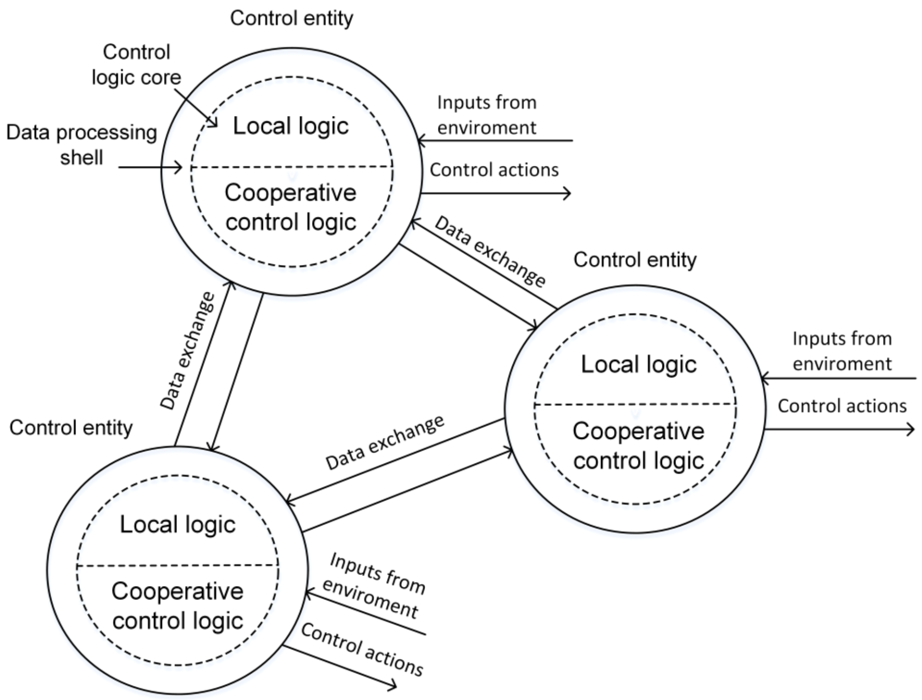

Figure 28: Illustration of cooperation concept between three control entities

3.4.2. Application of cooperative approach in urban motorway control

As was mentioned earlier, cooperative control applied on a particular urban motorway segment can be achieved by different traffic control methods or entities. Cooperation between different motorway control methods can produce the effect of synergy between them. In some urban motorway segments, the mentioned effect can produce better overall results in comparison with the results that would be achieved if the motorway control methods involved in cooperation are implemented as standalone applications. When the cooperative control approach on the urban motorway is planned, it is necessary to be aware of each motorway control method's restrictions which affect a particular traffic flow on the motorway system.

For example, ramp metering affects traffic flows related to on-ramps, while the VSLC affects mainstream flow. Furthermore, ramp metering with its traffic lights can completely stop on-ramp flows, while the VSLC can only reduce or increase the speed of mainstream traffic flows. In that case, it is possible to conclude that ramp metering as a motorway control method has the higher degree of restriction compared to the VSLC. As was mentioned earlier, the dependency between the on-ramp and mainstream flows is high on urban motorways, since the distance between on- and off-ramps is lower compared to classical motorways. This scenario demands more comprehensive control over urban motorway traffic flows. Since the VSLC and ramp metering affect different traffic flows on the urban motorway, they can provide more comprehensive control over it. The synergy between these two motorway control methods can be achieved by establishing cooperation between them. Each of the motorway control methods (or their individual components) involved in cooperation conducts analysis of the locally acquired traffic data. Based on this analysis it is possible to exchange specific data between different motorway control methods and adjust their previously computed control outputs. This adjustment can induce better final solutions, which will be more in line with a global goal or goals which have to be met.

Vehicles as potential control entities can enhance the cooperation between two different motorway control methods. All vehicles built upon "drive-by-wire" architecture and equipped with the OBU can exchange data with other control motorway method/s and other vehicles with the OBU. Each vehicle behaves differently according to the driver's psychophysical profile and depending in which type of motorway traffic flows the vehicle is located. Using the cooperation method, it is possible to have a complete or partial control over the "drive-by-wire" vehicles in predefined traffic scenarios. The most important impacts of mentioned cooperative control approach on urban motorways are the homogenization of speeds between on-ramp flows and the mainstream, speed limit obedience, and well timed vehicle inclusion from on-ramps into the mainstream.

Furthermore, it is possible to achieve cooperation between on-ramps (each on-ramp is one control entity) or it is possible to enable cooperation of the VSLC and PLUS. This chapter will go on to present the core topic of this thesis i.e. the cooperation between ramp metering and the VSLC. Furthermore, cooperation between ramp metering and vehicles, and the VSLC and vehicles will be also presented on a conceptual level.

3.4.2.1. Cooperation between ramp metering and VSLC

It was mentioned in previous sections that cooperation between the VSLC and ramp metering can provide more comprehensive control over urban motorway traffic flows. Both mentioned control methods applied as standalone implementations affect only one type of motorway traffic flows. In order to mitigate congestion at the mainstream, which has to appear near one on-ramp, and to prevent upstream shockwave propagation due to that congestion the VSLC is applied. The VSLC in cooperation with ramp metering can gradually decrease the speed of the upstream flow before congestion starts to form. Such an approach gradually decreases mainstream speed but enables a higher mainstream speed during the congestion period, unlike the scenario without the VSLC [38].

In this thesis the VSLC will be made to cooperate with the HELPER ramp metering algorithm. The mentioned cooperative approach uses the VSLC to decrease the speed of vehicles coming into the area between the last “slave” on-ramp and the congested one. It can be concluded that virtual queues provided by HELPER and speed reduction in the area between the last “slave” and the congested on-ramp induced by the VSLC significantly reduces traffic density upstream of the congested on-ramp. The lower upstream density of the congested on-ramp provides additional mainstream capacity to accept vehicles coming from congestion back-propagation. The concept and example of cooperation between HELPER ramp metering and the VSLC can be seen in Figure 29.

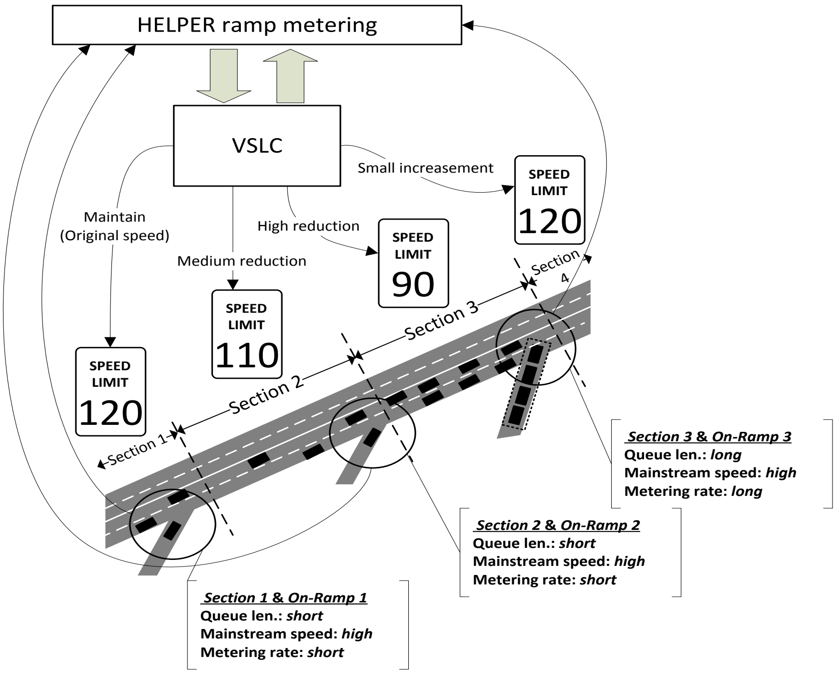

Figure 29: Concept and example of cooperation between HELPER ramp metering and VSLC

The research presented in [83] concludes that a system that uses the ALINEA ramp metering algorithm and the VSLC reduces travel time by around 1.62 % compared to the ALINEA standalone implementation. This is one of the first documented cases of cooperation between ramp metering and any other motorway control method.

3.4.2.2. Cooperation between ramp metering and vehicles

Cooperation between vehicles equipped with an OBU and the on-ramp control computer dedicated to communication with vehicles (RMS-r2v) is, due to its complexity, presented at this point on a conceptual level only. This research proposes enabling cooperative control at the moment when a vehicle is stopped at the on-ramp end and is waiting for the green light. The vehicle stopped at the on-ramp waiting zone sends information about its location, speed, and throttle, while the on-ramp computer delivers information to the vehicle about its current signal plan. When the green light is turned on, the on-ramp control unit obtains throttle control over the first vehicle in the queue. The vehicle movement starts automatically preventing inexperienced drivers from failing to leave the on-ramp during the short green light phase [35].

Additionally, the vehicle OBU can receive information about mainstream merging manoeuvres. This can be done when ramp metering cooperates with another motorway control method. The types of mainstream merging manoeuvres depend on the motorway control method which is cooperating with the ramp metering control system. If selective prohibiting of lane change is cooperating with ramp metering then the on-ramp computer will forward only simple merging trajectories to the vehicle OBU. When a vehicle becomes parallel with the mainstream direction in the acceleration lane, the on-ramp control computer terminates its control over the vehicle and the driver continues to manually control the vehicle. The OBU also provides appropriate information for the driver when the remote automatic control over the vehicle is established and when it is terminated [52]. A diagram of basic RMS-r2v activities can be seen in Figure 30.

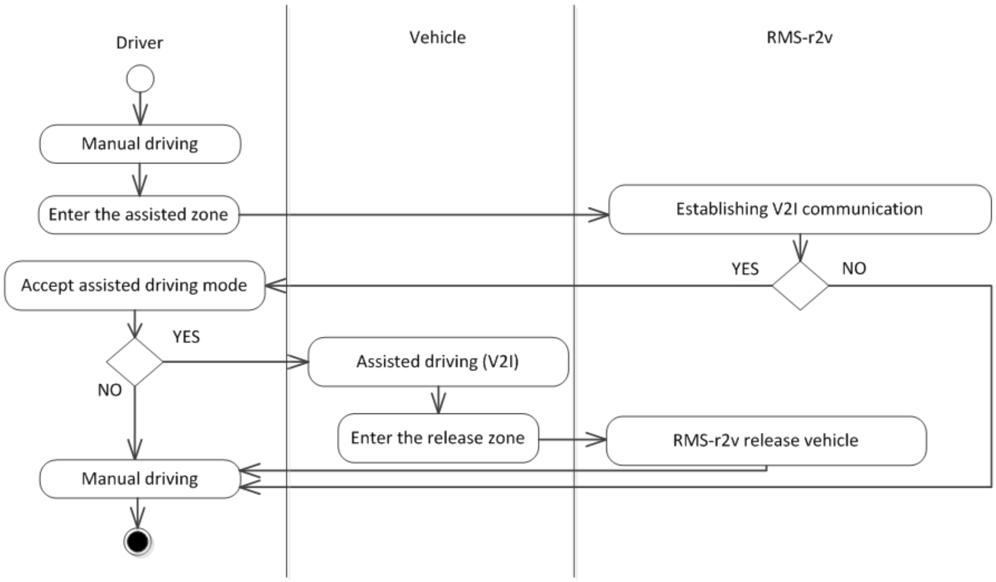

Figure 30: Basic activity diagram of on-Ramp Metering and assisted driving System based on ramp metering-to-vehicle communication [52]

3.4.2.3. Cooperation between VSLC and vehicles

The main VSLC infrastructure problem (e.g. the VMS) is that mainstream drivers are not fully obeying the posted speed limits. In order to boost the positive impact of this system on road safety, additional cooperation can be established. This cooperation is established between mainstream vehicles with an OBU and the VSLC. If the VSLC subsystem can directly communicate with the vehicle's OBU, a similar effect can be achieved as the abovementioned effect with vehicles waiting on an on-ramp. Two possible levels of influence on driver behavior are present in this system.

The first functions as an information system provides information about the current speed limit for a particular motorway segment. The second level provides an override process of the current vehicle speed by enforcing the speed limit should it not be obeyed. In this situation, the vehicle's OBU sends information about the current vehicle position and speed directly to the nearest on-ramp computer (RMS-r2v) in the Speed Limit Control (SLC) zone.

The VSLC computation unit computes the optimal speed limit value according to traffic situations and data obtained from vehicles. The computed results are sent to the VMS and the vehicle's OBU in the form of the optimal speed limit. The vehicle's OBU compares the value of the speed limit obtained from the VSLC computation unit to the current vehicle speed. Based on the difference between these two speed values, the OBU automatically adjusts current vehicle speed to the posted speed limit. The entire system is designed only to decrease the speed of vehicles driving faster than the currently valid speed limit at a critical segment of the urban motorway. The speed of other vehicles is left unchanged. The vehicle speed adjustment in an urban motorway mainstream based on a decision of a cooperating system can make a significant impact on traffic safety at critical motorway segments and have a significant influence on throughput and safety. These segments are on-ramps, tunnels, curvatures, etc. In Figure 31, it is possible to see basic activity in the case of cooperation between the VSLC and vehicles equipped with an OBU.

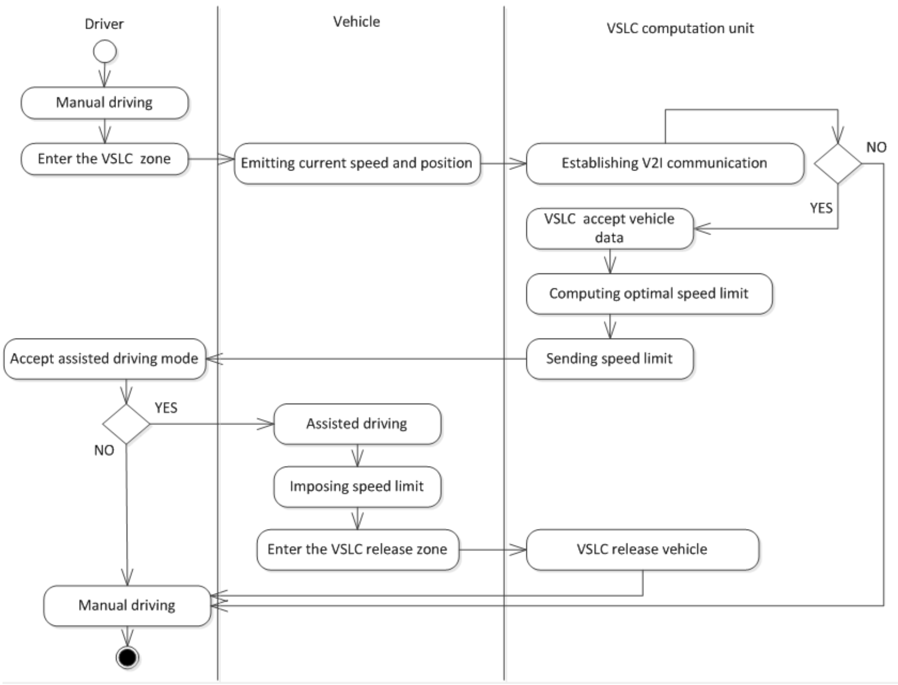

Figure 31: Basic activity in the case of cooperation between the VSLC and vehicles equipped with the OBU

4. Ramp metering based on machine learning

An urban motorway system is hard to model and it is almost impossible to build an exact model of all its traffic flows and their interactions due to their nonlinear stochastic nature and in complete information about them. Modern urban motorway simulators can provide only explanatory traffic models so programming and testing explicit ramp metering algorithms is unfeasible in some cases. The answer to these problems is building a ramp metering algorithm with the ability to adapt to the fluctuations in traffic demand which are hard to predict by using the simulator. Furthermore, this algorithm should not be procedurally programmed to react to traffic demand fluctuations or bounded for any traffic model because that would be inadequate in most cases. The structure of an adaptive ramp metering algorithm could overcome flaws of a simulator in a realistic representation of traffic flow on the motorway. This would be especially noticeable cases when such a ramp metering algorithm is actually removed from a simulation environment and set up in a real world environment.

The latest approach in ramp metering algorithm design usually involves methods that are part of the Artificial Intelligence (AI) domain. AI methods based on machine learning are especially interesting for ramp metering since they have the ability to adapt to common traffic patterns. These traffic patterns can be identified by using machine learning process with a presented traffic dataset (learning dataset). Since machine learning does not require an exact model of a system whose behaviour needs to be predicted or controlled, they are even more suitable for application in ramp metering design. Based on the control knowledge of common traffic patterns collected through the learning process, it is possible to detect sudden changes in a traffic system, such as incidents or other fluctuations. It is possible to conclude that with the machine learning approach it is possible to classify current traffic behaviour as stable (or predictable), bi-stable (or stable traffic state near critical density) or unstable (or unpredictable incident) if a large enough and representative historical traffic dataset is used in the learning process. The structure of the traffic dataset is very important for all machine learning methods since they are closely related to computational statistics. According to the current traffic state and learned control knowledge about adjusting metering rates, it is possible to provide an adequate solution in the form of a change in metering rates. The main goal of these computed solutions is to establish a stable traffic environment (stable or bi-stable traffic states).

It is important to mention that the performance of machine learning depends on the structure of a learning dataset. A learning dataset should contain a sufficiently comprehensive and representative quantity of traffic data collected from the controlled motorway system. The most widely used approaches in ramp metering based on machine learning are different types of ANNs, Reinforcement Learning (RL) methods, Iterative Learning Control (ILC), and hybrid AI system which involve one type of the ANN as the machine learning mediator. A brief overview of current approaches in ramp metering based on machine learning will be presented in this chapter.

This thesis will focus on the use of an ANFIS framework in order to develop a unique ramp metering control methodology. It was mentioned in the previous chapter that an ANFIS framework uses an adaptive ANN. After the learning process, an output of the mentioned ANN is an FIS that will actually provide a computation of metering rates based on the knowledge base contained in the ANFIS. The learning dataset for the ANFIS will contain data that is gathered after simulating three different ramp metering algorithms. These algorithms are simulated on the same urban motorway section, and under the same conditions (traffic demand, etc.). The described newly proposed ramp metering algorithm is named INTEGRA since it integrates control knowledge from different ramp metering algorithms through the process of machine learning.

Furthermore, this thesis will also deal with the application of the NARX network with machine learning capabilities. The network will be used in order to predict on-ramp traffic demand. The results of on-ramp traffic demand predictions will be used in order to adjust INTEGRA pre-computed values of metering rates.

4.1. Current approaches in ramp metering based on machine learning

The ANN is one of the most widely used methods for application of machine learning in ramp metering. All ANNs used for ramp metering algorithms are designed based on the spatiotemporal approach. This approach enables ANNs to deal with a time-correlated sequence of spatial patterns [55]. There are three types of spatiotemporal ANNs: the Multi-leg Network (MLN), the RNN, and the Spatiotemporal Pattern Recognition Network (STN). The MLNs learn from the learning dataset that contains a set of time sequenced data and generates time-dependent outputs [55].

The RN is characterized by an architecture that contains a feedback loop which brings a signal back to the same processing unit. This makes the RN trainable and adaptable in a non-linear system such as ramp metering applied on the urban motorway. The RN structure of the learning dataset is the same as for the MLN. The STN is still in the experimental phase. They are based on the dynamic associative memory for temporal patterns and require complex learning algorithms [55]. The MLN and the RN are most suitable for ramp metering algorithm design due to their methodology for the creation of learning datasets. The learning datasets of both ANN types are created upon the measured traffic data from the motorway. The aforementioned datasets are sorted in time-space sequences and adequately mapped with metering rates. It is important to mention that all ANN model types, instead of computing, estimate suitable metering rates. The estimations are derived according to a one type of rules that are produced by adjusting the ANN structure (learning process) according to the computed error rate. Figure 32 presents a self-adjusting ANN model for ramp metering which is described in more detail in [55].

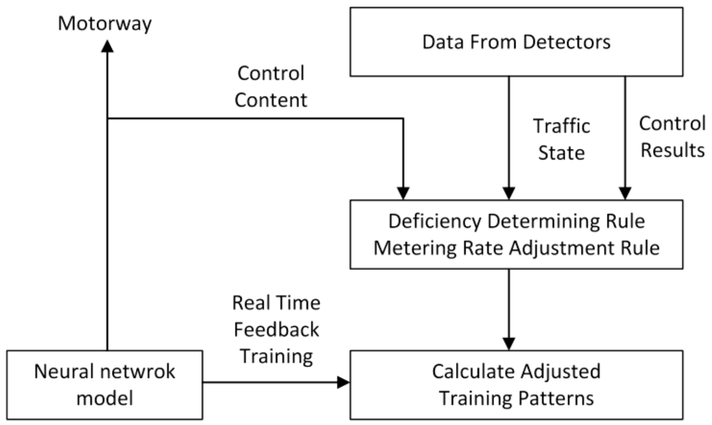

Figure 32: Self-adjusting ANN model for ramp metering [55]

The RL approach represents a type of learning which rewards an action if it achieves the desirable output result. The RL is one of the basic techniques of the Intelligent Agent (IA) technology. The learner or decision maker is named the agent, and everything that interacts with the agent is named the environment. The agent has a set of sensors, which are tasked with observing the state of the environment and performing a set of actions in order to change the state of the environment. The IA is a computation unit in the application of the RL in ramp metering dedicated to one on-ramp. The actions are metering rates, the environment in a wider context is an urban motorway, and the states describe the traffic state near the controlled on-ramp. The most important characteristics of the IA are trial and error search and delayed reward. The learner or an IA that senses its environment and/or acts in it, can learn by “trial and error” approach in order to select the optimal action. Optimal actions are the ones which result in the highest reward or the lowest penalty [84].

For example, in the application of the RL in ramp metering, the reward can be defined based on the on-ramp queue or by any other function which involves traffic parameters. Furthermore, it is possible to enable cooperation between several IAs by using communication channels between them and specialized logic. This cooperative logic will incorporate the status and decisions of other IAs in its final actions. These systems are commonly known under the term multi-agent systems. Figure 33 shows a single agent-environment interaction/communication with *N* agents that can communicate with each other in order to achieve global goal.

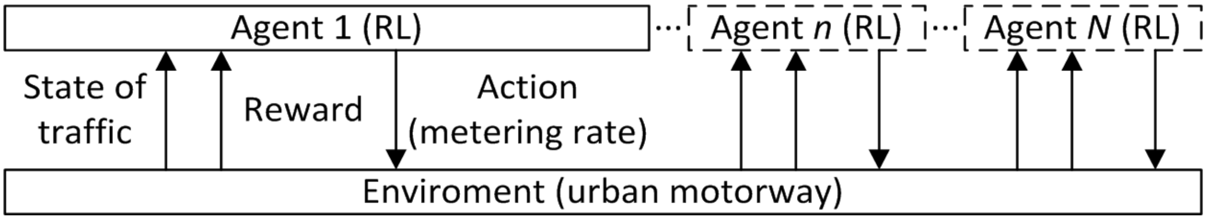

Figure 33: The agent-environment interaction/communication in case of one and in case of N agents that can communicate with each other

For a more accurate description of the interaction we can assume that the agent and the environment communicate in all sequences of the discrete time steps: $t = 0, 1, 2, \dots$ In each time step t, the agent receives a representation of the state of the environment, $s_t \in S$, where S is the set of possible states. In accordance with this, an action, $a_t \in A(s_t)$ is chosen, where $A(s_t)$ is a set of actions which are available in the state s_t. One step later, as a consequence of its action, the agent gets a numerical reward, $r_{t+1} \in R$ and finds itself in a new state, S_{t+1}. The agent obtains a reward or a penalty in order to evaluate the desirability of the current state [84], [85].

The Q-learning algorithm based on the Q-equation finds the optimal action-selection policy, essentially a type of rule, through the learning process. Based on the mentioned rule the Q-learning algorithm will select a given action for a given state. The learning process is governed by the old value computed by the Q equation, the estimated behaviour of the Q equation, the reward system and the learning rate. The Q-learning algorithm is the most widely used technique in the application of the RL in ramp metering.

According to [86] it is possible to apply the ILC for local ramp metering algorithm design. The learning process of the ILC uses data from a previous iteration in order to improve the control output. This action enables progress towards a suitable control action which can be found iteratively. The original ramp metering problem in the ILC approach must be formulated as an output tracking, disturbance rejection, and error compensation problem. The ILC can be implemented in the ramp metering algorithm design under the basic assumption that urban motorway traffic patterns are recurrent. Furthermore, it must be implicitly assumed that the ILC will provide a ramp metering algorithm based on fixed-time traffic control approach [86]. In order to develop an ILC which will be able to deal with non-recurrent behaviour (iteration-varying parameters, iteration-dependent trajectory and input constraints), it is necessary to introduce a set of control laws or design the ILC as a type of add-on to the other local ramp metering algorithms.

4.2. INTEGRA ramp metering algorithm

There are several experimental proposals for the application of the ANFIS framework in ramp metering algorithm design. For example, in [87] ANFIS is applied along with the ILC in order to compensate for the unknown traffic system nonlinearity and input gain respectively. In [88] ANFIS is trained on-line and the metering rate is computed each minute. The FIS is tuned according to the traffic data collected 15 minutes into the past in order to minimize the Total Time Spent (TTS) in the motorway system. The TTS takes into account mainstream density and on-ramp queues. It is expressed in vehicle-hour units.

In line with the mentioned research approaches, the author of this thesis proposed in [13] a concept of ramp metering algorithms based on the ANFIS framework, which is oriented towards the mitigation of various types of congestion by learning control from other ramp metering algorithms. The proposed concept is oriented towards the mitigation of congestion which is periodic, and for those who are varying in strength and in time. This is done by using ANFIS self-adaptation properties which can compensate for disturbances in traffic flow on urban motorways. The self-adaptation properties of the ANFIS framework applied in ramp metering are discussed and analysed in detail by the author of this thesis in [66]. The most prominent property of the ANFIS based ramp metering algorithm is the tuning of *Takagi – Sugeno* FIS parameters according to the learning dataset. The control and implementation of strategy of ramp metering based on the ANFIS framework will depend on the methodology for data gathering, and the structure and type of data used in the learning dataset. The learning dataset can be provided to the ANFIS framework during its operational work (on-line learning) or before its operational work (off-line learning). In [19] the author of this thesis proposed an off-line learning concept for ANFIS based ramp metering algorithm. The ANFIS ANN is trained by a hybrid learning algorithm (combination of feedback error propagation and least squares method). The FIS is the final product of the ANFIS based framework that actually provides metering rates for on-ramps. The author of this thesis introduced an FIS applied to several on-ramps that is tuned through the training process of the ANFIS framework with the control knowledge of different ramp metering algorithms in [13]. An augmentation of the CTMSIM simulator in [38] with a cooperative module enabled the creation of a unique FIS, which has the ability to compute metering rates according to the System-Wide cooperative approach on every on-ramp on the motorway model.

In this thesis the novel application of the ANFIS framework for the ramp metering algorithm design is presented. This new ramp metering algorithm is named INTEGRA after its control strategy. INTEGRA differs from other applications of the ANFIS framework in ramp metering due to its unique off-line methodology for gathering and structuring learning datasets. The introduction of criteria functions in post-processing of initially gathered learning datasets provides the mentioned structuring of data, and steers integrated knowledge towards specific goals. The core concept of INTEGRA is based on the integration of selected existing ramp metering algorithms into one comprehensive control strategy with a specific goal. In [52] the author of this thesis selected ALINEA as local, HELPER as cooperative and SWARM as the competitive ramp metering teaching algorithm. Those ramp metering algorithms are selected as the best representatives within their categories with respect to the implemented control strategy. The mentioned learning concept that is based on structured control knowledge, and proof of its operational work, which will be presented in Chapter 5. represents a core scientific contribution of this thesis. An example of the ALINEA and HELPER ramp metering algorithm integration in the proposed INTEGRA control working concept on a smaller section of urban motorway is shown in Figure 34.

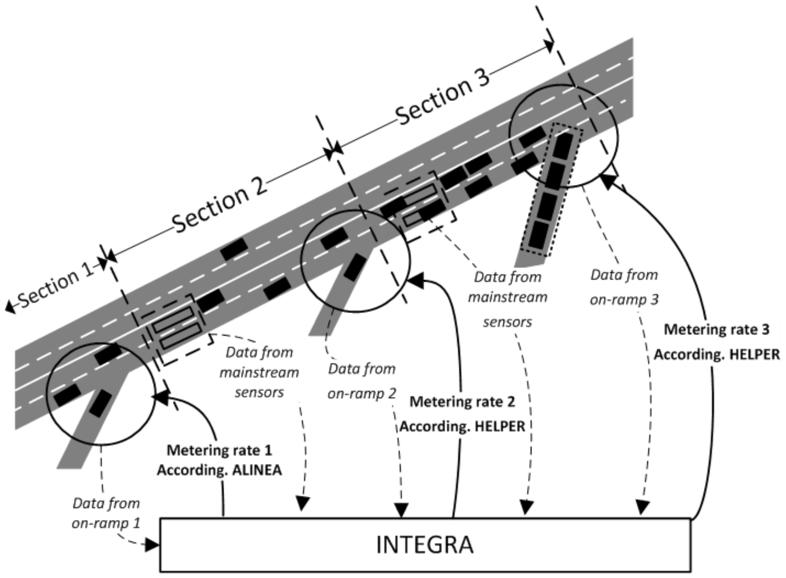

Figure 34: Example of proposed INTEGRA control working concept

The selected ramp metering algorithms have different control logic which enables INTEGRA to resolve different types of congestion. Basically, each of the selected ramp metering algorithms preforms best under specific traffic conditions. In order to provide adequate metering rates for a wide range of traffic scenarios it is necessary to teach the ANFIS framework of INTEGRA by using gathered knowledge. This knowledge is constructed by means of sufficiently different ramp metering algorithms used to prepare the learning data set with respect to their control logic. Since the selected ramp metering algorithms provide knowledge in the form of input (traffic parameters) – output (metering rates) pairs during a simulation run, they will be named teaching ramp metering algorithms. It is important to mention that teaching ramp metering algorithms must be simulated on the same simulation model (with the same constructional and traffic parameters). This simulation model must also be used for the simulation of the trained INTEGRA algorithm in order to verify its operational work. The idea is to verify the hypothesis that, in comparison to the three previously mentioned standalone ramp metering algorithms upon which integration is conducted, INTEGRA can produce better results in a similar traffic scenario. Based on simulation data obtained from the teaching ramp metering algorithms, the initial learning dataset will be created. This thesis will provide inclusion of criteria function and an analysis of their setup based on which particular learning pairs from the mentioned learning dataset are selected and stored in the final learning dataset. Selection of adequate learning pairs from the initially created learning dataset will enable convergence towards the desired ratio of the included MoS in criteria functions. In the following step, the mentioned learning dataset will be presented to the ANFIS framework of INTEGRA.

After the learning process, the FIS will be created and appropriately tuned. At this point, the FIS will be responsive to various traffic scenarios based on newly constructed knowledge learned from the different teaching ramp metering algorithms [89]. The functional scheme of INTEGRA learning can be seen in Figure 35.

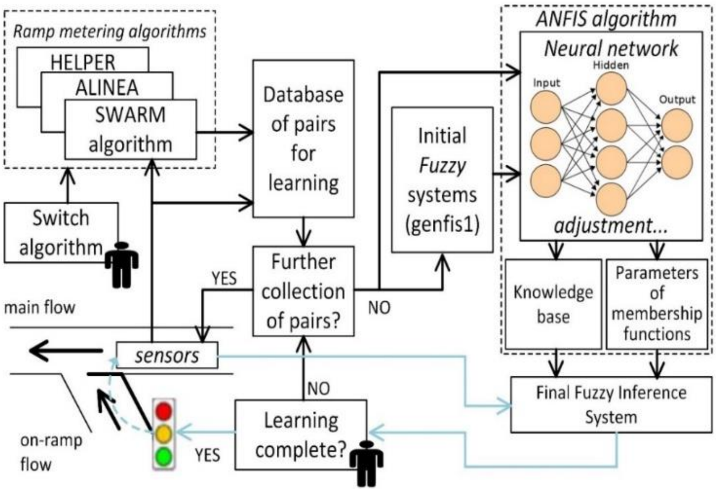

Figure 35: Functionality scheme of INTEGRA [52]

4.2.1. ANFIS framework for the INTEGRA ramp metering algorithm

INTEGRA uses the same ANFIS framework structure as is described in chapter 3. The setup of ANFIS framework is modified to produce the *Takagi – Sugeno* FIS model which has two input variables and one output in the form of metering rates. Each input variable has 5 membership functions. *Fuzzification* is achieved by the use of the Gaussian *fuzzifier* and the middle of maximum (MOM) method for *defuzzification*. In Figure 36, it is possible to see the graphical structure of the ANFIS framework (the adaptive ANN component) for described FIS output properties.

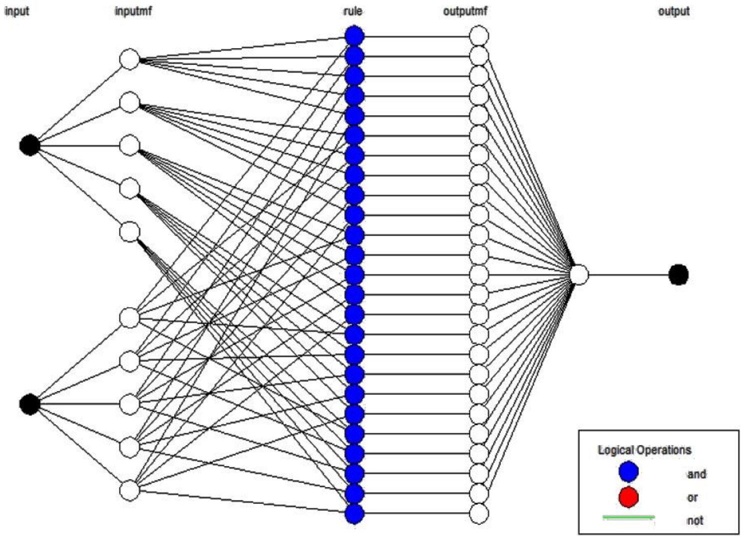

Figure 36: Graphical structure of the ANFIS framework (the adaptive ANN component) for the described FIS output properties

In order to learn the adaptive ANN of ANFIS according to the mentioned FIS specification, it is necessary to create an adequate learning dataset. The first step is to store and format the obtained simulation data from teaching ramp metering algorithms for data processing. The output of data processing will be a learning dataset. The learning dataset is organized in the form of an *L* matrix determinate by *N* x *B* dimensions. *N* denotes the number of on-ramps in a model multiplied by the number of 5 minutes intervals and the number of teaching ramp metering algorithms [66]. In this thesis, a simulation run will be conducted for a typical 24 hour day. The value *B* denotes the number of traffic parameters collected during all simulations runs. The last row in the *L* matrix is related to on-ramp metering rate.

$$L = \begin{bmatrix} {}^{1}tp_{1_1}^{\ 5} & {}^{2}tp_{1_1}^{\ 5} & \dots & {}^{n}tp_{1_1}^{\ 5} \\ {}^{1}tp_{2_1}^{\ 5} & {}^{2}tp_{2_1}^{\ 5} & \dots & {}^{n}tp_{2_1}^{\ 5} \\ \vdots & \vdots & & \vdots \\ {}^{1}tp_{k_1}^{\ 5} & {}^{2}tp_{k_1}^{\ 5} & & {}^{n}tp_{k_1}^{\ 5} \\ {}^{1}tp_{1_2}^{\ 10} & {}^{2}tp_{1_2}^{\ 10} & & {}^{n}tp_{1_2}^{\ 10} \\ {}^{1}tp_{2_2}^{\ 10} & {}^{2}tp_{2_2}^{\ 10} & & {}^{n}tp_{2_2}^{\ 10} \\ \vdots & \vdots & & \vdots \\ {}^{1}tp_{k_2}^{\ 10} & {}^{2}tp_{k_2}^{\ 10} & & {}^{n}tp_{k_2}^{\ 10} \\ {}^{1}tp_{1_a}^{\ t} & {}^{2}tp_{1_a}^{\ t} & & {}^{n}tp_{1_a}^{\ t} \\ {}^{1}tp_{2_a}^{\ t} & {}^{2}tp_{2_a}^{\ t} & & {}^{n}tp_{2_a}^{\ t} \\ \vdots & \vdots & & \vdots \\ {}^{1}tp_{k_a}^{\ t} & {}^{2}tp_{k_a}^{\ t} & \dots & {}^{n}tp_{k_a}^{\ t} \end{bmatrix}, \qquad (31)$$

where *tp* is the value of a traffic parameter which is defined by the traffic parameter marked with a number *n (e.g. 1 – Speed, 2 – Density, etc.)*, *t* is the 5 minute long interval in which *tp* is measured, $t \in \{5, 10, 15, \ldots, final\ number\ of\ 5\ minutes\ intervals\ in\ one\ month\}$, *k* is the number of on-ramps on the motorway model where tp is measured, $k \in \{1, 2, 3, \ldots, total\ number\ of\ on-ramps\ on\ motorway\ model\}$; and *a* is the number of the ramp metering algorithm involved in the process of creating *tp (e.g. 1 – ALINEA, 2 – HELPER, etc.)*.

Computation analysis of all permutations among available traffic parameters has to be made in order to determine which combination of traffic parameters has the most significant influence on on-ramp metering rate. That combination of traffic parameters will represent inputs to the future FIS that will be the product of the ANFIS framework. In this thesis, seven parameters are analysed: mainstream speed, density, flow, on-ramp queue, delay, travel time, and on-ramp demand. The decision, about the combination of these parameters, which have to be chosen is determined by using an exhaustive search technique. It is a type of a brute force technique and provides a list of every possible combination (all permutations) of traffic parameters respecting the maximum number of two FIS inputs constraint. According to the obtained list, every combination of traffic parameters is provided to the adaptive ANN of ANFIS. The combination of traffic parameters which has minimal learning error in the first learning iteration is chosen for FIS inputs [52].

4.2.2. Teaching ramp metering algorithms

In the previous section, it is explained that ALINEA, SWARM, and HELPER will be used as teaching ramp metering algorithms. All these ramp metering algorithms have different control logics and therefore perform best with a different type of congestion on urban motorways. For instance, the ALINEA ramp metering algorithm performs best if the on-ramps affected by this algorithm are fairly far from each other, and if on-ramp traffic demand is not high. It is possible to conclude that low dependency between on-ramps and lower traffic demand on them is ideal for local ramp metering. The SWARM ramp metering algorithm is one of the best algorithms to cope with congestion that occurs at regular intervals thanks to its predictive module. On the otherhand, the SWARM algorithm can potentially fail in the prevention of breakdown at an on-ramp with increased traffic demand in unexpected intervals of a day. This can consequently induce "shock waves". It is advisable to use the HELPER ramp metering algorithm in the mentioned traffic scenario. The HELPER ramp metering algorithm can effectively suppress upstream propagation of "shock waves" due to its ability to create upstream virtual queues. The HELPER algorithm is a relevant teaching algorithm for INTEGRA because this thesis is focused on analysing cooperative ramp metering algorithms. INTEGRA will also have the ability to provide a cooperative solution for the mentioned traffic scenarios since the HELPER algorithm, which is based on a cooperative control strategy, is included in the learning process.

It is possible to conclude that one ramp metering algorithm cannot respond to every traffic situation on the urban motorway with equal efficiency. This is the reason why it is imperative to develop a learning framework that will summarize knowledge from several different ramp metering algorithms into one control structure [7].

4.2.3. Discussion about INTEGRA criteria function

After creating the mentioned learning dataset, the best of all solutions provided by all teaching ramp metering algorithms for the same simulation step has to be selected. This is done by using the following function:

$$f(r) = X \cdot TT + Y \cdot D, \tag{32}$$

where *f(r)* is the metering rate function, X and Y are weighting factors for overall travel time (TT), and overall delay (D) respectively. Delay is defined as the difference between the actual amount of time spent by all vehicles on the motorway and the amount of time that would be incurred if vehicles travelled at free flow speed. Travel time is defined as the ratio between the length of the motorway model and achieved average speed. It is clear that a delay involves all vehicles on the motorway, including the vehicles from on-ramps, unlike the travel time which only considers vehicles in the motorway mainstream. By changing the value of weighting factors for these two parameters it is possible to favour mainstream vehicles (transit traffic) or merging vehicles from on-ramps.

In other words, solutions which enable a lower TT (better throughput of the mainstream) will be selected if the weighting factor of TT is higher than weighting factor of D. Conversely, if the weighting factor of D is higher than in the case of TT then solutions which enable higher metering rates will be selected. This situation will result with a higher TT since the throughput of the mainstream will be reduced due to large on-ramp flows. The objective is to find a balance between these two extreme situations. This thesis will also provide a comparative analysis which will include several different cases of the ratio between TT and D weighting factors. The solution, which provides the most balanced ratio between the achieved TT and D will be selected after analysing the results of the mentioned comparative analysis. After the application of the criteria function, the learning dataset is ready to be involved in the learning process.

4.3. The INTEGRA algorithm augmented with the traffic prediction function

In traffic control, the reaction to heavy and sudden congestion detected in this time step can be described by the following sentence: “If it's happening, it's probably too late”. The solution for this problem is the application of proactive control strategies. Proactive control strategies are based on spatial and temporal traffic demand predictions. The mentioned control strategies provide control actions in advance according to the spatial and temporal prediction of congestion. The latest approach in ramp metering algorithm design includes using various types of predictions in their final decisions concerning metering rates. The mentioned predictions are focused mainly on traffic flow evolution in time at the on-ramps and on the mainstream near the on-ramps. A ramp metering algorithm can introduce a set of decisions aimed mainly at preparing existent traffic flows for a traffic situation which is expected to arise in the near future, by using prediction data. Preparations are made in the form of changes in metering rates. For example, if a slow rise in traffic demand of a mainstream flow near a particular on-ramp is predicted, then metering rates can be reduced. The reduced metering rates, in comparison with metering rates computed by ramp metering which does not predict traffic flow tendencies pass less vehicles into the mainstream. This control action can consequently postpone or mitigate incoming congestion.

Using prediction data can provide a faster ramp metering response in the case when mainstream traffic density has a rising trend. This trend suggests that congestion is forming somewhere downstream and slowly back-propagating to the observed part of the motorway. In that case, a cooperative ramp metering algorithm can reduce the metering rate at several upstream on-ramps in order to reduce the impact of possible congestion back-propagation [15].

This thesis proposes augmentation of the INTEGRA algorithm with simple control logic which will include prediction results into the computation of the final metering rates. Furthermore, traffic demand at on-ramps is chosen for prediction since it has a direct impact on the traffic situation at on-ramps and consequently on urban the motorway mainstream. The mentioned predictions will be made by a special ANN model – the NARX network. The structure of the NARX network is explained in chapter 3.3. A detailed description of the NARX network tuning for the purposes of prediction in this thesis will be explained later in this chapter.

The INTEGRA algorithm augmented with traffic prediction enables the correction of the computed metering rates by the initial INTEGRA algorithm based on the predicted on-ramp traffic demand. The correction is made based on a set of four simple IF-THEN rules. The premise ("If part") of each rule compares the metering rate computed by the initial INTEGRA and the prediction of on-ramp traffic demand of a particular on-ramp. Furthermore, the mentioned part of the rule considers the comparison between critical and current density of the urban motorway segment that is related to a particular on-ramp. The consequence ("Then part") of the rule decreases or increases the metering rate computed by the initial INTEGRA ramp metering algorithm. The difference between the originally computed metering rate and the traffic demand prediction for a particular on-ramp can be subtracted or added to the originally computed metering rate value.

The decision whether or not it is necessary to subtract or add the mentioned difference to the originally computed metering rate is derived based on the two comparisons in the premise part of the particular rule [15]. The first comparison compares the current traffic density with critical density. Based on this comparison it is possible to make a decision whether the traffic flow is in free flow or a congested state. The second comparison compares the current traffic demand with the predicted one. Based on this comparison it is possible to tell whether one can expect an increase or a decrease in traffic demand. Generally speaking, it is possible to conclude that the originally computed metering rate has to be reduced if an increase in traffic demand is expected. The reduction of the same parameter must be made when a congested traffic state is detected. A block scheme of the augmented INTEGRA based on the traffic demand prediction ability is shown in Figure 37.

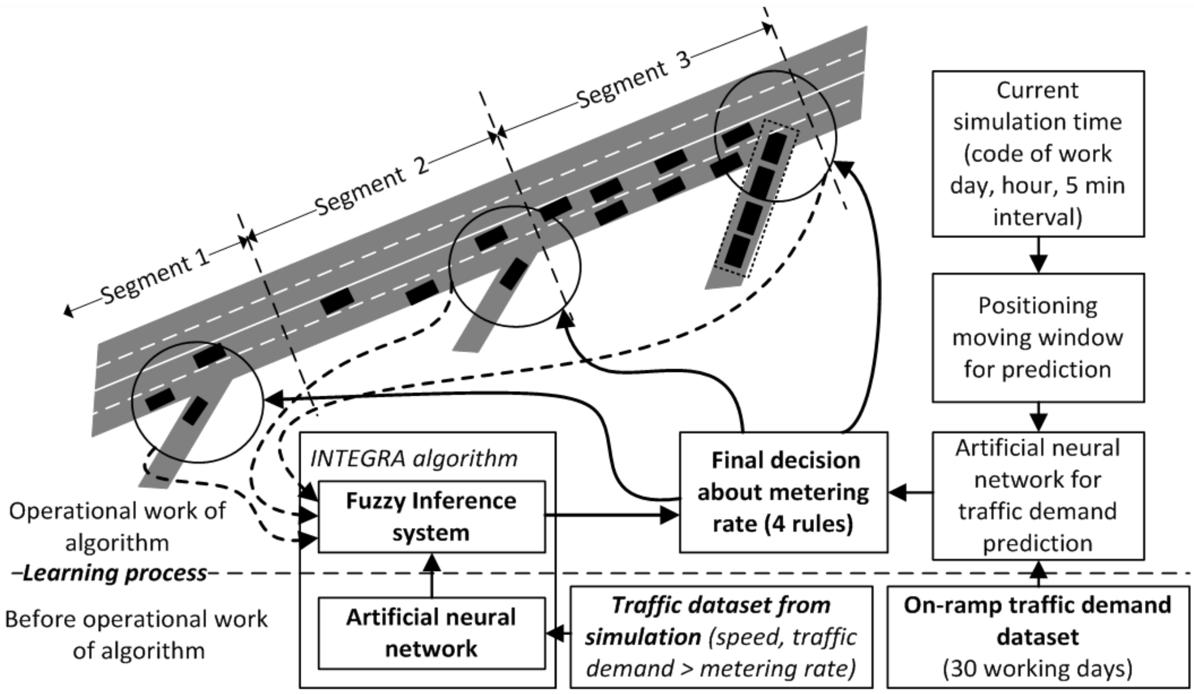

Figure 37: INTEGRA augmented with on-ramp traffic demand prediction

4.3.1. Traffic prediction based on NARX neural network

Traffic prediction of urban traffic flow has become one of the important modules of the ITS based services due to its impact on traffic control and continuous development. Traffic flows presented as time series contain a high amount of randomness and uncertainty. This is the main reason why traditional prediction techniques cannot meet the requirement for precise prediction in practice [66]. High forecast precision is especially important in advanced motorway control methods such as cooperative ramp metering.

In order to predict traffic demand at on-ramps for the purpose of ramp metering and use by the augmented INTEGRA algorithm respectively, it is very important to provide accurate short-term predictions. Firstly, it is imperative to choose an adequate approach to be used for prediction purposes. Traffic flow prediction approaches can be divided into four major categories. The first is based on the analysis of various mathematical prediction models such as the history average model, linear regressive model, Kalman filtering, etc. The second category of models includes knowledge-based intelligent models. They include non-parametric regressive models and specialized types of the ANN. The third group includes various traffic simulations which are used mainly to evaluate existing models. The fourth group contains models based on a combination of several previously mentioned prediction models [90].

The ANN-based models are selected because they provide better prediction results against non-linearity and uncertainty in traffic flow data. In this thesis, the NARX network, a type of RNN, is selected in order to predict long-term and short-term on-ramp traffic demand. Long-term predictions are used for purposes of accuracy testing and short-term prediction for actual use in adjustment of metering rates computed by the initial INTEGRA algorithm.

Short-term traffic flow prediction in general means real-time prediction for the next time interval $t + \Delta t$ (where Δt is less than 15 minutes), and even in later time intervals, based on the previously collected data [90], [91].

The proposed NARX network has 182 neurons in the hidden layer and is trained based on a learning dataset that contains on-ramp traffic demand obtained during 65 working days. The on-ramp traffic demand dataset is arranged in the form of time series. The NARX network predicts on-ramp traffic demand in the form of traffic flow for every on-ramp. The length of the prediction horizon can be changed in order to adapt the prediction to a particular application. The on-ramp traffic demand dataset was obtained from the Zagreb bypass traffic data.A detailed description of the Zagreb bypass use case model will be given in the next chapter.

In order to analyse the prediction performance of an ANN, an interpolated traffic demand dataset from one of the on-ramps of the mentioned use case model is used. The mentioned on-ramp is selected near the Lučko node since it exhibits common daily traffic characteristics (two peak hours) and is affected by a heavy traffic load that is distributed thorough the day. The on-ramp traffic demand dataset is divided into two groups. The first group of 60 working days from this dataset is used for the learning process, while 5 working days are used for validation purposes. Traffic data for Saturday and Sunday are not included in the prediction due to the fact that the traffic demand can be very low during these days, so it is unfeasible to apply predictive ramp metering or ramp metering in general [89]. The prediction inputs are: working day codes (1 - Monday, 2- Tuesday, 3 - Wednesday, 4 - Thursday, 5 - Friday), time of day (1, 2, 3, ... , 24 hours), 5 minute interval code (0, 5, 10, 15, ... , 55) and the traffic demand value data for the observed motorway on-ramp. This data contains traffic demand from *k*, *k-1* and *k-2* step of the simulation run due to the structure of the NARX network. The targeted output vector is defined by the *k+1* and *k+2* step of the simulation run due to the structure of the NARX network.

In the first input it is possible to emphasize unique characteristics of a particular day. Inputs related to the time of day and the 5-minute interval code enable the ANN to distinguish different parts of each day during the learning process. Using this additional inputs it is possible to increase the prediction accuracy of the existing approaches in on-ramp traffic demand prediction [90]. The resilient back-propagation method is used as the learning method.

Prediction results for long-term predictions are analysed by using 5 working days in the validation dataset. The prediction horizon is set to 10 minutes. The prediction results for 5 working days by using 10-minute prediction horizon are graphically presented in Figure 38.

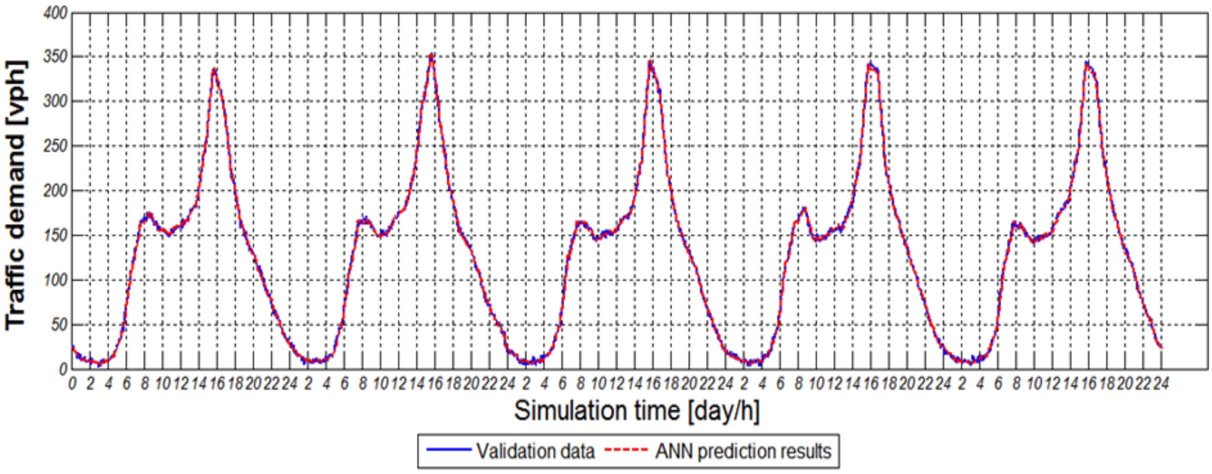

Figure 38: Graphical representation of prediction results for 5 working days by using 10-minute prediction horizon [89]

In the accuracy analysis of the mentioned case, the NARX network reached a 2.60 RMSE. Furthermore, the mentioned ANN reached a 2.05 Mean Absolute Error (*MAE*) and a 0.05 Mean Relative Error (*MRE*) value [89]. The MAE can be calculated according to the following equation:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|f_i - \hat{f}_i|, \tag{33}$$

where f_i is the real traffic demand, $\hat{f}_i$ is the predicted value of traffic demand, and *n* represents the number of the predicted traffic demand intervals. The MRE can be calculated according to the following equation:

$$MRE = \frac{1}{n}\sum_{i=1}^{n}\frac{|f_i - \hat{f}_i|}{f_i}. \tag{34}$$

In subsequent analyses, several short-term and long-term predictions are made using the same NARX network setup. The NARX network is trained for a 5, 60 and 75 minutes time period. Predictions are made for 30 minutes, 24 hours and five working days. A 5, 10, 15 and 30-minute prediction horizons are used for each of the mentioned training sequences and prediction lengths. In Table 2 the results of the NARX network prediction performance for different training times, prediction horizons and prediction lengths can be seen.

Table 2: NARX network prediction performance for different learning times, prediction horizons and prediction lengths [35]

Prediction length	Prediction horizon [min]	NARX (67 days, 5 $[min]$ learning)			NARX (67 days, 60 $[min]$ learning)			NARX (67 days, 75 $[min]$ learning)		
		RMSE	*MAE*	*MRE*	*RMSE*	*MAE*	*MRE*	*RMSE*	*MAE*	*MRE*
$30\ [min]$	5	0.75	0.58	0.03	0.35	0.25	0.01	1.17	1.05	0.06
	10	1.49	1.32	0.06	2.03	1.68	0.08	1.99	1.75	0.09
	15	2.35	1.84	0.08	2.66	2.17	0.10	2.62	2.42	0.13
	30	4.50	4.19	0.19	5.73	5.47	0.25	3.44	3.36	0.18
$24\ [h]$	5	5.29	4.25	0.09	154.19	91.34	0.88	31.54	19.89	0.18
	10	6.48	5.14	0.09	1279.98	715.39	6.24	31.76	19.26	0.16
	15	8.71	6.69	0.10	1176.25	657.97	5.74	35.01	20.98	0.16
	30	16.93	12.82	0.15	1909.96	1063.11	9.34	41.53	23.59	0.17
$5 \times 24\ [h]$	5	6.30	4.73	0.08	266.34	236.53	8.38	35	21.10	0.19
	10	7.81	5.66	0.09	2153.79	2003.49	51.36	35.49	20.75	0.17
	15	10.38	7.16	0.09	1981.43	1843.24	47.13	39	22.72	0.17
	30	19.30	13.21	0.15	3215.14	2991.56	77.75	46.61	26.56	0.19

The data in Table 2 indicates that the prediction accuracy for a longer period of time (prediction length) is lower in comparison with a 30-minute prediction length, regardless of prediction horizons. Shorter learning time produces better results in terms of long-term prediction. The best results for short-time prediction (30-minute prediction length) are produced when using the 5-minute prediction horizon with the NARX network that is learned for 60 minutes. This parameter of the NARX network will be used for INTEGRA augmentation. In Figures 39, 40 and 41, a graphical representation of the results regarding prediction duration and prediction horizon for the NARX network which learning for 60 minutes can be seen.

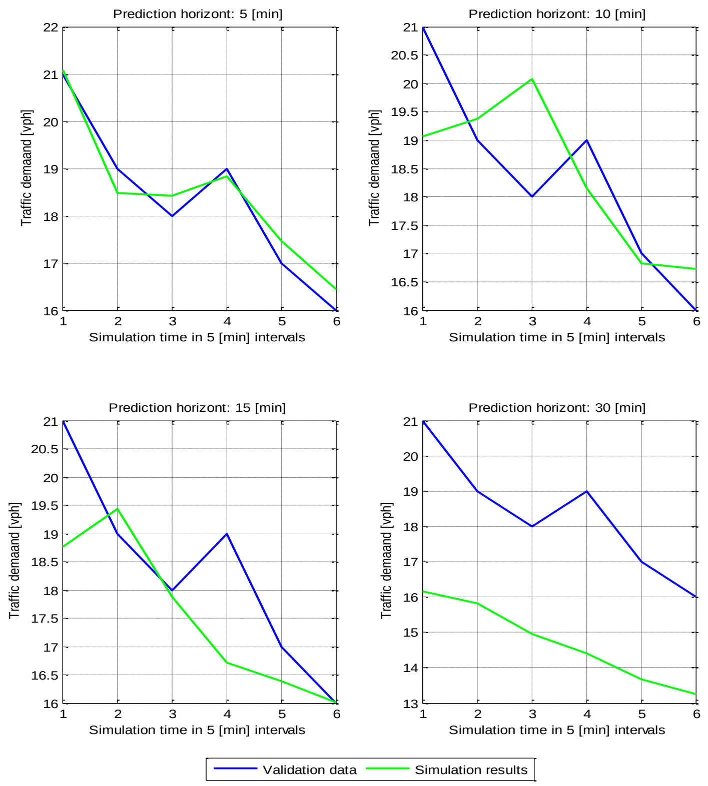

Figure 39: Predictions of traffic demand at an on-ramp for 30 [min] prediction length.

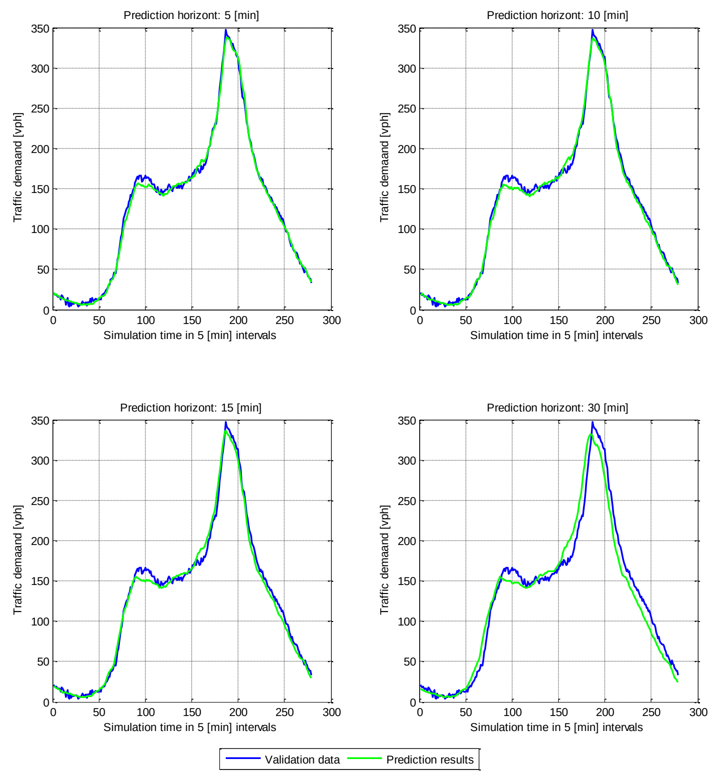

Figure 40: Predictions of traffic demand at an on-ramp for 24 [h] prediction length.

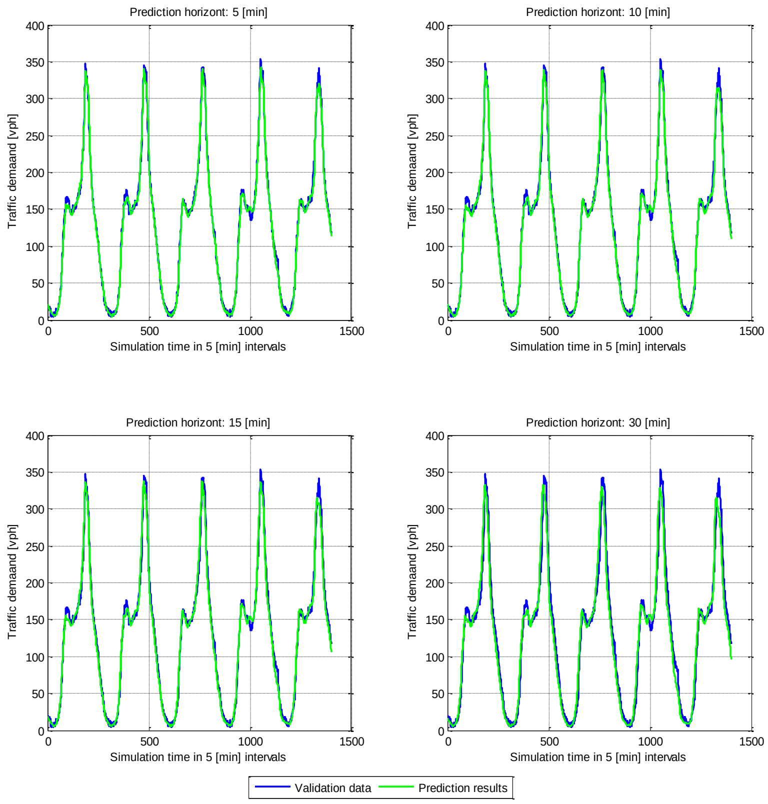

Figure 41: Predictions of traffic demand at an on-ramp for five working days prediction length.

In Figures 39, 40 and 41 the most prominent difference in prediction accuracy between 5 minutes and 30 minutes prediction horizons can be observed. This data provides an additional graphical proof that a shorter prediction horizon can produce on-ramp traffic demand predictions with higher precision. Results presented in Figure 39 for the 5 minute prediction horizon will be the focus of the next section of this thesis since the original INTEGRA is augmented in order to use short-time predictions.

4.3.2. Integration of the predictive function and the INTEGRA algorithm

As was mentioned earlier, the INTEGRA algorithm is augmented in order to take into account on-ramp traffic flow prediction. An adequate control module is created in order to enable correction of computed metering rates obtained by the initial INTEGRA algorithm. Corrections of the initially computed metering rates are based on the predicted traffic demand at a particular on-ramp. Traffic demand at on-ramps is predicted in time intervals of 30 minutes by using a 5-minute prediction horizon. Corrections of metering rates computed by the initial INTEGRA algorithm are made based on a set of four simple IF-THEN rules. The premise of each rule compares the metering rate computed by the initial INTEGRA and an on-ramp traffic demand prediction associated with a particular on-ramp.

The aforementioned part of each rule considers the comparison between the critical density and the current density of a particular motorway segment. The consequence part of the rules increases or decreases the value of the metering rate computed by the initial INTEGRA algorithm. The difference between the originally computed metering rate and the on-ramp traffic demand prediction for a particular on-ramp can be added to or subtracted from the mentioned metering rate value. This action is taken with respect to the comparisons which were made in the premise of a particular rule. The metering rate at a specific on-ramp is determined by four rules or cases [89]:

Case 1.

$$r_{i[k]} = \begin{cases} r_{i[k]}^{integra} + \left[r_{i[k]}^{integra} - \left(k_p d_{i[k]}^{p}\right)\right] & if\ d_{i[k]}^{p} < r_{i[k]}^{integra} \wedge n_{i[k]} < \bar{n}_{i[k]} \\ Case\ 2, & otherwise \end{cases}$$

where $r_{i[k]}^{integra}$ is the metering rate computed by the initial INTEGRA algorithm in cell *i*, during simulation time step *k*, $d_{i[k]}^{p}$ is the predicted traffic demand for the on-ramp in cell *i*, during simulation time step *k*, and k_p is the coefficient of prediction impact. The coefficient of prediction impact is set to 0.25.

Case 2.

$$r_{i[k]} = \begin{cases} r_{i[k]}^{integra} - \left[\left(k_p d_{i[k]}^{p}\right) - r_{i[k]}^{integra}\right] & if\ d_{i[k]}^{p} > r_{i[k]}^{integra} \wedge n_{i[k]} < \bar{n}_{i[k]} \\ Case\ 3, & otherwise \end{cases} \quad (33)$$

Case 3.

$$r_{i[k]} = \begin{cases} r_{i[k]}^{integra} - \left[r_{i[k]}^{integra} - \left(k_p d_{i[k]}^{p}\right)\right] & if\ d_{i[k]}^{p} > r_{i[k]}^{integra} \wedge n_{i[k]} > \bar{n}_{i[k]} \\ Case\ 4, & otherwise \end{cases} \quad (34)$$

Case 4.

$$r_{i[k]} = \begin{cases} r_{i[k]}^{integra} + \left[\left(k_p d_{i[k]}^{p}\right) - r_{i[k]}^{integra}\right] & if\ d_{i[k]}^{p} < r_{i[k]}^{integra} \wedge n_{i[k]} > \bar{n}_{i[k]} \\ r_{i[k]}^{integra}, & otherwise \end{cases} \quad (35)$$

5. Simulation results

This chapter will primarily be a discussion about the physical setup and traffic data used for the considered case model. The section between the nodes Jankomir and Lučko on the Zagreb bypass is selected for the *use case* model. This section of the Zagreb bypass can be considered an urban motorway due to its heavy traffic load (especially during the summer tourist season in Croatia), its proximity to the urban area of Zagreb (heavy recurrent congestions), and its specific constructional parameters. The heavy traffic load induced as recurrent traffic congestions, is especially noticeable during the afternoon peak hour which appears suddenly. The described traffic scenario, characterized by a sudden increase in daily traffic demand, is interesting as a use case scenario for studying the effect of motorway control methods. Congestion on this particular urban motorway section quickly reaches its maximum strength which makes this scenario suitable for a "stress test" of the motorway control methods being tested. A similar traffic scenario can arise during the summer tourist season, especially near the Lučko node where a motorway tollbooths are installed.

Constructional parameters of the mentioned sections are similar to other urban motorways near larger urban areas. In line with that statement, the mentioned use case model contains a lot of on- and off-ramps that are fairly close to each other. This setup will make a suitable testbed for all the analysed motorway control methods against the presence of a dominant dependency between on-ramp traffic flows which is characteristic of urban motorways.

A comparative analysis of commonly used urban motorway control methods that will be implemented in the mentioned use case model will be carried out. Special emphasis will be set on a cooperative approach between ramp metering and the VLSC and on both proposed versions of INTEGRA algorithms (with and without the traffic prediction function). The commonly used urban motorway control methods can be divided into standalone applications of ramp metering, standalone VSLC and the no control method. The following ramp metering algorithms will be used in the comparative analysis: ALINEA (local), SWARM (competitive) and HELPER (cooperative). Two types of VSLC algorithms are used: the temporal reactive VSLC (VSLCTR) and the density reactive VSLC (VSLCDR). Additionally, this section will include a comparative analysis of the results achieved by the INTEGRA algorithm using different criteria functions parameters for additional data selection from the initial learning dataset. A special emphasis will be set on analysing the results achieved by the cooperative approach between ramp metering, the VSLC and the INTEGRA algorithm since they are within the scope of this thesis. In order to assess and compare results achieved by all the analysed urban motorway methods, several Measures of Service (MoS) will be used. MoSs are used in order to assess urban motorway LoS. The MoS measures are explained in detail in Appendix 1.

5.1. Simulation setup and use case model

The Zagreb bypass is an urban motorway with marked seasonal overloads. As was mentioned earlier, the most significant problem occurs near the Lučko node due to waiting queues at tollbooths. These can induce intense and fast backpropagation of shockwaves which can consequently lead to vehicle queues that can reach more than 10 [km]. If the section between the Jankomir and the Lučko nodes is viewed in the context of urban motorways, the Lučko node has already become a part of the Zagreb urban road system. The fact that about 70% of traffic in this node is generated by the nearby city of Zagreb [92] supports this claim. The section between the Jankomir and Lučko nodes of the Zagreb bypass was used as the use case model due to the combination of increased traffic load during the entire day, long lasting increased traffic load and the significant effect of daily migrations during the afternoon peak hour. The impact of heavy congestion on traffic flows of this section can be used for another traffic scenario with heavy traffic load, e.g. studying the sudden and heavy increase of traffic load during the summer tourist season. This section can be seen in Figure 42.

Figure 42: The section between the Jankomir and Lučko nodes of the Zagreb bypass [92]

In Figure 43, a representation of the modeled section of the Zagreb bypass in The Google Maps tool (Accessed: 06. February, 2017.) can be seen. According to the data from the Google trip planner application (TRANSIT), an average of 4 – 6 minutes is required to navigate the section in a situation without a heavy traffic load. In this particular case, mild traffic congestion is detected near the Jankomir node. The total length of this section is 6.6 [km].

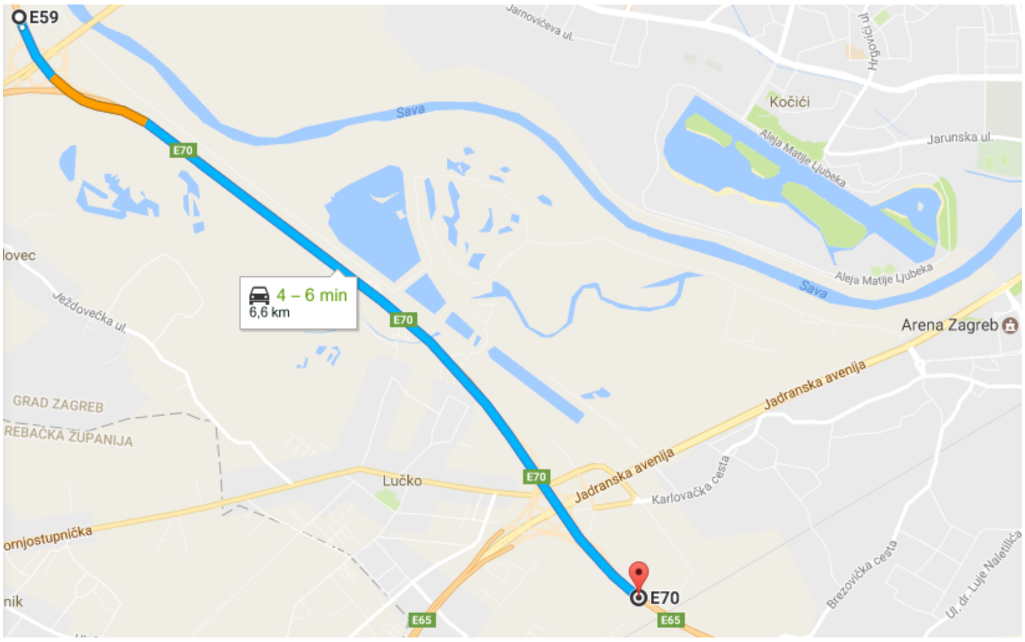

Figure 43: A representation of a Zagreb bypass section modeled in the Google Maps tool with the trip planner applied

The use case model used in this study considers only the traffic scenario with a heavy traffic load. The use case model will illustrate the impact of heavy traffic load on the Lučko node and on the Jankomir node. The impact of the heavy traffic load on the Lučko node is especially interesting since this node contains tollbooths and is directly connected with the Jadranska Avenija, which can be considered as the arterial road of the Zagreb urban traffic network. As it was mentioned earlier, during the tourist season, the waiting queues at the tollbooths and the corresponding waiting time can be very long.

The use case model is modelled and simulated in the augmented CTMSIM simulation environment in order to verify the functionality of the newly developed motorway control methods. Constant variables of the Zagreb bypass section model are related to its physical parameters. The physical parameters of the Zagreb bypass can be seen in Table 3.

Table 3: Physical parameters of the section between nodes Jankomir and Lučko on the Zagreb bypass

Number of the cell	Name of the cell	Length of the cell [miles]	Length of the cell [m]	Number of lanes	On-ramp	Off-ramp	Implemented VSLC
1	Lučko 1	0.29	465.13	3	1	1	No
2	Lučko 2	0.16	256.62	2	0	1	No
3	Lučko 3	0.13	208.51	2	1	1	No
4	Lučko 4	0.23	368.90	2	1	1	No
5	Lučko 5	0.2	320.78	2	1	1	No
6	Streach 1	0.2	320.78	2	0	0	No
7	POUPlitivice1	0.44	705.72	2	1	1	Implemented
8	Streach 2	0.5	801.95	2	0	0	Implemented
9	POUPlitivice2	0.24	384.94	2	1	1	No
10	Streach 3	0.83	1331.24	2	0	0	No
11	Jankomir 1	0.45	721.76	2	1	1	Implemented
12	Jankomir 2	0.12	192.47	3	1	1	No
13	Jankomir 3	0.1	160.39	3	1	1	No
14	Jankomir 4	0.24	384.94	2	1	1	No

A physical model of the Zagreb bypass is created based on 14 cells (10 cells have on-ramps, and 11 cells have off-ramps). The maximum capacity of every on-ramp is 600 [vph], while the maximum capacity of every mainstream cell depends on its length, number of lanes, etc. The constant variables also define the fundamental diagram for every mainstream cell. The variables of the motorway model are traffic demand (presented as a traffic flow) on every on-ramp and model input and output flows, [52]. Additionally, the mentioned section contains many on- and off-ramps close to each other, making it suitable for the implementation of the proposed cooperative control method due to the increased dependency between on- and off-ramps [92]. One motorway node, from the point of view of the macroscopic traffic models, contains several cells with on- and off-ramps, which have to be close to each other. The Jankomir and Lučko nodes contain the majority of cells with on- and off-ramps on the mentioned urban motorway section. The first five cells of the mentioned urban motorway model are parts of the Jankomir node, while the last four cells are parts of the Lučko node.

The on-ramp traffic demand characteristics of the Zagreb bypass simulation model is reconstructed using the daily characteristics of the Ljubljana bypass traffic. The traffic data is transformed in the form of a traffic demand dataset for each on-ramp separately. In order to adjust the daily traffic demand characteristic, the average daily traffic values from [92] are used to ensure that the daily vehicle number describes the traffic demand of the Zagreb bypass realistically, [52]. The urban motorway model in-flow and out-flow curve are shown in Figure 44. In Figure 45, the traffic demand for every on-ramp on the use case model is presented.

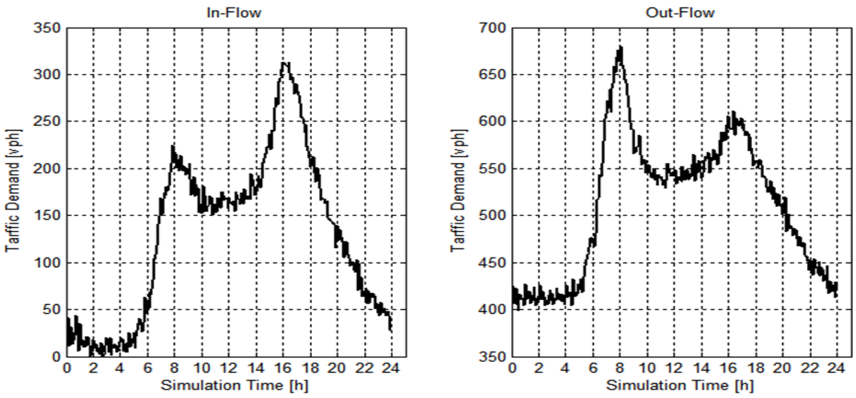

Figure 44: In-flow and out-flow data used in the use case model

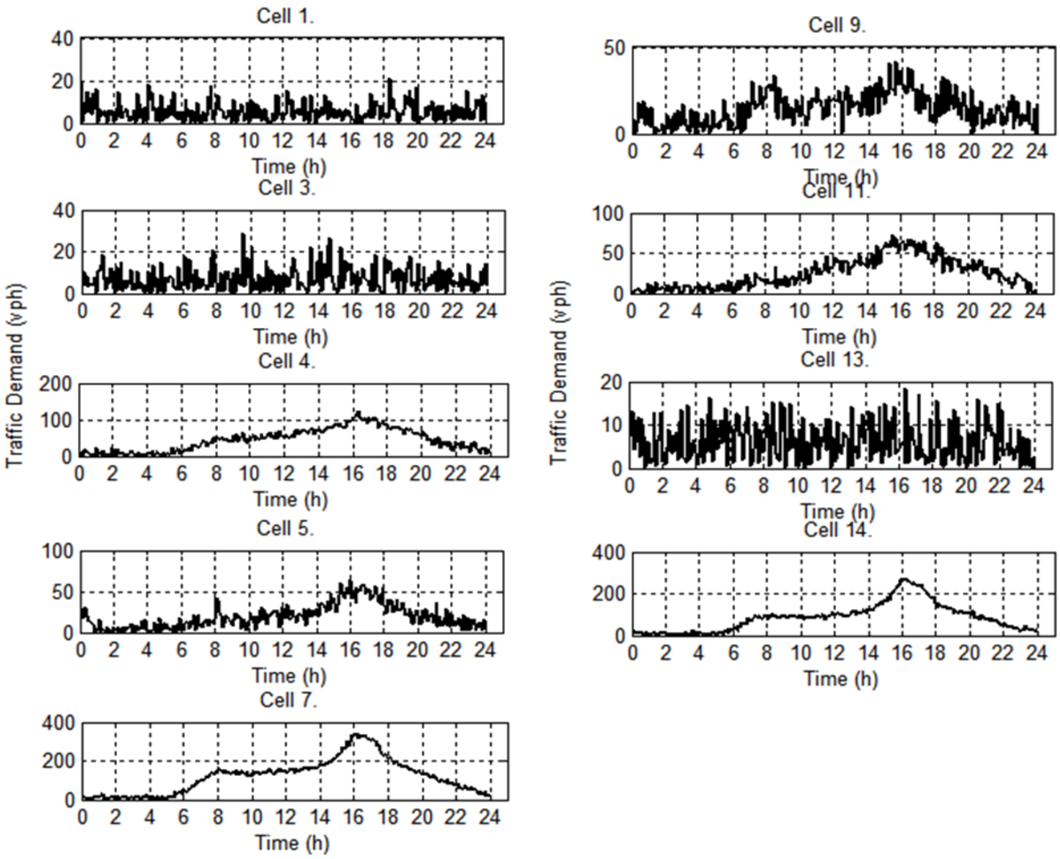

Figure 45: Traffic demand for on-ramps on the use case model.

In order to verify the operational work of the cooperative motorway control strategies, the 13th cell is set to generate high traffic demand. This creates downstream congestion resulting in a "shock wave" propagating upstream. This is done in order to observe the creation of upstream virtual queues during the simulation. These observations will be useful in the evaluation of the INTEGRA algorithm learned control properties.

5.2. Cooperation between ramp metering and VSLC

In order to implement direct cooperation between ramp metering and the VSLC, it is imperative to develop a ramp metering algorithm which reacts to changes in traffic parameters. In this thesis a VSLC algorithm which will compute speed limits based on the changes in traffic density in the motorway mainstream flow will also be developed. The mentioned VSLC algorithm will be named density reactive VSLC (VSLCDR). The VSLCDR algorithm computes the change in a posted speed limit value by using four different conditions. The conditions under which the VSLCDR algorithm changes speed limit values are based on fundamental diagram parameters which describe traffic behaviour on the motorway segment with the VSLC applied. The fundamental diagram is divided into four segments. Each of the mentioned fundamental diagram segments is defined by a specific range of traffic densities. Furthermore, each segment represents a specific state of the traffic flow for the observed motorway segment. The measured traffic density in a particular motorway segment can be allocated to one of the mentioned fundamental diagram regions.

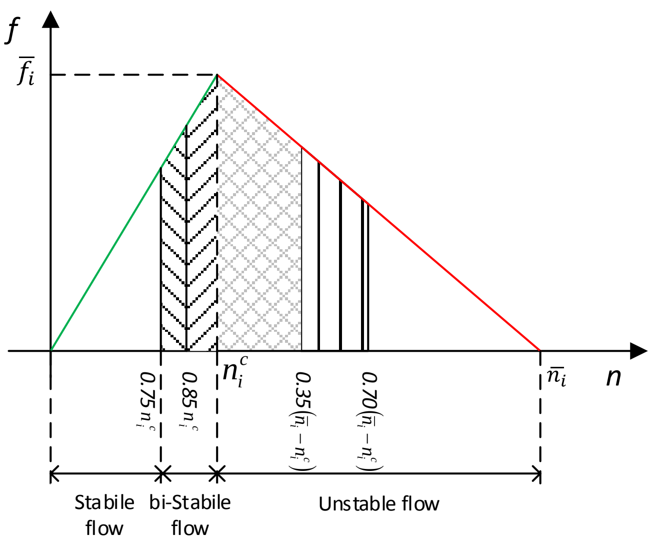

Figure 46: A representation of the fundamental diagram divided into segments based on which speed limit will be computed

An adequate speed limit will be assigned according to the value of the currently-measured traffic density. In Figure 46, the fundamental diagram is divided into four segments based on which speed limit will be computed.

The initial speed of the motorway mainstream traffic flow is set to 130 [km/h] (~81 mph). This is the maximum speed allowed on Croatian motorways. The speed limit values will be assigned according to the currently measured density that can be allocated in the segments defined by following borders:

- If [0.75 n_i^c, 0.85 n_i^c], then the speed limit of the mainstream flow will be set to 110 [km/h] (~50 mph);
- If [0.85 n_i^c, n_i^c] then the speed limit of the mainstream flow will be set to 90 [km/h] (~43 mph);
- If [g_k, 0.35 $(\bar{n}_i - n_i^c)$] then the speed limit of the mainstream flow will be set to 70 [km/h] (~37 mph);
- If [0.35 $(\bar{n}_i - n_i^c)$], 0,70 $(\bar{n}_i - n_i^c)$] then the speed limit of the mainstream flow will be set to 50 [km/h] (~31 mph);

The average speed on the section between Lučko and Jankomir on the Zagreb bypass is 95 [km/h]. According to the use case model, the VSLCDR algorithm has the ability to increase or decrease speed by a maximum of 20 [km/h]. This increment/decrement of a posted speed limit is selected according to the paper [93] published at the European Control Conference, and a paper [94] that described speed limit field tests by the same increment/decrement on Dutch motorways. The latter paper was published at the ITS World Congress in 2011. Both papers can be considered as good European guidelines for implementing the VSLC. The parameters of each segment are specified according to the research described in [93] and a series of experimental simulations on the current use case model.

The VSLCDR algorithm contains an imposed constraint which lets the value of a posted speed limit be valid for a minimum 10-minute interval. This time interval represents two time steps in the CTMSIM simulator. The mentioned constraint is very important since it is preventing that drivers do not change their speed too frequently. Frequent changes in speed limits can bring about less effective speed homogenization and a less comfortable driving experience which drivers can find annoying and therefore tend to ignore the speed limits. Another constraint of the VSLCDR algorithm is created in order to prevent frequent on / off algorithm switching. This constraint is especially effective during successive time steps in cases of a large fluctuations in traffic demand. Hysteresis is introduced in order to avoid negative effects of these fluctuations. In control systems, hysteresis can be used to filter input signals so that the output actions react less rapidly than they otherwise would, by taking into account recent history [95]. Basically, it switches an output between two constants. The region between the first border for turning a VSLCDR algorithm on or off is 0.75 n_i^c and the experimentally determined 0.75 n_i^c border will be used in order to switch among one of the two possible outputs. This will eliminate frequent oscillation between the VSLCDR algorithm being on or off.

The value of the speed limit cannot increase its value rapidly in case of a sudden increase in traffic demand. In the case of a rapid increase in traffic demand at on-ramps, the VSLCDR algorithm can "jump" from one fundamental diagram region to another in two successive time steps. This "jump" induces a double or larger decrement/increment of the speed limit. The described fluctuations of traffic demand are common on urban motorways. The mentioned traffic scenario with the rapid increase of traffic demand is incorporated into the use case model as it was described in the previous section.

Users of the urban motorway (drivers) would not comply with a posted speed limit if it were changed rapidly in short time intervals. This can be interpreted by motorway users as a system error, as "unfair" or "unnecessary" from their perspective which is pretty narrow since they can observe only small portions of the urban motorway from their vehicles. In order to mitigate this problem, the speed limit value can be increased or decreased only by a previously mentioned predefined value. This action will be conducted after the execution of a constraint which governs the minimum duration of a posted speed limit. Basically, the VSLCDR algorithm will gradually achieve a desired speed limit over several successive time steps by increasing the current speed limit with a predefined value. This sort of VLSCDR behaviour will enable a smooth transition from the current average mainstream speed to the desired one. In this case, the transition should be conducted without drastic changes in the speed limit which could be noticed by drivers.

In Figure 47, it is possible to see graphs that represent resulting mainstream speeds after the implementation of the VSLCDR algorithm. The dotted lines are introduced in this graph in order to represent periods when the VSLCDR algorithm is turned on and is posting speed limits. In other cases, it is turned off and shows the maximum possible mainstream speed on the motorway. Furthermore, it is possible to see additional cells 9 and 12, which are directly affected by the traffic flow originating from the cells affected by the VSLCDR algorithm. As it was mentioned earlier, the VSLC is applied in cells 7, 8, and 10. Additionally, the graphs depicting the cells affected by the VSLC are magnified in order to provide better insight into the VSLCDR algorithm behaviour. The magnified parts of the graph are related to the most prominent operational effect of the VSLCDR algorithm. The results of the standalone VSLCDR algorithm are compared to a no-control situation on the urban motorway.

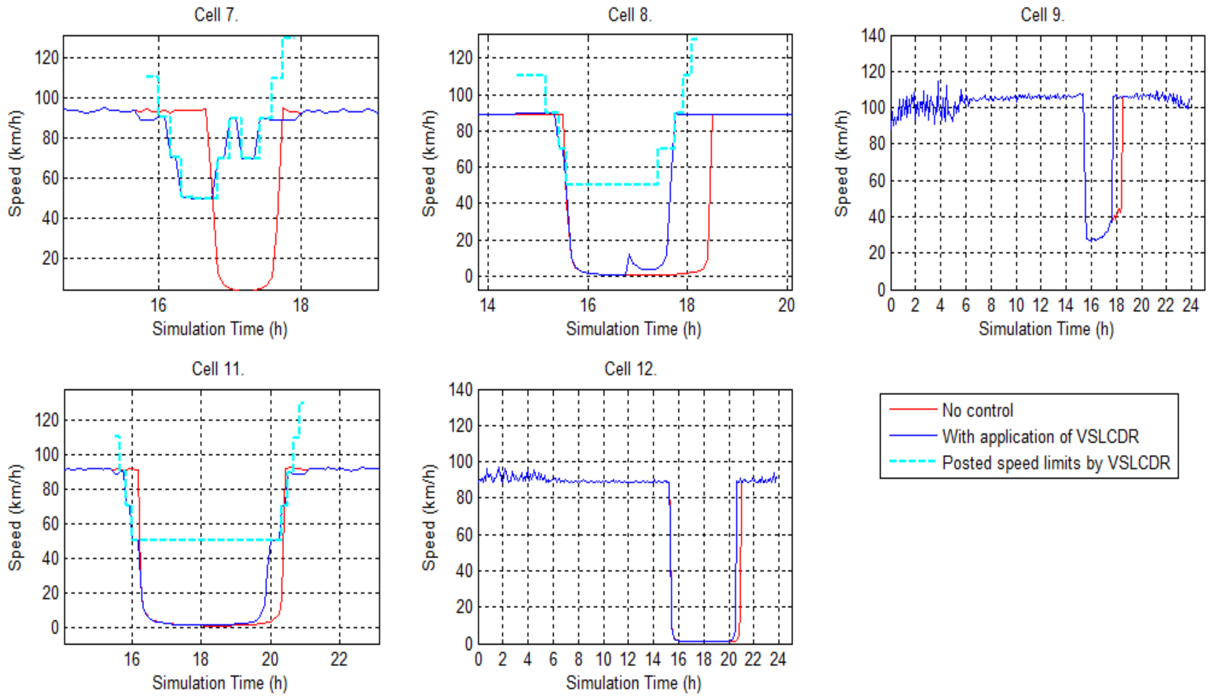

Figure 47: Resulting mainstream speeds after the implementation of the VSLCDR algorithm

According to Figure 47, it is possible to conclude that the VSLCDR algorithm manages to reduce the duration of congestion in every cell used in the analysis. In cell 7, congestion can be completely avoided by using the VSLC since the density in this cell is much lower compared to the other cells. The VSLCDR achieves the mentioned results by reducing the mainstream speed slightly before congestion arises. In another words, the speed limit reduces the mainstream speed when bi-Stabile traffic flow is detected. It is possible to conclude that within the bi-Stabile traffic state there is a critical time to react and reduce mainstream speed. In the fundamental diagram, the bi-Stabile traffic state is divided into the two segments in order to implement the VSLCDR algorithm. This action allows a smoother increase of a speed limit as congestion is forming. In Figure 48, it is possible to see a comparison of mainstream density achieved without control and with the application of the VSLCDR algorithm.

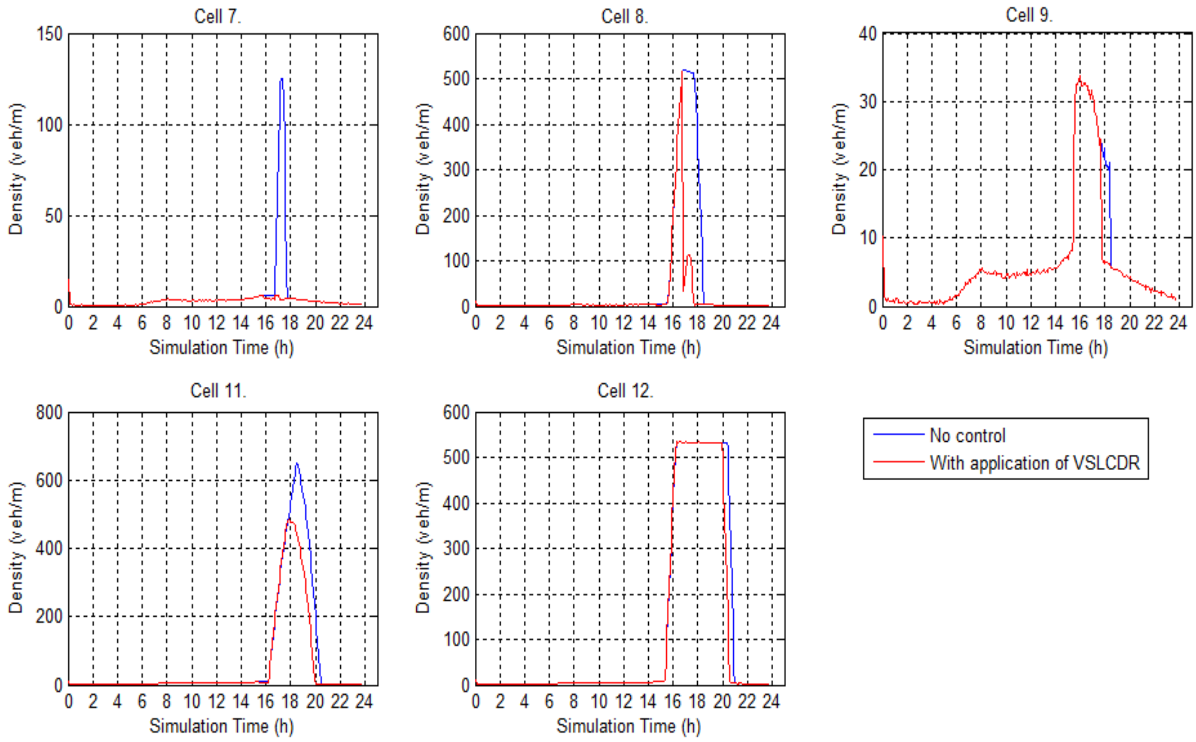

Figure 48: Comparison of mainstream density achieved by no control and with application of the VSLCDR algorithm

In Figure 48 a low mainstream density in the case of the VSLCDR algorithm implementation can be observed. The reduction of traffic congestion via the VSLCDR is especially noticeable after congestion starts to clear out. The reduction of speed in the case of congested traffic can provide a smoother transition towards a stable traffic state.

The VSLCDR algorithm will cooperate with the HELPER ramp metering algorithm directly. The HELPER ramp metering algorithm is chosen for cooperation because its main task is to reduce the impact of shockwave backpropagation effect on upstream mainstream flows by using several upstream on-ramps as “slave” on-ramps. These “slave” on-ramps provide a reduction of the additional inflow from the on-ramps into the upstream mainstream flow. It is possible to conclude that the HELPER algorithm can effectively reduce the impact of downstream congestion on the upstream traffic flow. The VSLCDR is tasked primarily with gradually slowing down mainstream flow at several upstream motorway sections. Additional slowdowns induced by the VSLCDR reduce the mean mainstream speed, and therefore the number of incoming vehicles to the place of congestion.

Direct cooperation between the VSLCDR and the HELPER ramp metering algorithm in a CTMSIM environment is explained in detail in section 3.1.2. It is necessary to mention that there are two specialized variables which the VSLCDR and the HELPER ramp metering algorithms exchange in each simulation step by using previously mentioned cooperative augmentation of the CTMSIM simulator. The HELPER ramp metering algorithm sends theVSLCDR algorithm a vector that indicates the activation status of the “master” and “slave” on-ramps. If the cooperative logic added to the VSLC algorithm detects the formation of the “master” and “slave” on-ramps, it will reduce the current computed speed limit by 20 [km/h] with respect to all built-in constraints. The posted speed limit cannot be lower than 50 [km/h].

The VSLCDR algorithm also sends a vector to the HELPER ramp metering algorithm in each simulation step. This vector contains data about the current categorization of the measured traffic density for each cell, which is initially used by the VSLCDR algorithm. The cooperative logic added to the HELPER ramp metering algorithm performs adjustment of previously computed metering rates according to the vector acquired form the VSLCDR algorithm. It is also important to emphasize that the provided vector contains the currently measured density categorization for cells with the implemented VSLCDR algorithm. If the density of all cells is in category zero, this means that there is no need for further adjustment of metering rates. In all other cases, the following adjustments of the metering rates are required (if the HELPER ramp metering algorithm has detected congestion and started to assign “master” and “slave” on-ramps):

- If [0.75 n_i^c, 0.85 n_i^c], then the categorization of density is 1. The HELPER ramp metering algorithm needs to decrease metering rates at “master” and “slave” on-ramps by 8 %;
- If [0.85 n_i^c, n_i^c] then the categorization of density is 2. The HELPER ramp metering algorithm needs to decrease metering rates at “master” and “slave” on-ramps by 12 %;
- If [g_k, 0.35 $(\bar{n}_i - n_i^c)$] then the categorization of density is 3. The HELPER ramp metering algorithm needs to decrease metering rates at “master” and “slave” on-ramps by 17 %;
- If [0.35 $(\bar{n}_i - n_i^c)$], 0.70 $(\bar{n}_i - n_i^c)$] then the categorization of density is 4. The HELPER ramp metering algorithm needs to decrease metering rates at “master” and “slave” on-ramps by 20 %;

The values for therequired metering rates reductions are derived by running several experimental simulations. The best results are achieved by the hereby presented setup. It is important to mention that the cooperative logic of the HELPER ramp metering algorithm is able to increase the metering rate by the same percentage. This action will be taken if the density categorization is decreasing in two or more successive time steps, i.e. when congestion is dissolving.

Figure 49 shows mainstream speeds brought about by cooperation between the VSLCDR and the HELPER ramp metering algorithm in the cells which are relevant for the VSLC. Furthermore, these results are compared with standalone applications of the VSLCDR algorithm and the HELPER ramp metering algorithm. Dotted lines are introduced in the same graph in order to represent time intervals in which the VSLCDR algorithm is turned on/operational. The light blue dotted line presents computed speed limits by the standalone VSLCDR algorithm. The magenta dotted line presents the speed limits computed by the VSLCDR algorithm in cooperation with the HELPER ramp metering algorithm. Additionally, graphs of the cells affected by the VSLC are magnified in order to provide better insight into the VSLCDR algorithm behaviour in both modes of implementation. All the results achieved are compared to the situation without any control on the urban motorway.

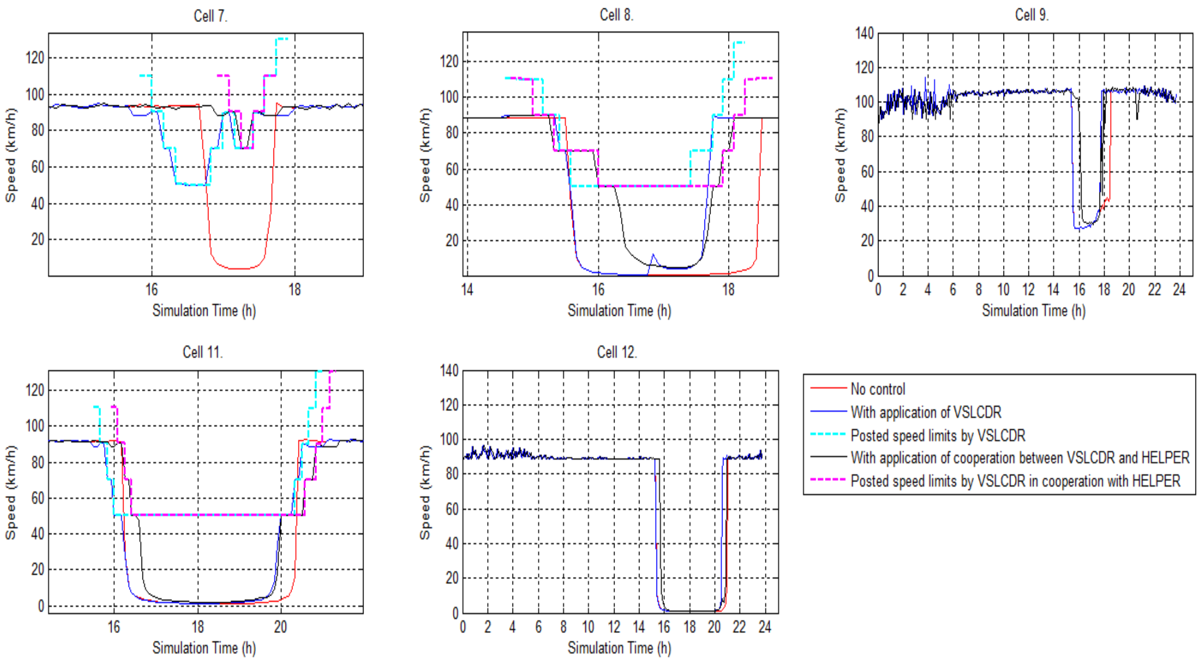

Figure 49: Achieved mainstream speeds in the cells relevant for the VSLC produced by the cooperation between the VSLCDR and the HELPER ramp metering algorithm

Figure 49 points to the conclusion that the VSLCDR algorithm which cooperated with the HELPER ramp metering algorithm preserved its ability to reduce the duration of congestion. The duration of congestion time is additionally reduced by the HELPER ramp metering algorithm's feature which enables the formation of virtual upstream on-ramp queues. The cooperation between the HELPER ramp metering algorithm and the VLSCDR algorithm can postpone congestion impact on motorway traffic flows by acting as congestion starts to form. This effect can be seen in cell 8 and cell 11. This is done due to the influence of the HELPER ramp metering cooperating with the VSLCDR algorithm. As was explained earlier in this thesis, the HELPER ramp metering algorithm creates virtual queues at upstream on-ramps known as "slave" on-ramps. This reduces the input of traffic flow from upstream on-ramps to the mainstream traffic flow. Information on the bi-Stabile traffic flow detected by the VSLCDR algorithm will be pieced together by its cooperative module and written into the vector which will be passed to the HELPER ramp metering algorithm. According to that information the cooperative module of the HELPER ramp metering algorithm will additionally reduce the value of metering rates. On the other hand, the HELPER ramp metering algorithm passes information about the activation of "slave" on-ramps to the VSLCDR algorithm which induces an additional reduction of the speed limit by the cooperative logic of the VSLCDR algorithm. It is necessary to mention that the VSLCDR will gradually increase speed limits according to the aforementioned constraints. It is also possible to conclude that "virtual" queues induced by the HELPER ramp metering algorithm create a discharge effect after congestion is cleared. This effect can be noticed in cells 8 and 11. The mentioned effect produces a form of slight offset in the case of a speed limit increase between a standalone VSLCDR and the one which is involved in the cooperation with the HELPER ramp metering algorithm. In Figure 50, it is possible to see a comparison of the metering rates from on-ramps with expressed on-ramp demand. Those metering rates are produced by the HELPER ramp metering algorithm which is cooperating with the VSLCDR algorithm and by its standalone operational work.

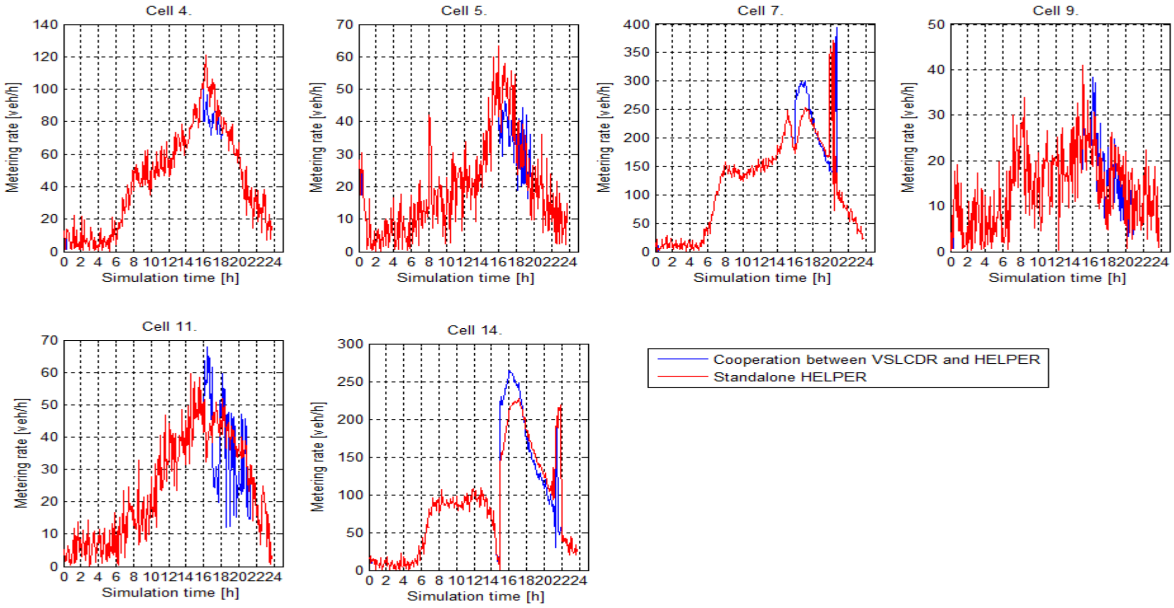

Figure 50: A comparison of the metering rates from on-ramps with expressed on-ramp demand produced by the HELPER ramp metering algorithm in cooperation with the VSLCDR algorithm and by standalone operational work

According to Figure 50, it is possible to conclude that HELPER in cooperation with the VSLCDR additionally decreases the values of metering rates at "slave" on-ramps 4 and 5. This action additionally reduces vehicle in-flow into the downstream mainstream. Furthermore, the VSLCDR additionally reduces the mainstream speed in that section (cells 7 and 8) of the motorway which produces a lower impact of congestion at the "master" cell 7, and enables higher metering rates on other downstream on-ramps. Those cells are not heavily affected by congestion because congestion is checked in the upstream part of the mainstream by using the VSLC and lower metering rates at "slave" on-ramps. All adjustments of metering rates are conducted during the congestion period. This proves its valid operational work which is adequately adjusted by the cooperative logic.

In Figure 51, a comparative analysis of mainstream densities achieved by cooperation between HELPER ramp metering algorithm and VSLCDR algorithm can be seen. Furthermore, results regarding mainstream density derived from standalone applications of HELPER and the VSLCDR algorithm, and situations that involves parallel work of HELPER and the VSLCDR algorithm are also included in the mentioned comparative analysis.

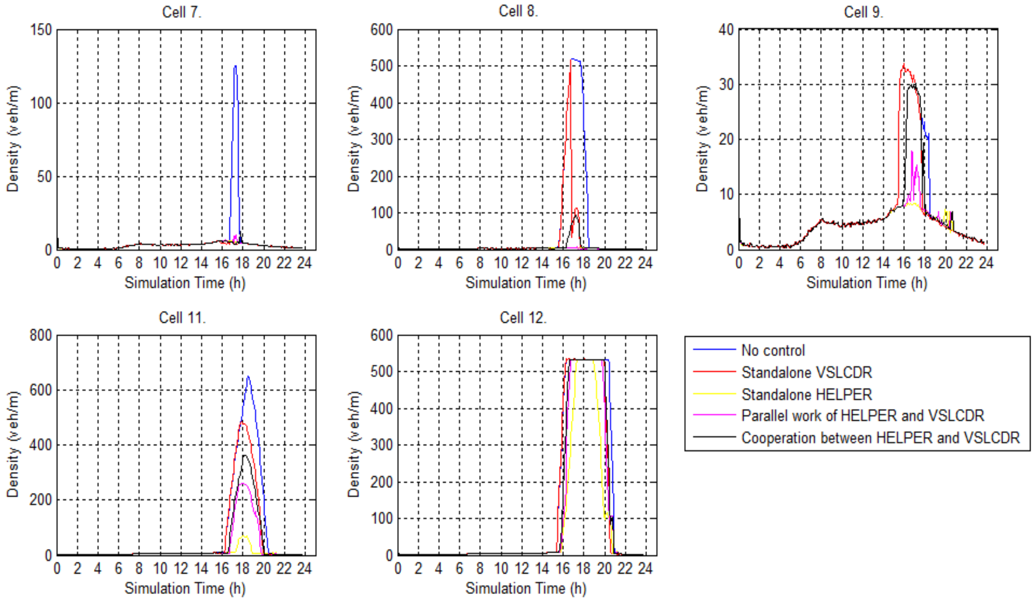

Figure 51: The comparison of mainstream density achieved by the cooperation between the HELPER ramp metering algorithm and the VSLCDR algorithm along with other involved motorway control methods

According to Figure 51, it is possible to conclude that standalone application of the HELPER ramp metering algorithm achieves the lowest mainstream density. This result is expected since the HELPER ramp metering algorithm reduces the overall input of vehicles from on-ramps into the mainstream. HELPER gradually releases vehicles into the mainstream after congestion starts to dissolve. The VSLC algorithms do not influence on-ramp flows since they regulate the motorway mainstream speed only. The cooperation between the VSLCDR and the HELPER ramp metering algorithm enables a synergy of both motorway control methods. The mentioned cooperative approach brings about lower mainstream density compared to the standalone VSLCDR algorithm, but higher density in comparison with a standalone HELPER ramp metering algorithm application. The VSLCDR in cooperation with the HELPER ramp metering algorithm can produce a higher density in comparison with the mentioned standalone ramp metering algorithm due to mainstream speed reduction. Parallel operation of the HELPER ramp metering algorithm and the VSLCDR algorithm has achieved lower mainstream density in comparison to the case in which they were cooperating. This can mean that parallel operation generates higher on-ramp queues by providing lower metering rates due to the absence of communication between these two types of motorway control methods. In order to provide a better evaluation of the mentioned cooperative approach, it is imperative to conduct an evaluation of a motorway according to the MoS such as travel time, delay, TTS, etc.

Figure 52 and Figure 53 show a comparative analysis regarding travel time and delay which includes a cooperative approach between the VSLCDR algorithm and the HELPER ramp metering algorithm, all three teaching ramp metering algorithms (ALINEA, HELPER and SWARM), a standalone application of the VSLCDR and the VSLCTR, parallel operation of the HELPER ramp metering algorithm and the VSLCTR/VSLCDR algorithm, and a situation with no control.

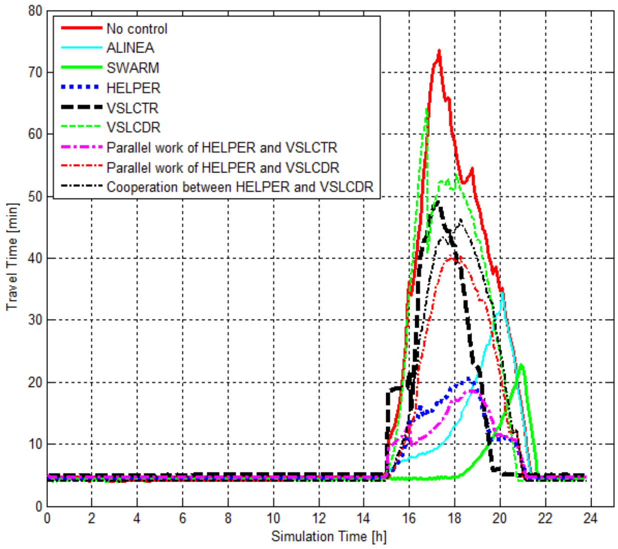

Figure 52: A comparative analysis regarding travel time which includes a cooperative approach between the VSLCDR algorithm and the HELPER ramp metering algorithm and other involved motorway control methods

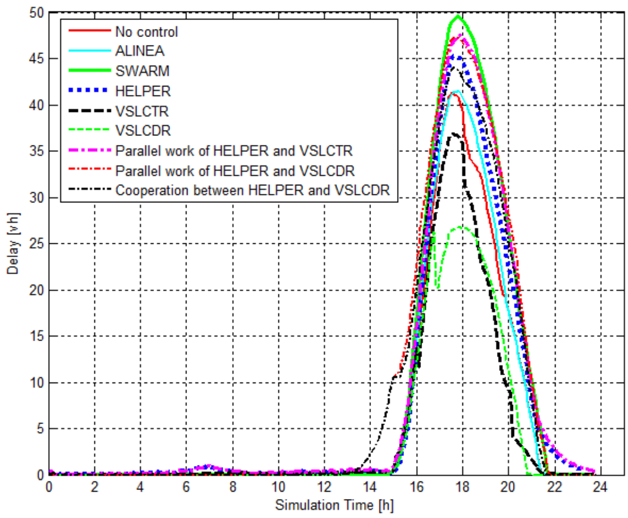

Figure 53: A comparative analysis regarding delay which includes a cooperative approach between the VSLCDR algorithm and the HELPER ramp metering algorithm and other involved motorway control methods

Figure 52 shows that cooperation between HELPER and VSLCDR provides two peaks in the resulting curve which represents travel time. After the first peak, it is possible to notice that the travel time is slightly reduced which suggests that HELPER started to induce on-ramp virtual queues. The second larger peak on the same curve represents vehicles which are released from the virtual queues into the mainstream. It is possible to notice that the delay curve for the same cooperative method in Figure 53. produces a lower delay during the second peak of the travel time curve. It is interesting that the VSLCDR delay curve produces two similar peaks as is the case with the travel time curve produced by the cooperation between HELPER and the VSLCDR. Table 4, shows the results of the comparative analysis of the cooperative approach and other relevant urban motorway control methods regarding the average MoS values.

Table 4: Results of comparative analysis between cooperative approach and other involved urban motorway control methods regarding average MoS values

	No control	ALINEA	SWARM	HELPER	VSLCTR	VSLCTR HELPER	VSLCDR	VSLCDR HELPER	Cooperation VSLCDR HELPER
Average Travel Time [min]	14.46	7.39	5.58	6.82	10.05	6.75	11.97	9.53	10.28
Average Delay [veh h]	6.06	8.8	8.03	7.29	4.85	7.59	4.20	8.75	7.02
TTS [veh h]	2949.90	2780.56	2857.70	2823.15	3005.28	3020.55	2610.97	3589.43	3001.98
Average on-ramp queue [veh]	0	16	18	17	13	18	13	18	16
Maximal on-ramp queue [veh]	0	40	49	40	15	42	13	36	31

Table 4 shows the results of a comparative analysis of the mentioned cooperative approach and other relevant urban motorway control methods regarding average MoS values. It is possible to conclude that cooperation between the VSLCDR algorithm and thr HELPER ramp metering algorithm has achieved the lowest average TTS and delay results compared to the standalone and parallel operation of these two control methods. Furthermore, mentioned cooperation has achieved the lowest average and maximal on-ramp queue. On the other hand, it achieved a slightly higher travel time compared to the mentioned parallel and standalone approaches. Standalone application of the VSLC algorithm produces a huge difference between travel time and delay in favour of the delay. The HELPER algorithm imposes additional restrictions on the on-ramp flows which drastically reduces travel time but increases delay and maximum on-ramp queues.

The overall lowest delay was achieved in the simulation scenario without ramp metering and with the use of the standalone VSLCs. This result can be explained by the setup of the CTMSIM simulator which enabled immediate merging of on-ramp flows with the mainstream - under the condition that the maximum mainstream capacity is not exceeded in the particular cell [34].

One can conclude that cooperative approaches provide smaller delays compared to the standalone HELPER ramp metering algorithm, and a much lower travel time compared to the standalone VSLCDR algorithm. These two motorway control methods affect different traffic flows on the urban motorway, so cooperation between them is tasked with providing a sort of a “fix” for their individual weaknesses. The results achieved by the mentioned cooperative approach provide an optimal ratio between travel time, delay and the average number of vehicles in waiting queues at on-ramps in comparison with the individual standalone application of the mentioned control methods and parallel operation of ramp metering and the VSLC. This proves that the cooperation between the VSLCDR and the HELPER ramp metering algorithm can maintain balanced control over the mainstream flow and on-ramp flows.

Based on these findings the next step is to develop a platform based on machine learning which will enable the integration of several different ramp metering algorithms into one control methodology. The result of the integration will be a unique ramp metering algorithm which will be able to provide metering rates for different traffic scenarios. In other words, that ramp metering algorithm, which is named INTEGRA, will replace the weaknesses of individual ramp metering algorithms with the strengths of other ramp metering algorithms.

5.3. The INTEGRA ramp metering algorithm

The first step in teaching the INTEGRA algorithm is simulating various teaching ramp metering algorithms on the same use case model. After the simulation process with the selected teaching ramp metering algorithms is complete, traffic parameters such as speed, density, flow, etc. are stored for each simulation step. Some of these traffic parameters are used by the control logic of a particular teaching ramp metering algorithm in order to compute metering rates. At this point it is not important which traffic parameters a particular teaching ramp metering algorithm uses in its control logic. What is important is to collect a set of parameters for each simulation step and for each cell with an on-ramp and associate it with the control action taken – the metering rate.

The following input traffic parameters are initially selected in order to be collected and stored during the simulation run: mainstream speed, density, flow, on-ramp queue, delay, travel time and on-ramp demand. These traffic parameters are collected and stored for each simulation step, and for all cells which contain an on-ramp. These sets of traffic parameters are associated with computed metering rates. These traffic parameters are selected because they are generally used for traffic control purposes. Additionally, these traffic parameters can be relatively easy to measure or compute from previously collected traffic data.

An adequate structure of a target vector and an input vector should be selected in order to create a learning dataset which will be presented to the adaptive ANN of the ANFIS framework as the part of INTEGRA algorithm. The FIS is the final output of the learning process and a component of the INTEGRA algorithm which actually provide control over the metering rates. The FIS requires inputs based on which it will compute outputs. Obviously, the metering rate will be its single output. It is necessary to select a number of inputs and adequate traffic parameters which will be used as the inputs for future FISs.

The number of inputs should not be large because that would induce exponential growth of the IF-THEN rules, and therefore increase computational time. For adequate input selection, it is necessary to select inputs which have the most notable impact on the control system output (metering rates). The selection process will be conducted by exhaustive search or brute force approach since there are not many possible combinations of seven possible traffic inputs. The specific dataset will be created for each possible combination of traffic inputs. This dataset will be divided into the learning and the validation dataset in a 3:1 ratio.

Input vectors for each of the created learning datasets will be created based on all possible combinations of the traffic parameters. The metering rate will be the output in all the cases. The adaptive ANN of the ANFIS framework will conduct the learning process for each of the created learning datasets during one epoch only and compute the resulting learning error. After the completion of this first learning process, the FIS trained during one epoch only will be evaluated against a validation dataset. The validation error, which is computed after the evaluation process, will also be presented.

The mentioned methodology will be applied for one, two, and three parameter sets as the potential inputs for the FIS. The number of inputs, along with a combination of traffic parameters for the analysed number of inputs, which produces the minimum cumulative error will be selected as the final configuration of inputs for a future FIS. In Figures 54, 55 and 56, one can see a graphical representation of the exhaustive search approach based on the error produced in an ANFIS learning iteration for selecting traffic parameters in the case of one, two and three inputs.

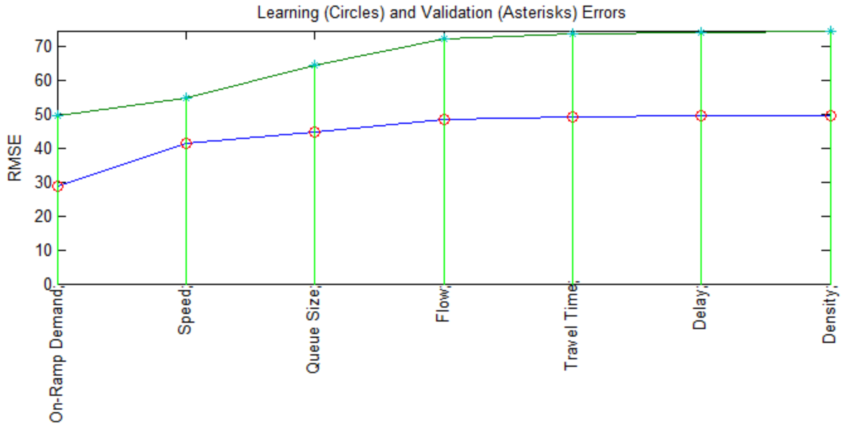

Figure 54: An exhaustive search approach based on the error produced in one ANFIS learning iteration for selecting traffic parameters in the case of one input

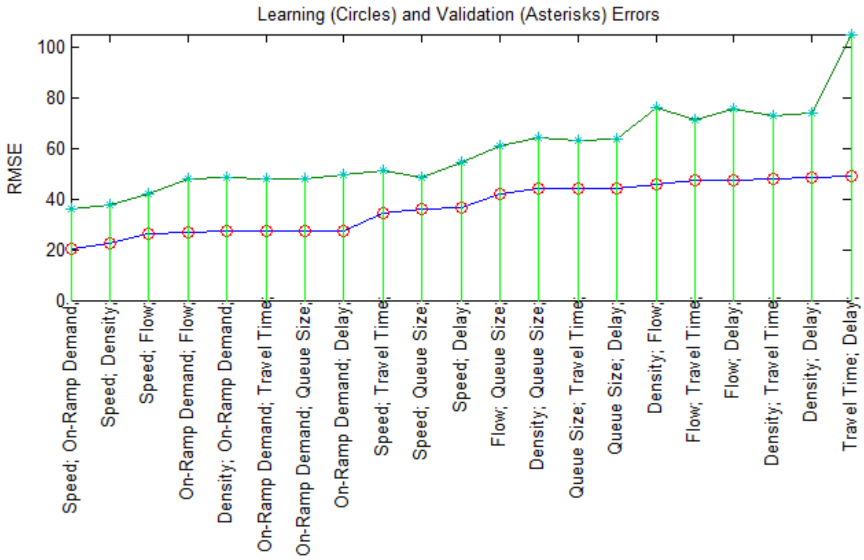

Figure 55: An exhaustive search approach based on the error produced in one ANFIS learning iteration for selecting traffic parameters in the case of two inputs

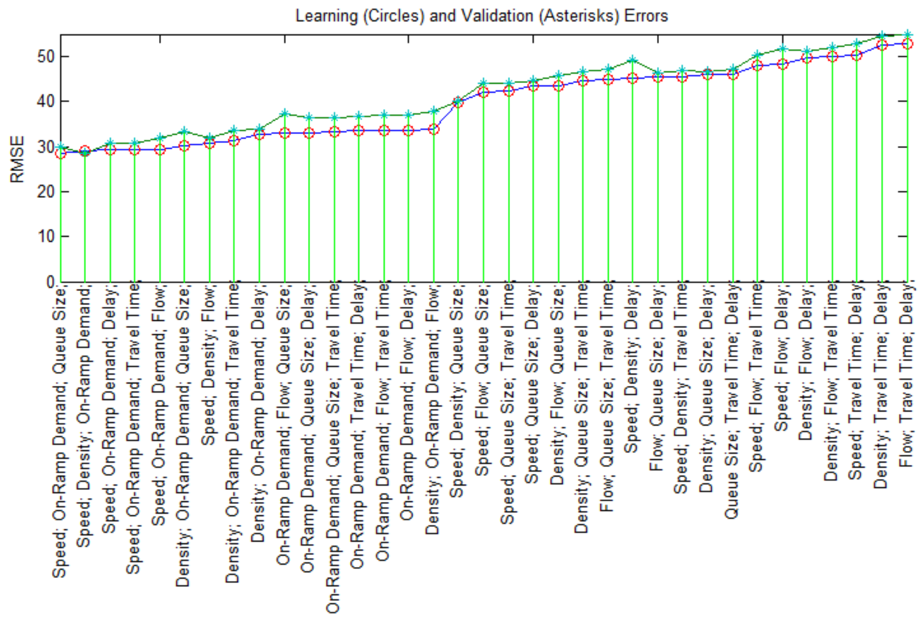

Figure 56: An exhaustive search approach based on the error produced in one ANFIS learning iteration for selecting traffic parameters in the case of three inputs

In Table 5, a representation of the best solution according to the exhaustive search for one, two or three inputs can be seen. The results are derived after a learning epoch of the ANFIS model with the adequate number and type of inputs.

Table 5: A representation of the best solution according to the exhaustive search (after one learning epoch) for one, two or three inputs

RMSE	**One input**	**Two inputs**	**Three inputs**
Learning	41.37	20.11	28.95
Validation	55.02	36.33	28.50
Cumulative	96.39	56.71	57.45

The lowest cumulative reward is detected in the case of two inputs by using mainstream speed, and on-ramp demand as the traffic parameters respectively. This case is also suitable for further examination due to a lower number of inputs, which greatly reduces the time necessary to compute metering rates.

The next step in designing the INTEGRA algorithm is to define the criteria function. The criteria function main task is to select solution among three possible solutions for each time step. Each of the three possible solutions is derived by one of the three ramp metering algorithms. In this thesis, the following configuration of the criteria function is used:

$$f(r) = 0.6 \cdot TT + 0.4 \cdot D. \quad (35)$$

The presented parameters of the criteria function are used due to a specific role of urban motorways. Urban motorways such as the Zagreb bypass are tasked mainly with serving transit traffic, but they must also deal with on-ramp traffic flows. The discrimination of on-ramp traffic which has its origin in a nearby urban area can induce large on-ramp queues and consequently induce massive spillback effects.

It is possible to conclude that in the defined criteria function, the travel time parameter is multiplied by a higher weight in comparison with the delay. This means that the INTEGRA algorithm should learn solutions which emphasize lower travel time values. Naturally, these solutions will give an advantage to solutions which provide lower travel time compared to the solution which produces lower delay. Lower travel time values are suitable for the main purpose of urban motorways since this MoS takes into account only the mainstream flow, while the MoS delay takes into account the on-ramp queue size as well. Later in this chapter an additional analysis of different criteria function setups will also be provided.

In order to assess the quality of the INTEGRA learning process, it is necessary to compare the learned outputs with the outputs in learning dataset based on which the process of learning is conducted. In Figure 57, it is possible to observe a comparative analysis of the INTEGRA learning outputs ($r_{INTEGRA}$) and outputs from which INTEGRA learns (r_{ld}). Both outputs are presented in the form of metering rates computed on identical input sets. An input set contains learning data for 5 working days. Higher RMSE values are reported during the learning process due to a lack of accurate Zagreb bypass traffic data, so only a relatively small set is used.

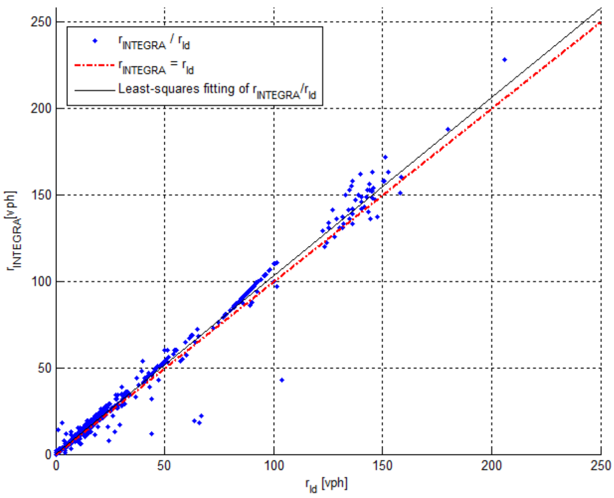

Figure 57: A comparative analysis of the INTEGRA learned outputs and outputs based on which INTEGRA is learned

Figure 58 and Figure 59 show a comparative analysis of travel time and delay which includes INTEGRA and all three teaching ramp metering algorithms. This analysis is important since it provides results based on which it is possible to conclude whether INTEGRA has learned behaviour similar to the teaching ramp metering algorithms or not. Furthermore, this analysis includes the standalone application of the VSLCDR and the VSLCTR, parallel operation of the HELPER ramp metering algorithm and the VSLCTR/VSLCDR algorithm, the previously described cooperative approach between the VSLCDR algorithm and the HELPER ramp metering algorithm, and the situation with no control.

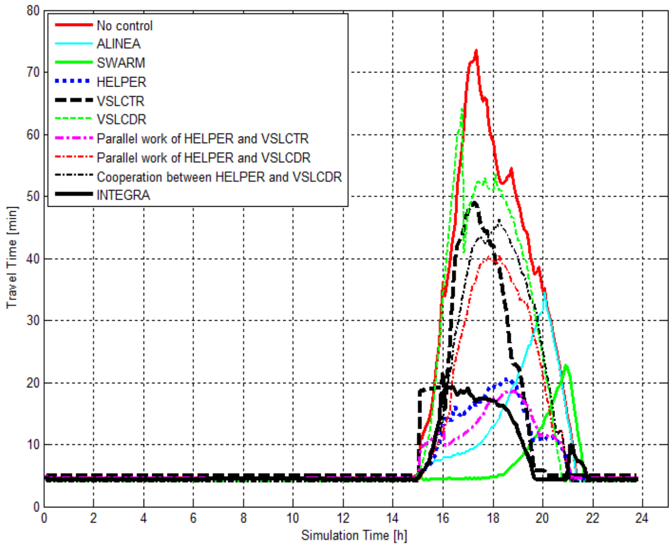

Figure 58: A comparative analysis of travel time which includes INTEGRA and other involved motorway control methods

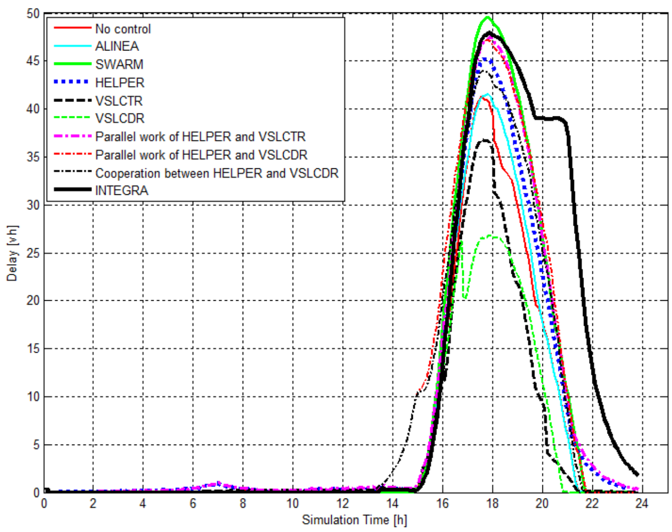

Figure 59: A comparative analysis of delay which includes INTEGRA and other involved motorway control methods

Figure 58 points to the conclusion that the travel time curve achieved by INTEGRA produces a slight increase in the time period when the increase is also present in curves produced by the ALINEA and SWARM teaching algorithms. The rest of the travel time curve produced by INTEGRA exhibits behaviour similar to the HELPER ramp metering algorithm. Figure 59 points to the conclusion that the delay curve is larger in comparison with other motorway control methods. This result must be evaluated with respect to the criteria function that gives an advantage to the solutions with a lower travel time over the solutions which favour a lower delay.

Looking at Table 6, one can conclude that the SWARM competitive ramp metering algorithm achieved the best average travel time value among all stand-alone ramp metering algorithms due to its restrictive nature. The lowest delay was achieved in the simulation scenario without ramp metering. The reason for the lowest delay in the no control scenario is related to the CTMSIM simulator restrictions which were explained in the previous chapter of this study.

Table 6: The results of a comparative analysis of INTEGRA and other involved urban motorway control methods regarding average MoS values

	No control	Teaching ramp metering algorithms			VSLCTR	VSLCTR HELPER	VSLCDR	VSLCDR HELPER	Cooperation VSLCDR HELPER	INTEGRA
		ALINEA	SWARM	HELPER						
Average Travel Time [min]	14.46	7.39	5.58	6.82	10.05	6.75	11.97	9.53	10.28	6.43
Average Delay [veh h]	6.06	8.8	8.03	7.29	4.85	7.59	4.20	8.75	7.02	10.01
TTS [veh h]	2949.90	2780.56	2857.70	2823.15	3005.28	3020.556	2610.97	3589.43	3001.98	3436.10
Average on-ramp queue [veh]	0	16	18	17	13	18	13	18	16	19
Maximal on-ramp queue [veh]	0	40	49	40	15	42	13	36	31	42

The proposed INTEGRA ramp metering algorithm produced the second lowest average travel time value in comparison to the other motorway control methods which are covered in this analysis. These results are expected due to the lowest average travel time achieved by the teaching ramp metering algorithms and the setup of a criteria function which enables the selection of lower travel time solutions. On the other hand, INTEGRA has achieved the highest delay values compared to the other ramp metering algorithms. This is a direct consequence of a generally low average travel time value. Furthermore, average on-ramp queue length and the TTS are higher in comparison to the other motorway control methods for to same reasons. The difference of the delay achieved by INTEGRA and other used motorway control methods is still within acceptable boundaries. INTEGRA managed to reduce the highest number of the maximum queue length produced by the SWARM teaching ramp metering algorithm. This is very important since the maximum capacity of on-ramps in this use case scenario is set to 50 vehicles. In Table 7, key learning dataset features after the application of a criteria function are shown.

Table 7: Key INTEGRA learning dataset features after application of criteria function [52]

INTEGRA learning dataset	Teaching ramp metering algorithms		
	ALINEA	SWARM	HELPER
Average metering rate [vph]	17.99	34.80	24.76
Variance	52.09	50.36	41.46
Number of times when algorithm solution is chosen	6587	419	1634

It can be concluded that INTEGRA, with the current setup of the criteria function, learned the majority of its control actions according to the ALINEA teaching ramp metering. Furthermore, this teaching ramp metering algorithm produced the lowest average metering rate, which consequently produced longer on-ramp queues and therefore longer delays. The SWARM and the HELPER teaching ramp metering algorithms manage to achieve lower travel time but they simultaneously produced higher delay compared to the other analysed motorway control methods.

The control strategy of the HELPER teaching ramp metering algorithm maintains increased mainstream throughput by distributing vehicles, and consequently the waiting time at "slave" on-ramps’ queues. This behaviour causes longer queues at "slave" on-ramps and consequently extends average delay at the controlled segment of the motorway. The SWARM teaching ramp metering algorithm produces the longest on-ramp queues due to its predictive techniques. This algorithm can reduce metering rates drastically in situations when an increase of on-ramp traffic demand is predicted.

INTEGRA showed promising results in learning control behaviour according to the learning dataset which was created from selected solutions with lower travel time values compared to the delay. The negative aspect of the learning process is that higher delay, the TTS, and average on-ramp queue values were produced. In order to alleviate these negative effects, two approaches are considered. The first is related to the augmentation of the existing INTEGRA setup with predictive abilities. The second is based on the analysis which includes several different setups of the INTEGRA criteria function. The results of the second approach will be presented in next chapter. In the continuation of this chapter, the results of the first approach will be evaluated in more detail.

INTEGRA augmented with the on-ramp traffic demand prediction function, (predictive INTEGRA), is based on the previously described INTEGRA setup. This augmentation of INTEGRA is significant since it uses the same setup of criteria function as the original version of this algorithm. Using the same criteria function is significant since it adequately describes the role of the majority of urban motorways. Predictive INTEGRA's main goal is to reduce some of the negative aspects of the results achieved by the original INTEGRA. This is done by introducing a proactive control approach. Changing metering rates before congestion arises can lead to a better overall result. As was mentioned earlier in this chapter, metering rates computed by the original INTEGRA are adjusted based on on-ramp traffic flow predictions.

The results achieved by the mentioned ramp metering algorithm will be compared to results of other motorway control methods used in this study. The given comparative analysis will focus on the comparison of the results achieved by the predictive INTEGRA and the INTEGRA without predictive capabilities, and all teaching ramp metering algorithms. In Table 8, the results of a comparative analysis between the predictive INTEGRA and all other involved motorway control methods can be seen.

Table 8: A comparative analysis of the predictive INTEGRA and all other involved motorway control methods

	No control	Teaching ramp metering algorithms			VSLCTR	VSLCTR HELPER	VSLCDR	VSLCDR HELPER	Cooperation VSLCDR HELPER	INTEGRA	Predictive INTEGRA
		ALINEA	SWARM	HELPER							
Average Travel Time [min]	14.46	7.39	5.58	6.82	10.05	6.75	11.97	9.53	10.28	6.43	6.69
Average Delay [veh h]	6.06	8.8	8,03	7.29	4.85	7.59	4.20	8.75	7.02	10.01	7.03
TTS [veh h]	2949.90	2780.56	2857.70	2823.15	3005.28	3020.55	2610.97	3589.43	3001.98	3436.10	3102.43
Average on-ramp queue [veh]	0	16	18	17	13	18	13	18	16	19	16
Maximal on-ramp queue [veh]	0	40	49	40	15	42	13	36	31	42	41

In Figure 60 and Figure 61, the relationship between travel time and delay in the comparative analysis which includes the predictive INTEGRA and other relevant motorway control methods can be observed.

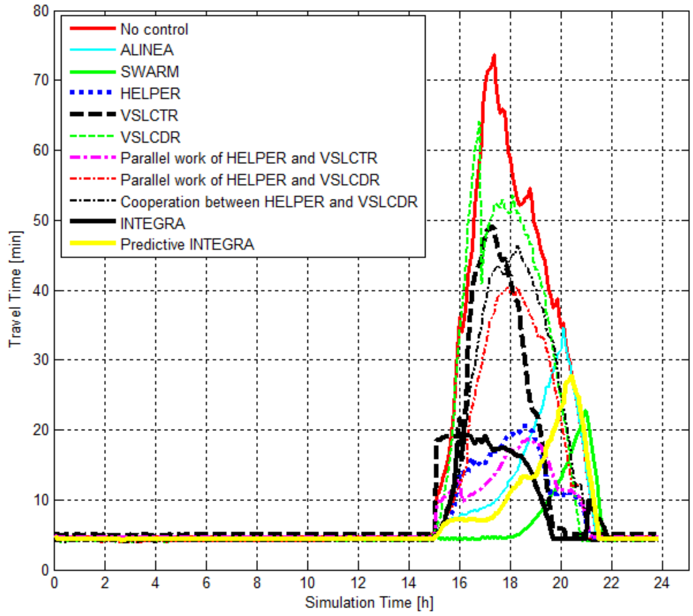

Figure 60: A Comparative analysis of travel time which includes the predictive INTEGRA and other involved motorway control methods

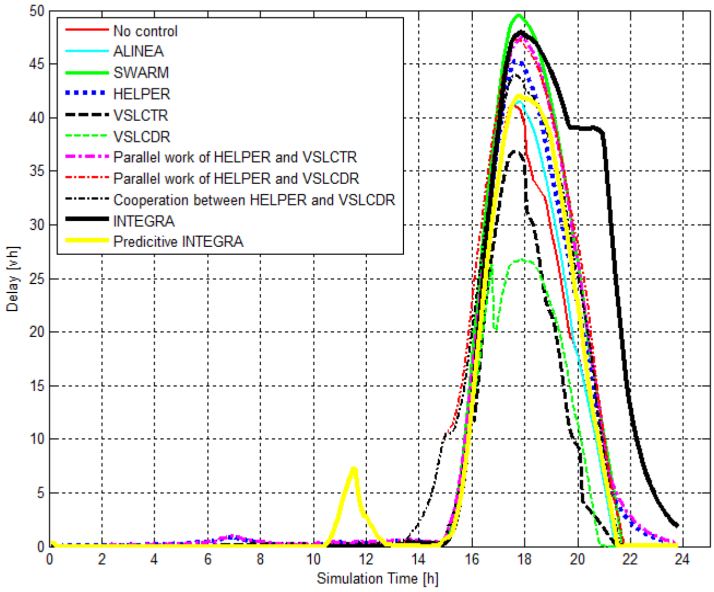

Figure 61: A Comparative analysis of delay which includes the predictive INTEGRA and other involved motorway control methods

Figure 61 shows that the predictive INTEGRA creates a higher delay in the form of a peak before congestion starts to form. This is a significant result since it indicates its ability to correctly detect congestion in the near future and reduce metering rates with respect to that information. It is noticeable that the delay produced by the predictive INTEGRA during congestion is significantly lower compared to the original INTEGRA. The reason for this can be found in the ability of the predictive INTEGRA to provide metering rate reduction before congestion arises. Imposing metering rate restrictions in the mentioned interval provides a lesser inflow of traffic from on-ramps into the mainstream before congestion arises. With this control action, the predictive INTEGRA prepares the mainstream flow for the upcoming congestion by reducing traffic flow into the mainstream. Additionally, the predictive INTEGRA produced a minimum increase in travel time compared to the original INTEGRA. The predictive INTEGRA also reduced the TTS, the average and the maximum on-ramp queues in comparison with the original INTEGRA. These results can be considered as the direct consequence of reduced delay. An on-ramp queue length comparison which includes the described predictive INTEGRA, teaching ramp metering algorithms and a no control scenario can be seen in in Figure 62.

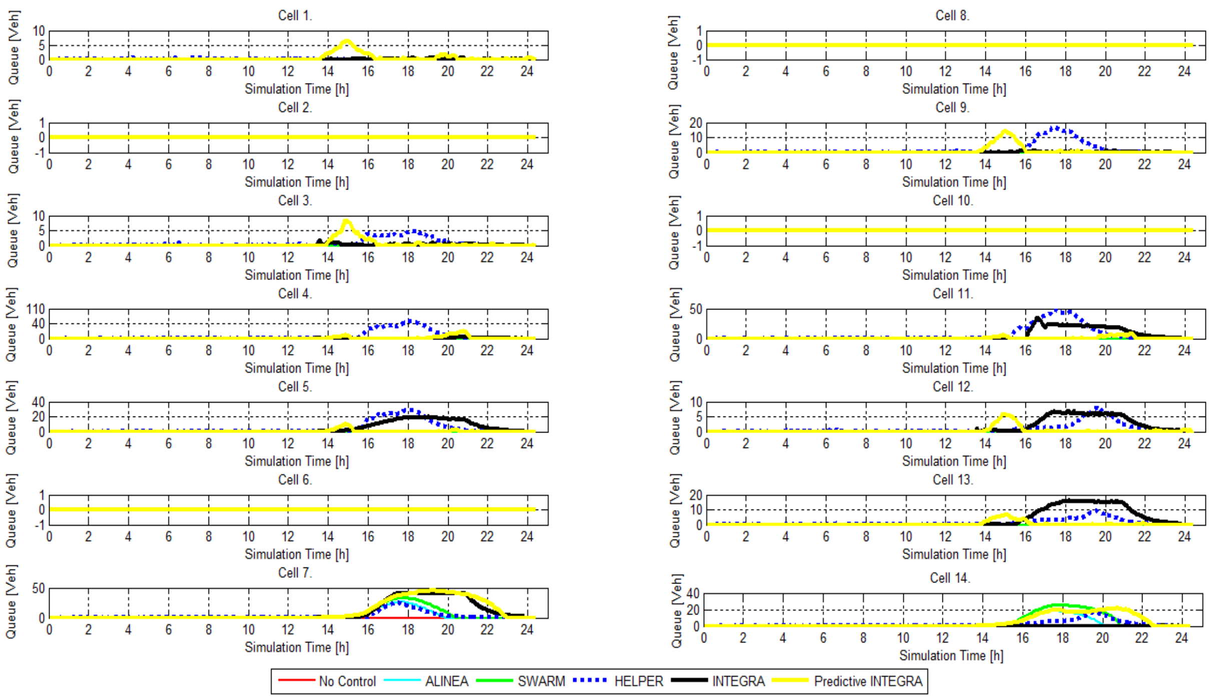

Figure 62: On-ramp queue length comparison which includes described predictive INTEGRA, teaching ramp metering algorithms and a no control scenario

It is noticeable that control actions of the predictive INTEGRA create on-ramp queues before congestion forms. This is direct evidence that supports the claim that the predictive INTEGRA produces lower metering rates before congestion forms. The mentioned action consequently produces higher on-ramp queues before congestion. In cells 4, 11 and 13, proactive action can completely mitigate congestion that would be produced by virtual queues of the original INTGERA or HELPER.

Furthermore, it is possible to assume that the predictive INTEGRA creates virtual queues at upstream on-ramps before real congestion forms. This can mean that the predictive INTEGRA learned similar control behaviour that the HELPER teaching ramp metering algorithm exhibits. It is important to emphasise that the predictive INTEGRA, unlike HELPER, creates those virtual on-ramp queues before congestion starts to form. This behaviour provides the means for the integration of cooperative and proactive control strategies in order to mitigate certain types of congestions. In comparison with the original INTEGRA, it has produced a somewhat higher travel time due to “pre-congestion” at several on-ramps which were not directly affected by congestion at that time. The average delay produced by the predictive INTEGRA is significantly lower in comparison with the original INTEGRA application. These results make the predictive INTEGRA and its proactive approach a considerable upgrade of the original INTEGRA.

5.4. Impact analysis of criteria function weighting factors

Changing the weighting factors of the criteria function was additionally analysed in order to increase the quality of the results yielded by the original INTEGRA algorithm. By changing the parameters of the criteria function, a different learning dataset can be created from the different ratios of selected solutions from teaching ramp metering algorithms. This is important since each teaching ramp metering algorithm produces different traffic parameters for each simulation step, which directly represent individual traffic solutions,. One of the three solutions derived from three different teaching ramp metering algorithms for one particular simulation time step must be selected for the inclusion in a learning dataset. The goal of this analysis is to test different ratios of two weighting factors in the criteria function.

As was mentioned earlier, the INTEGRA criteria function contains two variables or weighting factors/values: travel time and delay. The sum of weighting factors assigned to each of those variables must be one. From the results presented so far, it is possible to conclude that with an increased weighting factor assigned to a travel time parameter it is possible to achieve better throughput at the mainstream. On the other hand, if a weighting factor assigned to the delay has a larger value compared to the travel time weighting factor, it is possible to achieve better throughput at the on-ramps. Increased delay will consequently decrease the throughput of the mainstream since the significant input of vehicles is produced by on-ramps.

Depending on the motorway's key role, it is possible to go in favour of a mainstream flow or in favour of on-ramp flows. In previous chapters, the travel time weighting factor was higher compared to the delay weighting factor. This setup of the criteria was chosen since the majority of urban motorways primarily serve transit traffic. Consequently, this approach will produce lower metering rates which usually discourages drivers from using urban motorways for short journeys (with their origin and destination in the same urban area).

In this thesis, the previously mentioned statements will be validated by changing weighting factors of criteria function parameters. This should be most noticeable in cases when the difference between travel time and delay weighting factors of the INTEGRA criteria function is most prominent. Furthermore, this study will try to find an optimal solution for weighting factors of the criteria function which will provide the optimal ratio between travel time and delay values achieved by the specific setup of the INTEGRA criteria function.

Furthermore, it is possible that an urban motorway's main role is to serve traffic originating from the same urban areas. In this case, it is necessary to increase the weight factor of delay compared to the one assigned to travel time. That action will increase metering rates and make an urban motorway mainstream more accessible for short journeys. The results of this analysis will make selecting a setup of the criteria function easier.

Six different cases are considered in order to provide an analysis of the relation between delay and travel time weighting factors within the INTEGRA criteria function. Each of the different setups of the criteria function will be used for the design of a special INTEGRA ramp metering algorithm type. The analysed types of the INTEGRA ramp metering algorithms and the criteria function setups based on which they are created can be seen in Table 9.

Table 9: An analysis of the relationship between delay and travel time ponder in the INTEGRA criteria function

Type of INTEGRA algorithm	Value of travel time ponder	Value of delay ponder
INTEGRA T01D09	0.1	0.9
INTEGRA T03D07	0.3	0.7
INTEGRA T05D05	0.5	0.5
INTEGRA T06D04	0.6	0.4
INTEGRA T07D03	0.7	0.3
INTEGRA T09D01	0.9	0.1

All these types of INTEGRA ramp metering are learned according to the learning dataset created by a different setup of criteria function parameters that are presented in Table 9. Each of the learned INTEGRA types was tested on the Zagreb bypass section, which was used as the use case model in this study. All the mentioned types of INTEGRA algorithms are simulated using the same simulation model and traffic data for a typical working day (24 hours).

According to Table 10, it is possible to conclude that INTEGRA type INTEGRA T06D04 represents the original INTEGRA. In Figure 63, one can see the impact of different INTEGRA criteria function parameters on a) travel time and b) delay.

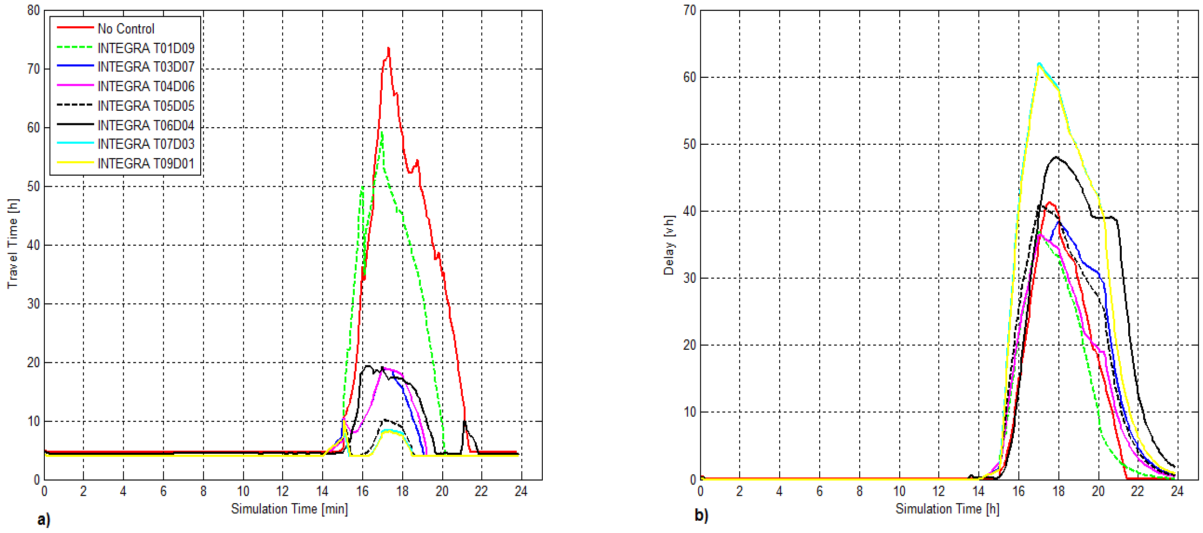

Figure 63: The impact of different INTEGRA criteria functions parameters on a) travel time and b) delay

In Figure 63 one can notice that different types of the INTEGRA algorithm show similar behaviour regarding travel time and delay. The difference between the curves that describe travel time and delay are in line with the setup of the criteria function used for each type of the INTEGRA algorithm. The results of the comparative analysis of different types of INTEGRA algorithms according to the average values of the TT, Delay, the TTS, queue length and maximum queue length are shown in Table 10.

Table 10: Results of the comparative analysis of different types of INTEGRA algorithms

	INTEGRA T01D09	INTEGRA T03D07	INTEGRA T04D06	INTEGRA T05D05	INTEGRA T06D04	INTEGRA T07D03	INTEGRA T09D01
Average Travel Time [min]	11.10	5.55	5.69	4.52	6.43	4.37	4.36
Average Delay [veh h]	5.41	7.63	6.39	7.68	10.01	11.76	11.66
TTS [veh h]	2129.5	2919.5	2186.5	2536.3	3436.10	4893.8	4849.7
Average TTS [veh h]	19.4	22.07	28.26	20.97	19.48	23.92	24.82
Average Queue [veh]	15	19	18	19	19	23	22
Max. Queue [veh]	22	40	36	44	42	61	61

In Figure 64, a representation of a graphical comparative analysis which includes all INTEGRA algorithm types according to the achieved results presented in Table 6. can be seen.

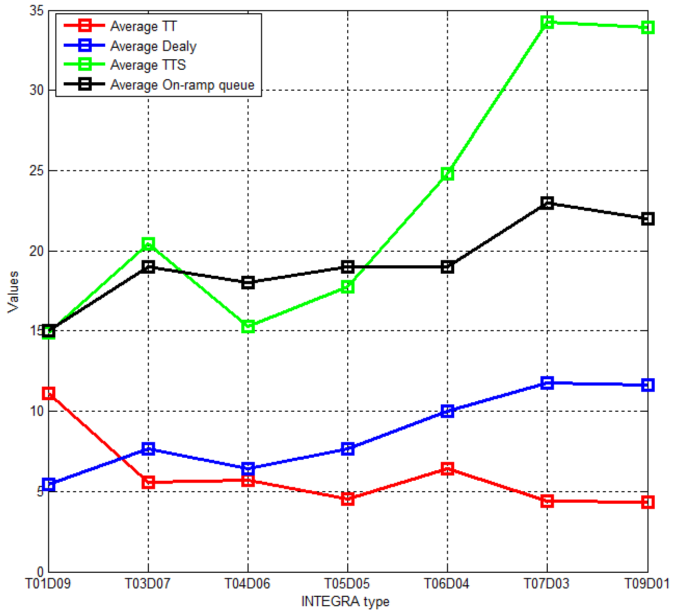

Figure 64: A graphical comparative analysis that includes all INTEGRA algorithm types

In Figure 65, a graphical representation of the TTS during the entire simulation run for each tested INTEGRA algorithm type can be seen. It is possible to conclude that the curve in Figure 65 exhibits the same behaviour as the curve which represents the average TTS in Figure 64.

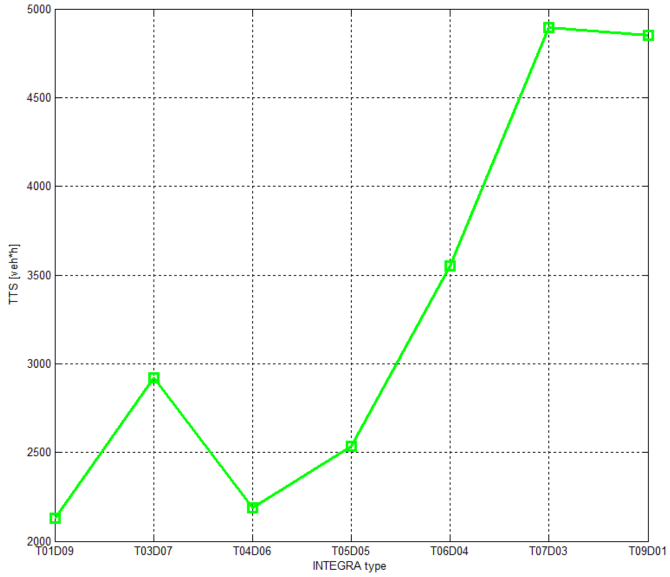

Figure 65: A graphical representation of the TTS during the entire simulation run for each tested INTEGRA algorithm type

According to the results presented in Table 10 and Figure 64 it is possible to conclude that the highest difference between the parameters of the INTEGRA criteria function is in the cases of the two most diverse INTEGRA algorithm types: INTEGRA T01D09 and INTEGRA T09D01.

Therefore, these two types of the INTEGRA algorithm have the greatest difference between the achieved travel time and delay. INTEGRA T09D01 achieves the lowest travel time value, but on the hand, it achieves the highest delay. INTEGRA T01D09 achieves the lowest delay, but consequently, the highest travel time.

Furthermore, it is possible to conclude that those types of INTEGRA algorithm which are created based on the criteria function with a higher delay weighting factor compared to the travel time achieve a much higher TTS. This can be explained as being due to the chosen solutions, which provide shorter metering rates and therefore longer queues at on-ramps. Longer queues at on-ramps produce longer waiting time which consequentially produces higher values of the TTS measure. The trend of increasing an average on-ramp queue value can be observed from INTEGRA T01D09 to INTEGRA T09D01 types of the INTEGRA algorithm with some minor exceptions.

According to Figure 64, it is possible to divide all the analysed INTEGRA algorithm types (marked at x-axis) into two regions. The first region includes algorithms from INTEGRA T01D09 to INTEGRA T05D05, and the second from INTEGRA T05D05 to INTEGRA T09D01. In order to find the type of INTEGRA algorithm with the optimal weighting factors of travel time and delay it is necessary to narrow the analytic search down and select one of the two mentioned regions for a detailed examination. It is possible to conclude that the region between INTEGRA T01D09 and INTEGRA T05D05 is the most interesting to observe. This region is interesting for further analysis since the types of the INTEGRA algorithms in this region produce lower values of the TTS, delay, and average on-ramp queue compared to the other region. Compared to the other region, the increase of travel time in this region is noticeable, but not too drastic.

Furthermore, it is possible to conclude that INTEGRA T04D06 achieved much lower values of all the involved MoSs compared to the original INTEGRA (INTEGRA T06D04). These results suggest that the INTEGRA T04D06 criteria function configuration can select solutions from the teaching ramp metering algorithm that produce better overall results than the solutions selected by the criteria function setup used in the original INTEGRA. In conclusion, INTEGRA T04D06 criteria function setup is selected as optimal among all the analysed types of INTEGRA algorithms. The reason for this is based on the fact that the mentioned INTEGRA algorithm type achieved the second lowest values of average on-ramp queue, TTS, and delay. INTEGRA T01D09 achieved the best values for all the mentioned parameters, but consequently produced the highest travel time which is unacceptable for an urban motorway. On the other hand, INTEGRA T04D06 type achieved nearly half the value of the travel time compared to INTEGRA T01D09. The key features of the learning dataset created by different setups of the criteria function used in all the analysed types of the INTEGRA algorithm are shown in Table 11.

Table 11: Key features of the learning dataset created by different setups of the criteria function used in all the analysed types of the INTEGRA algorithm

		Teaching ramp metering algorithms		
		ALINEA	**SWARM**	**HELPER**
INTEGRA T01D09	Average metering rate [vph]	17.31	37.01	21.68
	Variance	33.43	111.18	49.84
	Number of times when algorithm solution is chosen	5809	1757	1074
INTEGRA T03D07	Average metering rate [vph]	17.99	34.80	24.76
	Variance	37.12	108.28	51.54
	Number of times when algorithm solution is chosen	5769	1375	1496
INTEGRA T04D06	Average metering rate [vph]	18.73	35.04	23.43
	Variance	41.08	108.77	48.30
	Number of times when algorithm solution is chosen	5861	1299	1480
INTEGRA T05D05	Average metering rate [vph]	20.72	28.68	21.36
	Variance	49.04	89.41	42.32
	Number of times when algorithm solution is chosen	6220	916	1504
INTEGRA T06D04	Average metering rate [vph]	17.99	34.80	24.76
	Variance	52.09	50.36	41.46
	Number of times when algorithm solution is chosen	6587	419	1634
INTEGRA T07D03	Average metering rate [vph]	21.93	19.04	21.86
	Variance	53.21	31.63	41.03
	Number of times when algorithm solution is chosen	6520	346	1774.00
INTEGRA T09D01	Average metering rate [vph]	22.61	18.17	20.14
	Variance	54.42	31.00	39.13
	Number of times when algorithm solution is chosen	6339	351	1950

Figure 61. to 71. show the comparative analysis of learned outputs ($r_{INTEGRA}$) and outputs based on which all INTEGRA types are trained (r_{ld}). Both measures are expressed in metering rates. The results achieved by INTEGRA T06D04 are not displayed since they were presented earlier in this chapter.

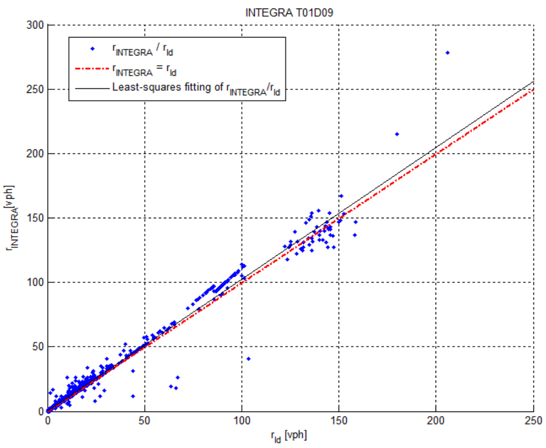

Figure 66: Comparative analysis of INTEGRA T01D09 learned outputs and outputs based on which INTEGRA is learned

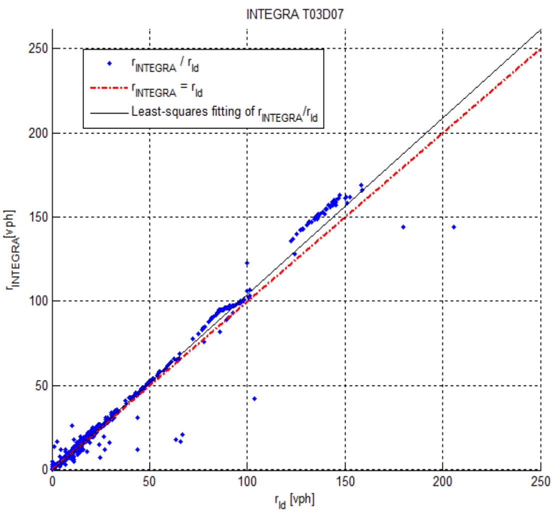

Figure 67: Comparative analysis of INTEGRA T03D07 learned outputs and outputs based on which INTEGRA is learned

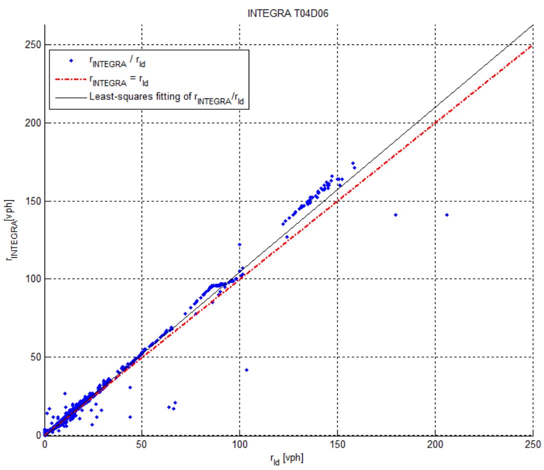

Figure 68: Comparative analysis of INTEGRA T04D06 learned outputs and outputs based on which INTEGRA is learned

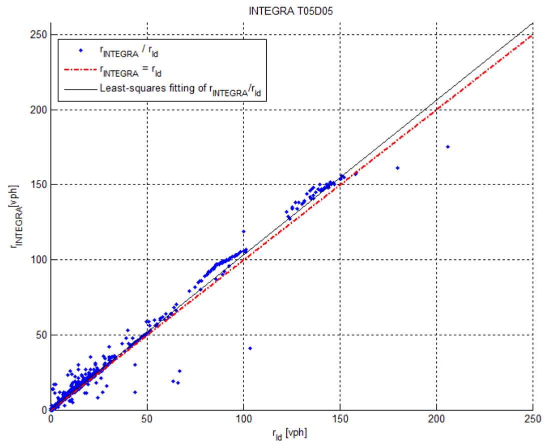

Figure 69: Comparative analysis of INTEGRA T05D05 learned outputs and outputs based on which INTEGRA is learned

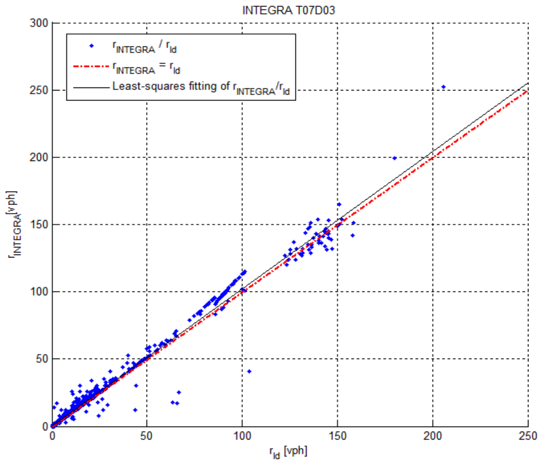

Figure 70: Comparative analysis of INTEGRA T07D03 learned outputs and outputs based on which INTEGRA is learned

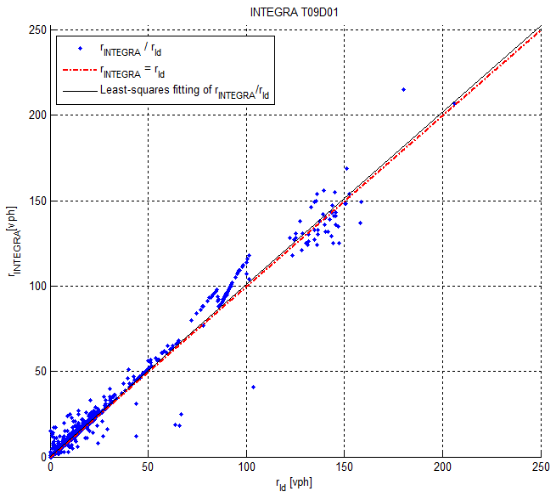

Figure 71: Comparative analysis of INTEGRA T09D01 learned outputs and outputs based on which INTEGRA is learned

Figure 66. to 71. shows that the lowest difference between the least-squares fitting of outputs based on which all INTEGRA types are learned and the line which connects the mentioned outputs, and INTEGRA learning outputs is in the case of INTEGRA T09D01 and INTEGRA T01D09 types. These two types of INTEGRA algorithms are the two most extreme cases in the tested group of INTEGRA algorithms so it is possible to conclude that this kind of control knowledge can be learned with higher precision.

Table 12 shows the comparative analysis of the INTEGRA T04D06 type and the results provided by the teaching ramp metering algorithms such as ALINEA, SWARM, HELPER and the standalone VSLCs. Furthermore, in this comparative analysis scenario which involves the parallel operation of the HELPER ramp metering algorithm and the VSLC, the cooperative approach between HELPER and the VSLCDR, and the no control is also included.

Table 12: The results of the comparative analysis of INTEGRA T04D06 and other involved motorway control methods

	No control	ALINEA	SWARM	HELPER	VSLCTR	VSLCTR HELPER	VSLCDR	VSLCDR HELPER	Cooperation VSLCDR HELPER	Predictive INTEGRA	INTEGRA T06D04	INTEGRA T04D06
Average Travel Time [min]	14.46	7.39	5.58	6.82	10.05	6.75	11.97	9.53	10.28	6.69	6.43	5.69
Average Delay [veh h]	6.06	8.8	8.03	7.29	4.85	7.59	4.20	8.75	7.02	7.03	10.01	6.39
TTS [veh h]	2949.90	2780.56	2857.70	2823.15	3005.28	3020.55	2610.97	3589.43	3001.98	3102.43	3436.10	2186.5
Average on-ramp queue [veh]	0	16	18	17	13	18	13	18	16	16	19	18
Maximal on-ramp queue [veh]	0	40	49	40	15	42	13	36	31	42	42	36

In Figure 72 and Figure 73 the relation between travel time and delay in the comparative analysis which involves INTEGRA T04D06 and other involved motorway control methods can be observed.

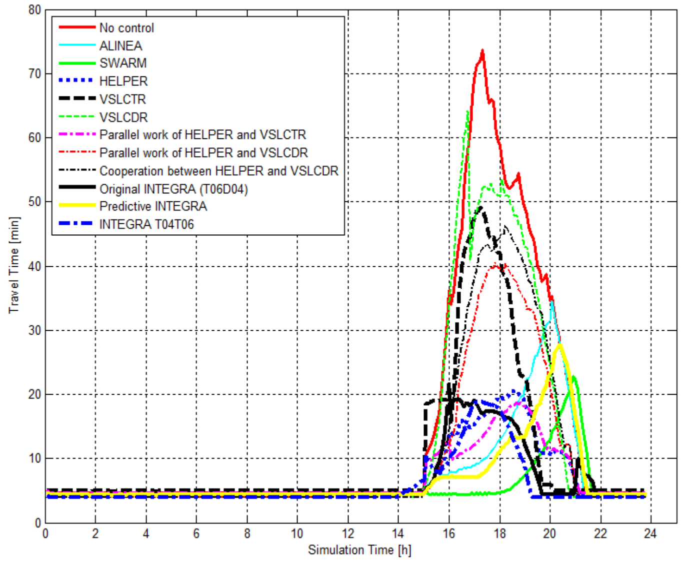

Figure 72: The comparative analysis of travel time which includes INTEGRA T04D06 and other involved motorway control methods

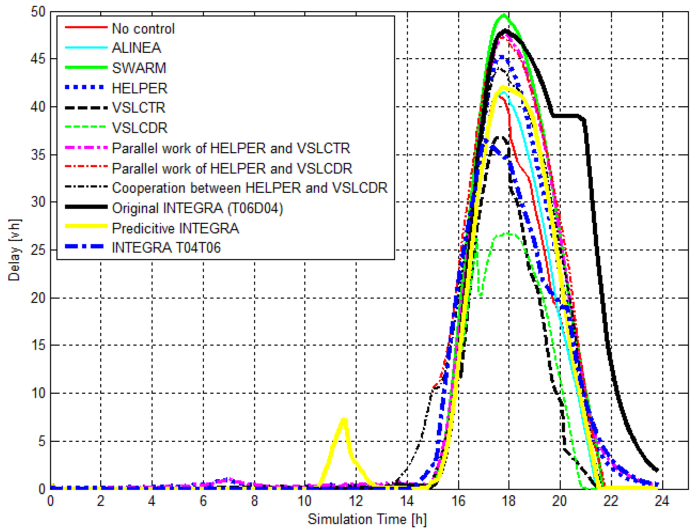

Figure 73: The comparative analysis of delay which includes INTEGRA T04D06 and other involved motorway control methods

Figure 72 shows that INTEGRA T04D06 travel time curve is slightly elevated before congestion starts. Following the elevation, it shows similar behaviour such as the HELPER ramp metering algorithm when creating virtual queues. This behaviour suggests that INTEGRA T04D06 learned the strategy of creating virtual queues in situations when shockwave backpropagation is detected, the difference being that the discharge of virtual on-ramp queus is done more eficently. In Figure 73, INTEGRA T04D06 exhibited similar behaviour compared to the initially tested INTEGRA, but with much lower values. According to Table 12. it is possible to conclude that INTEGRA T04D06 has outperformed the two concurrent INTEGRA modifications: the original INTEGRA and the predictive INTEGRA. The predictive INTEGRA managed to produce lower TTS and an average on-ramp queue compared to INTEGRA T04D06. It is interesting that the predictive INTEGRA did not manage to produce lower delay compared to INTEGRA T04D06 despite a lower average on-ramp queue. The reason for these results lies in the fact that the predictive INTEGRA produces a generally higher maximum queue and higher travel time compared to INTEGRA T04D06. Furthermore, INTEGRA T04D06 produced the new second best travel time in comparison with other involved motorway control methods. The SWARM ramp metering algorithm produced lower travel time compared to INTEGRA T04D06. On the other hand, the SWARM ramp metering algorithm produced a much larger delay compared to INTEGRA T04D06. INTEGRA T04D06 achieved the lowest delay compared to the all other involved motorway control strategies with the exception both VSLC algorithms and a no control situation (which does not involve the creation of on-ramp queues due to simulator limitations).

It is possible to conclude that the selection of adequate weighting factors in the INTEGRA criteria function can produce better overall results compared to the original INTEGRA augmentation which involves on-ramp traffic predictions. INTEGRA augmentation which involves on-ramp traffic predictions provides corrections of metering rates previously computed by the original INTEGRA. On the other hand, INTEGRA T04D06 is the product of the selection process of adequate weighting factors. This enables a searching process within the space of all previously collected/computed solutions derived from all teaching ramp metering algorithms. According to the achieved results of the searching process, it is possible to find the optimal solution for the selection of the criteria function weighting factors. INTEGRA augmentation which involves on-ramp traffic predictions provides promising results as well. This is especially noticeable in the case of average queues at on-ramps and the average TTS. In future work, it is necessary to develop a more comprehensive logic which will adjust previously computed metering rates based on on-ramp traffic flow predictions.

6. Conclusion

The roles of specific road classes within the urban area and its immediate vicinity changed with the expansion of urban regions. Urban bypasses have undergone the most interesting transformation regarding their role in the traffic systems of nearby urban areas. The transformation was due to the trend of expansion of urban areas and the fact that urban bypasses were affected by larger traffic loads. The originally projected MoS for urban bypasses was compromised by a constant increase in traffic demand originating from nearby urban areas. The solutions to the this problem were in constructional operations such as the expansion of urban bypasses via new traffic lanes and their improvement by new nodes with several on- and off-ramps. Urban bypasses can be considered urban motorways when they become surrounded by urban infrastructure so there is no space for the "build only" approach and/or if they are assigned a new function, e.g. to serve traffic originating from an urban area. In this thesis, the segment between Lučko and Jankomir of the Zagreb bypass is selected as the use case scenario, since it shows characteristics of an urban motorway. It is characterized by strong connections with the Zagreb urban network and an increased traffic load during the afternoon peak hour. The rest of the Zagreb bypass is in a process of transition between an urban bypass and an urban motorway.

The research described in this thesis is motivated by the search for a motorway control method which will enable urban motorways to better fulfill their roles. The urban motorway as a part of the urban road network has the role to serve traffic demand originating from the urban area with respect to the higher LoS. The higher LoS must be ensured for mainstream traffic flows (e.g. transit traffic). The maximum length of queues at on-ramps has also been taken into account since a spill back effect must be avoided. Ramp metering, as the chosen motorway control method, is the focus of this thesis. The first approach sought to establish cooperation of ramp metering and another motorway control method. In this thesis, the VSLC is selected as the motorway control method which will cooperate with ramp metering. The HELPER ramp metering algorithm is described as the suitable ramp metering algorithm which can be used in cooperation with the VSLC. The mentioned ramp metering algorithm creates "virtual" on-ramp queues in the upstream region of the controlled urban motorway with regard to the place of congestion. By acting in this manner, HELPER produces the effect of cooperation between on-ramps in order to achieve the common goal – better throughput of the motorway mainstream. In addition, the VSLC affects the previously mentioned upstream part of the urban motorway by changing speed limits. This produces a synergetic effect between two different motorway control methods, and potentially produces slower backpropagation of shock waves. The cooperation of the VSLC and ramp metering has produced better delay and a shorter length of maximum queue lengths in comparison with other ramp metering algorithms that are included in the comparative analysis. On the other hand, the mentioned cooperative approach has achieved better travel time compared to the other VSLC algorithms. Based on these findings, it can be concluded that the second hypothesis is confirmed and the novel cooperative approach between ramp metering and the VSLC constitutes a valid scientific contribution.

The described cooperative approach is effective in specific traffic scenarios when the place of a bottleneck is known and when the upstream section relative to the place of the bottleneck is covered by the VSLC. Furthermore, the cooperative approach can produce unnecessary slowdowns at critical places of an urban motorway system (where they are initially implemented) in the case of low traffic demand. In the urban motorway, it is not unusual that traffic demand suddenly increases on different segments with on-ramps. At that point, it is necessary to develop a ramp metering algorithm which will effectively resolve traffic congestion related to high fluctuations in traffic demand that are characteristic of urban motorways.

The next step towards the development of a ramp metering algorithm which will enable a more comprehensive dealing with congestion on urban motorways was based on the fact that each ramp metering algorithm produces better overall results in specific traffic scenarios. Considering this fact, this thesis is using an approach that utilizes an adaptive ANN and FIS in order to integrate several different ramp metering control behaviours into a single control behaviour. This approach was made possible by using the ANFIS structure based on an adaptive ANN in order to produce a tuned FIS. The mentioned structure was used as the framework for the ramp metering algorithm named INTEGRA according to its main role – integration of several different ramp metering control behaviours. The integration of several different ramp metering control behaviours into a single control behaviour is one of the main scientific contributions of this thesis. INTEGRA has the main goal to build a learning dataset upon which the adaptive ANN will create a calibrated FIS with metering rates as its outputs. The initial learning dataset contains outputs of three teaching ramp metering algorithms which are based on different control behaviours. All ramp metering algorithms are simulated on the same use case scenario so each produces different solutions for the same traffic scenario.

INTEGRA uses a criteria function in order to select the solutions that are in line with its parameters. The first experiment included a criteria function which has selected solutions that give a slight advantage to the solutions which favour travel time over delay. This setup of the criteria function adequately describes the main role of the urban motorway. In comparison with the cooperative approach, the results showed that INTEGRA has produced a much lower travel time, but on the other hand, it has also produced higher delay. Furthermore, INTEGRA did not manage to produce lower values of on-ramp queues and the TTS compared to the cooperative approach due to higher delay values. At this point, it is possible to say that the INTEGRA criteria function selected solutions that are in line with the criteria function, but those solutions do not provide the best overall results.

In order to achieve better MoS values compared to the other analysed urban motorway methods, there are two possible directions toward the improvement of INTEGRA. The first is an augmentation of the existing INTEGRA with the current criteria function, and the second is a selection of a different criteria function. In the latter case, the original INTEGRA is augmented in order to adjust its output metering rates according to the traffic demand prediction for an on-ramp for which the metering rate is computed. In this case, the setup of criteria function remains the same. Test results showed that the predictive INTEGRA achieved lower delay, TTS and average queue length in comparison with the original INTEGRA and all teaching ramp metering algorithms. Additionally, the average travel time was increased slightly in comparison to all the previously mentioned ramp metering algorithms. These results show that the augmentation of the predictive INTEGRA can provide better overall control compared to the original INTEGRA. This approach provides additional value to the research related to the first hypothesis and the creation of an advanced learning framework for ramp metering constitutes a valid scientific contribution.

The second direction towards the improvement of the overall MoS results of the original INTEGRA, and consequently the full confirmation of the first hypothesis, is the changing of weighting factors of travel time and delay in the criteria function. Several different setups of the criteria function are selected, and adequate learning datasets were created based on them. Based upon the mentioned learning datasets, learning processes were conducted, and several different INTEGRA ramp metering algorithms were created. A comparative analysis between all the mentioned INTEGRA ramp metering algorithms was carried out. The analysis has shown that the INTEGRA ramp metering algorithm that was created by the criteria function with a weighting factor of 0.4 for travel time and 0.6 for delay produced the best overall results. The INTEGRA ramp metering algorithm created based on the mentioned setup of the criteria function produced lower values of all involved MoSs compared to the original INTEGRA. It also produced better MoS related results compared to the predictive INTEGRA, to the cooperative approach between ramp metering and the VSLC, and to other teaching ramp metering algorithms, with the exception of the average on-ramp queue and the average TTS which are slightly higher. At this point, it is possible to conclude that the first hypothesis is fully confirmed and that all the related scientific objectives are met. Furthermore, the results have shown the importance of an appropriate data selection process in creating a learning dataset later used in the INTEGRA machine learning process. Considering the relationship setup of weighting factors assigned to travel time and delay in the criteria function, it is possible to produce results which will go in favour of one of the two MoSs used in criteria function. Furthermore, the presented results suggest that the mentioned "biased" approach in the setup of the criteria function, does not always yield the best possible overall results.

Due to the limitations of this research and conclusions that where reached during it, there are several courses which could be feasibly pursued in the future, such as the use of a macroscopic traffic simulation model with more accurate traffic data on the Zagreb bypass section between Lučko and Jankomir nodes, the expansion of the use case model on the entire Zagreb bypass, considering integration of the VSLC and ramp metering by using the augmented ANFIS framework, analysing additional different criteria function setups for the INTEGRA ramp metering algorithm, the integration of an INTEGRA designed on the setup of criteria function which enables best overall results, and the prediction of the traffic demand on each on-ramp, conducting a simulation which will enable the inclusion of connected vehicles in cooperation with the ramp metering system, impact analysis of the penetration rate of autonomous and connected vehicles in motorway systems with applied cooperative ramp metering etc.

Bibliography

[1] J. Barcelo, Fundamentals of Traffic Simulations, Springer Science Business Media, 2010.

[2] M. Papageorgiou, C. Diakaki, V. Dinopoulou, A. Kotsialos and Y. Wang, “Review of Road Traffic Control Strategies,” *Proceedings of the IEEE,* vol. 91, no. 12, pp. 2043-2067, 2003.

[3] M. Treiber and K. Arne, Traffic Flow Dynamics - Data, Models andSimulation, New York, NY, USA: Springer-Verlag, 2013.

[4] J. R. Scariza, “Evaluation of Coordinated and Local Ramp Metering Algorithms Using Microscopic Traffic Simulation,” Master of Science in Transportation Thesis, Massachusetts Institute of Technology (MIT), Massachusetts, USA, Jun 2003.

[5] R. Murphey and P. Pardalos, Cooperative Control and Optimization, Springer Science & Business Media, 2002.

[6] A. Ghods, K. Hosein, R. Ashkan, and M. Tabibi, “A Genetic-Fuzzy Control Application to Ramp Metering and Variable Speed Limit Control,” in *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, Montreal, Quebec, 2007, pp. 1723-1728.

[7] M. Gregurić, E. Ivanjko, I. Galić, S. Mandžuka and H. Gold, “Urban Highways Level of Service Improvement Based on Intelligent Ramp Metering,” in *25th Central European Conference on Information and Intelligent Systems CECIIS*, Varaždin, Croatia, 17– 19. Sep 2014, pp. 296-303.

[8] American Association of State Highway and Transportation Officials, “A Policy on Geometric Design of Highways and Streets,” Washington, D.C., 2001.

[9] T. R. Board, Highway Capacity Manual, Washington D.C: National Research Council, 2000.

[10] B. Huzjan, S. Mandžuka and G. Kosa, “Real-time traffic safety management model on motorways,” *Tehnički vjesnik,* vol. 24, no. 5, pp. 1457-1469, 2017.

[11] A. Kesting, Traffic flow dynamics, Springer Science Business Media, 2013.

[12] IBI Group UK Ltd (IBI), »Deliverable D7.5 project EURAMP, Handbook of ramp metering,« 2007.

[13] M. Gregurić, M. Buntić, E. Ivanjko and S. Mandžuka, “Improvement of highway level of service,” in *Proceedings of 21st International Symposium on Electronics in Transport ISEP 2013*, Ljubljana, Slovenia, 25-26. Mar 2013.

[14] C. Lee and B. V. Hellinga, “Evaluation of variable speed limits to improve traffic safety,” *Transportation Resreach Part C,* vol. 14, no. 3, pp. 213-228, 2016.

[15] N. Mahajan, A. Hegyi, G. S. Weg and S. P. Hoogendoorn, “Integrated Variable Speed Limit and Ramp Metering Control Against Jam Waves -A COSCAL v2 Based Approach,” in *Proceedings of IEEE 18th International Conference on Intelligent Transportation Systems*, Las Palmas, Spain, 2015, pp. 1156-1162.

[16] “Motorways network huka.hr,” Croatian Association of Motorway Concessionaires (HUKA), [Online]. Available: http://www.huka.hr/mreza-autocesta. [Accessed 28 January 2016].

[17] M. Wang, W. Daamen, S. P. Hoogendoorn and B. van Arem, "Cooperative Car-Following Control: Distributed Algorithm and Impact on Moving Jam Features," *IEEE Transactions on Intelligent Transportation Systems,* vol. 17, no. 5, pp. 1459-1471, May 2016.

[18] B. Kerner, The Physics of Traffic, Empirical Freeway Pattern Features, Engineering Applications, and Theory, Springer Complexity, Springer-Verlang Berlin Heidelberg, 2010.

[19] M. Gregurić, E. Ivanjko and S. Mandžuka, "New concepts for urban highways control," in *Proceedings of REAL CORP 2014. 21. - 23. May,* Vienna, Austria, 2014, pp. 423-432.

[20] M. Abdel-Aty, A. Dhindsa and V. Gayah, "Considering various ALINEA ramp metering strategies for crash risk mitigation on freeways under congested regime Transportation Research Part C," *Transportation Research Part C,* vol. 15, no. 2, p. 113–134, 2007.

[21] I. Bošnjak, Intelligent transportation system I, Zagreb: Faculty of transport and traffic sciences, University of Zagreb, 2006.

[22] A. Hegyi and S.P. Hoogendoorn, "Dynamic speed limit control to resolve shock waves on freeways – Field test results of the SPECIALIST algorithm," in *13th International IEEE Annual Conference on Intelligent transportation systems,* Madeira Island, Portugal, 19-22. Sep 2010, pp. 519-524.

[23] L. Martine, L. Gilbert, T. Katalin and T. Philippe, Operations Research and Decision Aid Methodologies in Traffic and Transportation Management, Springer Berlin - Heidelberg, 1998. pp. 46-83.

[24] E. Marie-Christine, J. Polak, R. Krishnan and M. Pleydell, "A Global Comparison of Ramp-Metering Algorithms Optimizing Traffic Distribution on Motorways and Arterials," in *Road Transport Information and Control Conference and the ITS United Kingdom Members' Conference (RTIC 2010) - Better transport through technology, IET*, London, UK, 2010, 25-27 May 2010, P07.

[25] A. Justice, N. Bhaven and S. Scott, "Calibration of Microsimulation Models for Multimodal Freight Networks, Techical Report # MATC-UNL: 42," Mid-America Transportation Center, University of Nebraska-Lincoln, USA, Jun 2012.

[26] Federal Highway Administration, "Revised Monograph on Traffic Flow Theory, Technical Report," Jul, 2005.

[27] S. Rajeswaran and S. Rajasekaran, "A Study of Vehicular Traffic Flow Modeling Based on Modified Cellular Automata," *IOSR Journal of Mathematics (IOSR-JM),* vol. 4, no. 5, pp. 32-38, 2014.

[28] T. Tamás and T. H. Márton, "A practical manual for Vissim COM programming in Matlab, lectures," Budapest University of Technology and Economics Dept. for Control of Transportation and Vehicle Systems, Budapest, Hungary, 2015.

[29] P. G. Gipps, "A behavioural car-following model for computer simulation," *Transportation Research Board Part B,* vol. 15, pp. 105-111, 1981.

[30] M. J. Lighthill and G. B. Whitham, "On kinematic waves II: A theory of traffic flow on long, crowded roads," in *Proceedings of The Royal Society of London Ser. A 229*, London, UK, 1955, pp. 317-345.

[31] P. I. Richards, “Shock waves on the highway,” *Operations Research,* vol. 4, pp. 42-51, 1956.

[32] M. Papageorgiou, “Some remarks on macroscopic traffic flow modeling,” *Elsevier Science, Ltd,* vol. 32, no. 5, pp. 323-329, 1998.

[33] M.Di Francesco and M.D.Rosini, “Rigorous Derivation of Nonlinear Scalar Conservation Laws from Follow-the-Leader Type Models via Many Particle Limit,” *Archive for Rational Mechanics and Analysis,* vol. 217, no. 3, pp. 831-871, 2015.

[34] A. Kurzhanskiy, “CTMSIM Traffic Macro-Simulator for MATLAB User Guide,” Technical manual, UC Berkeley, 2008.

[35] xxx, “Rules, traffic models, travel strategies,” Techical Report. FP7 ICSI, D5.1.1, 2014.

[36] “Caltrans – California Department of Transportation,” [Online]. Available: http://pems.dot.ca.gov. [Accessed 15 April 2013].

[37] M. Gregurić, M. Buntić, P. Škorput, E. Ivanjko and S. Mandžuka, “State of the Art: Ramp Metering Algorithms, Technical report Nr. A1-01,” Faculty of Transport and Traffic Sciences, University of Zagreb, Zagreb, 2013.

[38] M. Gregurić, E. Ivanjko and S. Mandžuka, “Cooperative Ramp Metering Simulation,” in *International convention on information and communication technology, electronics and microelectronics (MIPRO)*, Opatija, Croatia, May 2014, pp. 1204-1209.

[39] C. F. Daganzo, “The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory,” *Transportation Research, Part B,* vol. 28, no. 4, pp. 269-287, 1994.

[40] G. Gomes and R. Horowitz, “Optimal freeway ramp metering using the asymmetric cell transmission model,” *Transportation Research, Part C,* vol. 14, no. 4, p. 244–262, 2006.

[41] G. Gomes, R. Horowitz, A. Kurzhanskiy, P. Varaiya and J. Kwon, “Behavior of the cell transmission model and effectiveness of ramp metering,” *Transportation Research Part C: Emerging Technologies,* vol. 16, no. 4, pp. 485-513, Aug 2008.

[42] M. van den Berg, A. Hegyi, B. De Schutter and J. Hellendoom, “A macroscopic traffic flow model for integrated control of freeway and urban traffic networks,” in *42nd IEEE International Conference on Decision and Control IEEE Cat. No. 03CH37475, Vol.3*, Mad, Hawaii USA, Dec 2003, pp. 2774-2779.

[43] D. Helbing, A. Hennecke, V. Shvetsov and M. Treiber, “MASTER: macroscopic traffic simulation based on a gas-kinetic, non-local traffic model,” *Transportation Research Part B: Methodological, Volume 35, Issue 2,,* vol. 35, no. 2, pp. 183-211, February 2001.

[44] [Online]. Available: http://vision-traffic.ptvgroup.com/en-uk/products/ptv-vissim/use-cases/mesoscopic-and-hybrid-simulation/. [Accessed January 2017].

[45] [Online]. Available: http://www.its.uci.edu/ctss/sim_models/dynasmart.html. [Accessed January 2017].

[46] R. Jayakrishnan, O. Jun-Seok and A.-E.-K. Sahraoui, “Calibration and Path Dynamics Issues in Microscopic Simulation for Advanced Traffic Management and Information Systems,”

Transportation Research Record: Journal of the Transportation Research Bord, vol. 1771, pp. 9-17, 2001.

[47] [Online]. Available: http://www.its.uci.edu/ctss/sim_models/dynasmart.html.

[48] Official Journal of the European Union, "Directive 2010/40/EU: On the framework for the deployment of Intelligent Transport Systems in the field of road transport and for interfaces with other modes of transport," European Parliament and the European Council, Brussels, Belgium, 2010.

[49] S. Mandžuka, M. Žura, B. Horvat, D. Bićanić and E. Mitsakis, "Directives of the European Union on Intelligent Transport Systems and their impact on the Republic of Croatia," *Promet - Traffic&Transportation*, vol. 25, no. 3, pp. 273-283, May 2013.

[50] A. Hegyi, B. De Schutter and H. Hellendoorn, "Model predictive control for optimal coordination of ramp metering and variable speed limits," *Transportation Research Part C,* vol. 13, no. 3, pp. 185-209, Jun 2005.

[51] N. Geroliminis, A. Srivastava and P. Michalopoulos, "Coordinated ramp metering algorithm for Minnesota's freeways based on density," in *Proceedings of the 13th International IEEE Annual Conference on Intelligent Transportation Systems*, Madeira Island, Portugal, 19-22 Sep 2010 .

[52] M. Gregurić, E. Ivanjko and S. Mandžuka, "The Use of Cooperative Approach in Ramp Metering," *PROMET-Traffic&Transportation,* vol. 28, no. 11, pp. 11-22, 2016.

[53] M. Zhang, K. Taewan, N. Xiaojian and J. Wenlong, "Evaluation of On-ramp Control Algorithms," California PATH Program - Research Report UCB-ITS-PRR-2001-36, University of California, Berkeley, 2001.

[54] S. Sariel, "An integrated planning, scheduling and execution framework for multi-robot cooperation and coordination [PhD thesis]," İstanbul Technical University: Institute of Science and Technology, İstanbul, Turkey, 2007.

[55] C.-H. Wei, "Analysis of artificial neural network models for freeway ramp metering control," *Artificial Intelligence in Engineering,* vol. 15, no. 3, pp. 241-252, Jul 2001.

[56] X. Li and S. Wang, "Adaptive PIP controller based on neural network for time-delay systems and its applications," in *2001 International Conferences on Info-Tech and Info-Net. Proceedings* , Beijing, China, pp. 346-351 ., 2001.

[57] S. Ahn, R. L. Bertini, B. Auffray and O. Eshel, "Evaluating Benefits of System-Wide Adaptive Ramp-Metering Strategy in Portland - Oregon," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 2012, no. 1, pp. 47-56, Dec 2008.

[58] N. Mahajan, "Integrated Approach to Variable Speed Limits and Ramp Metering," Master thesis, Faculty of Civil Engineering and Geosciences(CiTG), Delft University of Technology, Department of Transport and Planning, 2015.

[59] M. Papageorgiou, E. Kosmatopoulos and I. Papamichail, "Effects of Variable Speed Limits on Motorway Traffic Flow," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 1, no. 2047, pp. 37-48, 2008.

[60] P. Rämä and A. Schirokoff, "Effects of weather-controlled variable speed limits on injury accidents, " FIN-02044 VTT, VTT Technical Research Centre of Finland Box 1800, 2010.

[61] M. Gregurić, E. Ivanjko, N. Korent and K. Kušić, "Short review of approaches for variable speed limit control," in *International Scientific Conference Perspectives on Croatian 3pl Industry in Acquiring International, Cargo Flows ZIRP2016*, Zagreb, Croatia, 12. Apr 2016, pp. 41-52.

[62] K. Kušić, N. Korent, M. Gregurić and E. Ivanjko, "Comparison of two controllers for variable speed limit control," in *International Symposium ELMAR*, Zadar, Croatia, 2016, pp. 101-106.

[63] B. Huzjan, M. Šoštarić and S. Mandžuka, "Analysis of speed limit obedience on Croatian highways - Technical report," Faculty of Transport and Traffic Sciences, University of Zagreb, Zagreb, Croatia, 2013.

[64] H. Zoghi, M. Hajali, M. Dirin and K. Malekan, "Evaluation of Passive & Active Intelligent Speed Adaption System," in *International Conference on Computer and Automation Engineering (ICCAE)*, Singapore, Vol. 4, Feb 2010, pp. 182-186.

[65] Y. Zhang and P. A. Ioannou, "Combined Variable Speed Limit and Lane Change Control for Truck-Dominant Highway Segment," in *EEE 18th International Conference*, Las Palmas, Spain, 15-18 Sep 2015.

[66] M. Gregurić, E. Ivanjko and S. Mandžuka, "Learning-Based Control Algorithm for Ramp Metering," in *Autonomic Road Transport Support Systems*, Springer International Publishing, 2016, pp. 197- 213.

[67] R. Kohavi and F. Provost, Glossary of Terms - Special Issue on Applications of Machine Learning and the Knowledge Discovery Process, vol. 30, Boston, USA, Manufactured in The Netherlands: Kluwer Academic Publishers, 1998 pp. 271–274, p. 271–274.

[68] L. Duo and P. Ranjitkar, "A fuzzy logic-based variable speed limit controller," *Journal of Advanced Transportation*, vol. 49, pp. 913-927, 2015.

[69] V. Vichuzhanin, "Realization of a fuzzy controller with fuzzy dynamic correction," *Central European Journal of Engineering*, vol. 2, no. 3, pp. 392-398, 12 Apr 2012.

[70] F. Camastraa, A. Ciaramellaa, V. Giovannellib, M. Lenerb, V. Rastellib, A. Staianoa, G. Staianob and A. Staracec, "A fuzzy decision system for genetically modified plant environmental risk assessment using Mamdani inference," *Expert Systems with Applications*, vol. 42, no. 3, pp. 1710-1716, 15 Feb 2015.

[71] F. Cavallaro, "A Takagi-Sugeno Fuzzy Inference System for Developing a Sustainability Index of Biomass," *Sustainability+*, vol. 7, no. 9, pp. 12359-12371, 2015.

[72] C. P. Pappis and C. Siettos, "Fuzzy reasoning.," in *Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques*, Boston, USA, Springer, Burke, E.K., Kendall, G., Eds.; Kluwer, 2005, pp. 437-474.

[73] V. Spudić, C. Conte, M. Baotić and M. Morari, "Cooperative distributed model predictive control for wind farms," *Optimal Control Applications and Methods*, vol. 36, no. 3, pp. 333-352, 2015.

[74] Mathworks, "Fuzzy logic toolbox documentation," January 2016. [Online]. Available: https://www.mathworks.com/help/fuzzy/fuzzy-inference-process.html.

[75] J. Kacprzyk and W. Pedrycz, Handbook of Computational Intelligence, Berlin Heidelberg: Springer-Verlag, 2015.

[76] H. T. Siegelmann, B. G. Horne and L. C. Giles, “Computational Capabilities of Recurrent NARX Neural Networks,” *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics,* vol. 27, no. 2, pp. 208-215, Apr 1997.

[77] W. Suparta and K. M. Alhasa, Modeling of Tropospheric Delays Using ANFIS, Springer Briefs in Meteorology, 2016.

[78] A. O. Cruz, *ANFIS: Adaptive Neuro-Fuzzy Inference Systems,* Authorized lectures: Mestrado NCE, 2001.

[79] A. Ghods, H. Kian, A. R and M. Tabibi, “GenetiLearning of the Knowledge Base of a Fuzzy System by Using the Linguistic 2-Tuples Representation,” in *The 2005 IEEE International Conference on Fuzzy Systems*, Reno, Nevada, USA, 22-25 May 2005.

[80] European commission, “Intelligent transport systems - Cooperative, connected and automated mobility,” January 2017. [Online]. Available: http://ec.europa.eu/transport/themes/its/road/action_plan/cooperative_systems_en.

[81] S. Butenko, R. Murphey and P. M. Pardalos, Recent Developments in Cooperative Control and Optimizatio - Cooperative systems, New York, USA: Springer, 2004.

[82] Business Dictionary, “Prisoner's dilemma, Game Theory,” [Online]. Available: http://www.businessdictionary.com/definition/prisoner-s-dilemma.html. [Accessed 2017].

[83] M. Abdel-Aty, H. K. Kirolos, R. Cunningham and V. Gayah, “Application of variable speed limits and ramp metering to improve safety and efficiency of freeways,” in *2nd International Symposium on Freeway and Tollway Operations*, Honolulu, Hawaii, 21-24. Jun 2009 pp. 1-13.

[84] E. Ivanjko, D. Koltovska-Nečoska, M. Gregurić, M. Vujić, G. Jurković and S. Mandžuka, “Ramp metering control based on the Q-learning algorithm,” *Cybernetics and Information Technologies; Special Issue on Control in Transportation Systems,* vol. 15, no. 5, pp. 88-97, 2015.

[85] D. Koltovska-Nečoska and K. M. Bombol, “Intelligent Agent Based Traffic Signal Control on Isolated Intersections,” *TEM Journal,* vol. 3, no. 3, pp. 216-222, 2014.

[86] Z. Hou, J. Yan, J. X. Xu and Z. Li, “Modified Iterative-Learning-Control-Based Ramp Metering Strategies for Freeway Traffic Control With Iteration-Dependent Factors,” *IEEE Transactions on Intelligent Transportation Systems,* vol. 13, no. 2, pp. 606-618, 2012.

[87] Y. C. Wang, C. J. Chien and C. H. Wang, “A Fuzzy-Neural Adaptive Iterative Learning Control for Freeway Traffic Flow Systems,” in *International MultiConference of Engineers and Computer Scientists 2016*, Hong Kong, 16 - 18 Mar 2016.

[88] K. Bogenberger, H. Keller and S. Vukanovic, “A neuro-fuzzy algorithm for coordinated traffic responsive ramp metering,” in *ITSC 2001 (Cat. No.01TH8585),* Oakland, USA, 25-29 Aug 2001, pp. 94-99.

[89] M. Gregurić, E. Ivanjko and S. Mandžuka, “A Neuro-fuzzy Based Approach to Cooperative Ramp Metering,” in *IEEE 18th International Conference on Intelligent Transportation Systems*, Las Palmas, Spain, 2015, pp. 54-59.

[90] I. Grabec and F. Šveg, “Statistical forecasting of high-way traffic jam at a bottleneck,” *Advances in Methodology and Statistics*, vol. 9, no. 1, pp. 81-85, Jun 2012.

[91] Z. Chunmei, X. Xiaoli and Y. Changpeng, “The research of method of short-term traffic flow forecast based on ga-bp neural network and chaos theory,” in *International Conference on Information Science and Engineering (ICISE)*, Hangzhou, China, 4 - 6 Dec 2010, pp. 1617-1620.

[92] G. Štefančić, D. Marijan and S. Kljajić, “Capacity and Level of Service on the Zagreb Bypass,” *Promet - Traffic and Transportation*, vol. 24, no. 3, pp. 261-267, 2012.

[93] H. Lenz, R. Sollacher and M. Lang, “Standing waves and the influence of speed limits,” in *European Control Conference (ECC)*, Porto, Portugal, 4-7. Sep 2001, pp. 1228-1232.

[94] H. Stoelhorst, M. Schreuder and S. Polderdijk, “Summary results of dutch field trials with dynamic speed limits (Dynamax),” in *18th ITS World Congress*, Orlando, USA, 16-20 Oct 2011, pp. 1-13.

[95] “Wikipedia,” [Online]. Available: https://en.wikipedia.org/wiki/Hysteresis.

Appendix 1 – Definitions of used Measures of Services

Measures of service (MoS) can be defined as the set of measures for the assessment of the overall motorway LoS. The basic MoS for the assessment of motorway mainstream traffic flow is travel time (TT). The TT is a simple measure which describes the time needed for one vehicle to travel through the observed motorway segment. It is usually measured in minutes. All MoSs which will be explained in this appendix are formulated as part of the ACTM microscopic traffic model. A TT is computed using the following equation:

$$TT = T \sum_{i=1}^{N} 60 \frac{L_i}{v_{i[k]}}, \tag{36}$$

where $v_{i[k]}$ denotes the traffic velocity at the motorway segment i, L_i is the length of the segment i, T is the normalized time step in hours, N is the total number of segments, k is the simulation step, and T is the simulation step length. It is possible to conclude that a high value of a TT is a clear sign of the LoS's quality drop. There are several other quality measures derived from the TT. One of the simplest measures derived from the TT is the Total Travel Time (TTT). The TTT sums up values of the TTs on all observed motorway segments and simulation steps during the entire simulation run [52]. A TT is computed using the following equation:

$$TTT = T \sum_{k=1}^{K} \sum_{i=1}^{N} 60 \frac{L_i}{v_{i[k]}}. \tag{37}$$

Furthermore, it should be emphasised that a TT only provides information about the motorway mainstream throughput. In order to assess other traffic flows on a motorway, it is neccessary to introduce other MoSs. The Total Time Spend (TTS) is the most comprehensive measure, which is originally derived from the TT. The TTS takes into account mainstream density and on-ramp queues. It is expressed in vehicle-hour units. The equation (38) describes the TTS [37]:

$$TTS = T \sum_{k=1}^{K} \sum_{i=1}^{N} (L_i \, n_{i[k]} + T r_{i[k]}), \tag{38}$$

where $n_{i[k]}$ denotes the number of vehicles in motorway segment in time step k, and $r_{i[k]}$ is the number of vehicles merging with the mainstream in cell i. The Total Travel Distance (TTD) presents another measure derived from the TT. It represents the total travelled distance in vehicle-kilometres [veh·km] and can be obtained as:

$$TTD = T \sum_{k=1}^{K} \sum_{i=1}^{N} L_i \, f_{i[k]}, \tag{39}$$

where f_i denotes the number of vehicles leaving the motorway segment *i*. In the assessment of motorway LoSs, it is possible to use the Vehicle Hours Travelled (VHT) and Vehicle Kilometres Travelled (VKT) MoS. The VHT indicates the amount of time spent by all of the vehicles on the motorway in hours. The VKT is defined for a given unit of time and a given section of the motorway. It indicates the sum of kilometres driven by each vehicle on a motorway. The Measure of travel Delay can be computed as the difference between the actual VHT and the respective VHT value a vehicle would travel at free flow speed [9]. It is computed only if the number of vehicles in the motorway segment *i+1* is larger than the critical number of vehicle in the same cell. It is expressed in vehicle-hour units. Equation (40) presents Delay for the motorway segment *i* in time step *k+1* [37]:

$$D_{i[k+1]} = \begin{cases} 0 & if \quad n_{i[k+1]} \leq n^{C}_{i[k+1]} \\ \left(n_{i[k+1]} \times L_i + l_{i[k+1]} \times T - \frac{n_{i[k+1]} \times v_{i[k+1]} \times L_i \times T}{v_i} \right) & if \quad n_{i[k+1]} > n^{C}_{i[k+1]} \end{cases}. \tag{40}$$

Author biography

Martin Gregurić was born in 1988 in Sisak, Croatia. He received his Bachelor and Master Degrees in traffic and transport engineering, course: Intelligent transportation systems in 2008. and 2011, respectively. In November 2011 he enrolled in the Ph.D. study Technological systems in traffic and transport at the Faculty of Transport and Traffic Sciences University of Zagreb. From July 2013 to September 2015 he was employed as a research assistant in the project: Intelligent Cooperative Sensing for Improved Traffic Efficiency, FP7-317671 ICSI. In 2014 he started working at the Faculty of Transport and Traffic Sciences University of Zagreb as a resreach and teaching asistent where he remains to this day. He is involved in lectures of courses: Automatic control in traffic and transport on the undergraduate level, and Intelligent transportation systems I and II, Transport Telematics, and Artificial Intelligence at the graduate level. He received the University of Zagreb Rector award, award "Best graduate student in the course of Intelligent transportation systems in 2010", and won the third prize at the Autonomic Road Transport Support Systems Early Career Researcher Conference (La Valletta, Malta, 2015). He participated in several international COST Training Schools with the main topic of autonomic road transport support systems, first two summer schools of the Croatian Centre of Research Excellence for Data Science and Advanced Cooperative Systems: Research Unit Data Science and the summer school "Intelligent Cars on Digital Roads – Frontiers in Machine Intelligence" organized by the BMW group. His resreach focuses on application of artificial intelligence in traffic control systems, big data analytics in traffic systems, and traffic simulations.

Recent resreach outputs:

Gregurić, M., Ivanjko, E., Mandžuka, S.: The Use of Cooperative Approach in Ramp Metering, PROMET - Traffic&Transportation, Vol. 28, No. 1, Str. 11-22, 2016.

Kušić, K., Korent, M., Gregurić, M., Ivanjko, E.: Comparison of Two Controllers for Variable Speed Limit Control, Proceedings of ELMAR-2016, Zadar, Croatia, Str. 101-106, 2016.

Gregurić, M., Ivanjko, E., Mandžuka, S.: Learning-Based Control Algorithm for Ramp Metering, poglavlje u Autonomic Road Transport Support Systems, Birkhäuser, Str. 197-213, 2016.

Ivanjko, E., Koltovska, N. D., Gregurić, M., Vujić, M., Jurković, G., Mandžuka, S.: Ramp metering control based on the Q-learning algorithm, Cybernetics and Information Technologies, Vol. 15, No. 5, Str. 88-97, 2015.

Mandžuka, S., Ivanjko, E., Vujić, M., Škorput, P.: Gregurić, M.: The Use of Cooperative ITS in Urban Traffic Management, poglavlje u Intelligent Transport Systems: Technologies and Applications, NewYork, USA, Str. 272-288, 2015.

Gregurić, M., Ivanjko, E., Mandžuka, S.: A neuro-fuzzy based approach to cooperative ramp metering, Proceedings of 2015 IEEE 18th International Conference on Intelligent Transportation Systems, Canarias Islands, Spain, 15-18.09.2015, Str. 54-59.

Vujić, M., Mandžuka, S., Gregurić, M.: Pilot implementation of public transport priority in the City of Zagreb, PROMET - Traffic&Transportation, Vol. 27, No. 3, Str. 257-265, 2015.