Road Traffic State Classification Based on Speed Transition Matrix

Tišljarić, Leo

University of Zagreb

Faculty of Transport and Traffic Sciences

Leo Tišljarić

ROAD TRAFFIC STATE CLASSIFICATION BASED ON SPEED TRANSITION MATRIX

DOCTORAL THESIS

Zagreb, 2023

University of Zagreb

Faculty of Transport and Traffic Sciences

Leo Tišljarić

ROAD TRAFFIC STATE CLASSIFICATION BASED ON SPEED TRANSITION MATRIX

DOCTORAL THESIS

Supervisor:

Professor Tonči Carić, Ph.D.

Zagreb, 2023

Sveučilište u Zagrebu

Fakultet prometnih znanosti

Leo Tišljarić

KLASIFIKACIJA STANJA PROMETNE MREŽE PRIMJENOM MODELA ZASNOVANOGA NA PRIJELAZNOJ MATRICI BRZINA

DOKTORSKI RAD

Mentor:

Prof. dr. sc. Tonči Carić

Zagreb, 2023

Doctoral thesis has been made at the University of Zagreb, Faculty of Transport and Traffic Sciences, Department of Intelligent Transport Systems, Chair of Applied Computing.

Mentor: Professor Tonči Carić, Ph.D.

Doctoral thesis type: set of published scientific papers (Scandinavian model).

Doctoral thesis has: 129 pages.

Doctoral thesis committee:

1. Associate Professor Luka Novačko, Ph.D. (committee chair)
   Faculty of Transport and Traffic Sciences, University of Zagreb
2. Professor Tonči Carić, Ph.D.
   Faculty of Transport and Traffic Sciences, University of Zagreb
3. Senior Researcher Tomislav Šmuc, Ph.D.
   Ruđer Bošković Institute
4. Associate Professor Edouard Ivanjko, Ph.D. (replacement)
   Faculty of Transport and Traffic Sciences, University of Zagreb

Date: obo

About the mentor

Tonči Carić received his B.Sc. and the M.Sc. degree in Computer Science from Faculty of Electrical Engineering and Computing in 1993 and 2000 respectively. He defended his doctoral dissertation in 2004 at the Faculty of Traffic and Transport Sciences. He currently works at the Department of Intelligent Transport Systems at the Faculty of Traffic and Transport Sciences as a professor and the head of the Department of Applied Computing. He mentors two postgraduate students and one postdoctoral researcher. In the past five years, he has co-authored 18 scientific papers and collaborated on two projects funded by the European Commission, national research projects and Croatian Science Foundation and one project of the Scientific Center of Excellence for Data Science and Cooperative Systems. His scientific activities successfully connect two fields of technical sciences - computer science and traffic and transportation technology. He has participated in numerous EU projects, professional projects, and technological projects. He was the project leader of the SORDITO project funded by the structural funds of the European Fund. He teaches several courses at the undergraduate, graduate, and doctoral levels. He has mentored five successful students in postgraduate studies.

Web: http://www.fpz.unizg.hr/detalji_nastavnika.asp?izbID=75&user=&korID=26

List of published papers: http://bib.irb.hr/lista-radova?autor=213324

O mentoru

Tonči Carić diplomirao je 1993. i magistrirao 2000. na Fakultetu elektrotehnike i računarstva. Obranio je doktorsku disertaciju 2004. na Fakultetu prometnih znanosti. Radi na Zavodu za Inteligentne transportne sustava Fakulteta prometnih znanosti kao profesor i voditelj Katedre za primijenjeno računalstvo. Mentor je dvojici poslijediplomskih studenata i jednom poslijedoktorandu. U zadnjih pet godina koautor je na 18 znanstvenih radova i surađivao je na dva projekta financiranih od Hrvatske zaklade za znanost i jednom projektu Znanstvenog centara izvrsnosti za znanost o podatcima i kooperativne sustave. Njegovu znanstvenu djelatnost uspješno povezuje dva polja tehničkih znanosti - računarstvo i tehnologiju prometa i transporta. Sudjelovao je u nizu EU projekta, stručnih projekata i tehnoloških projekta. Bio je voditelj projekta "Sustav za Optimizaciju Ruta u Dinamičkom Transportnom Okruženju-SORDITO" koji se financirao iz strukturnih fondova Europskog fonda. Nositelj je više kolegija preddiplomskog, diplomskog i doktorskog studija. Bio je mentor petorici uspješnih studenata na poslijediplomskim studijima.

Web: http://www.fpz.unizg.hr/detalji_nastavnika.asp?izbID=75&user=&korID=26

Popis objavljenih radova: http://bib.irb.hr/lista-radova?autor=213324

If we could change ourselves, the tendencies in the world would also change. As a man changes his own nature, so does the attitude of the world change towards him. We need not wait to see what others do.

Mahatma Gandhi

Acknowledgements

This research has been supported by the European Regional Development Fund under the grant KK.01.1.1.01.0009 (DATACROSS). This research is partially supported by Croatian Science Foundation under project IP-2018-01-8323. Data used for this research is collected during project SORDITO, European Regional Development Fund under contract RC.2.2.08-0022.

I would like to thank my thesis and life mentor, Tonči Carić for guiding me for the last seven years from the lost student on the second bachelor's year to the end of the PhD path. With his advice, sometimes consciously, sometimes unconsciously, he helped me to become the person that I am today. Thank you, Tonči!

Family, obviously without them and their way of supporting me, this thesis wouldn't exist. Their whole concept of life, which was transferred to my brothers and me, helped us to become tough machines that just keep going whatever you put in front of them. Thank you all!

The Erdelić family, Martina and Tomislav, were like my mentors when Tonči was not around. They supported me all the time no matter my rookie questions, over ambitious ideas, and the will to change the world. Thank you, Erdelići!

Dominik Cvetek, my friend, this whole academic journey would be much different (or it would not exist) if he was not around. Everything started on the second bachelor's year when we started working together with the Pioneer robot in a basement laboratory at Faculty. It continued with all kinds of joint undertakings which led to us both working as research assistants. Thank you, Dominik!

And last, but most important, the love of my life, Lidija. Who supported me from the moment we met, and I hope that she will support me to the end of our days. She showed me how to use my hidden emotional potential, which many of us just dig deep into ourselves. And of course, maybe the most important thing, she showed me that life is not all about work, it is much, much more. Thank you, Lidija!

Abstract

The accelerated development of communication technologies in this decade has allowed the use of large spatiotemporal traffic datasets for research related to intelligent transport systems (ITS). ITS applications based on large sets of traffic data enable the solution to problems arising from the growth of urban centers around the world. Some of the problems of increasing traffic demand in urban centers are manifested through traffic congestion that leads to increased travel time on the transport network and environmental pollution due to increased exhaust gases leading to reduced quality of life. The aim of this doctoral thesis is to develop methods for classifying different states of the traffic network using the proposed traffic data representation named the Speed Transition Matrix (STM). The thesis will present three methods of classifying the state of road traffic networks based on the proposed STM traffic data representation. The methods relate to classifying traffic network congestion in urban areas, detecting anomalies in traffic patterns on the urban transport network, and detecting bottlenecks on motorways due to increased traffic demand. The result indicates that STM can be used for mentioned tasks with high accuracy. Obtained methods can be used in traffic monitoring and control systems in a variety of use cases like bottleneck prevention, traffic lights optimization, and delivery vehicles routing.

Keywords: speed transition matrix, machine learning, anomaly detection, traffic state estimation, traffic congestion estimation, motorway bottlenecks, traffic data modeling

Prošireni sažetak

KLASIFIKACIJA STANJA PROMETNE MREŽE PRIMJENOM MODELA ZASNOVANOGA NA PRIJELAZNOJ MATRICI BRZINA

Ubrzani razvoj komunikacijskih tehnologija u ovom desetljeću omogućava korištenje velikih skupova prostorno-vremenskih prometnih podataka za istraživanja vezana za inteligentne transportne sustave (ITS). ITS aplikacije zasnovane na velikim skupovima prometnih podataka omogućavaju rješavanje problema koji nastaju zbog rasta urbanih centara diljem svijeta. Neki od problema povećanja prometne potražnje u gradskim središtima se očituju kroz prometna zagušenja koja dovode do povećanja vremena putovanja na prometnoj mreži i zagađenja okoliša zbog povećane količine ispušnih plinova što dovodi do smanjenja kvalitete života. Cilj doktorskog rada je razviti metode za klasifikaciju različitih stanja prometne mreže modeliranjem prometnih podataka prijelaznim matricama brzina (eng. SpeedTransition Matrix - STM). U radu će se prikazati tri metode klasifikacije stanja cestovne prometne mreže koje se temelje na predloženom STM prikazu podataka. Metode se odnose na klasifikaciju zagušenja prometne mreže u gradskim sredinama, detekciju anomalija u prometnim uzorcima na gradskoj prometnoj mreži i detekciju uskih grla na autocestama zbog povećane prometne potražnje.

STM je nova metoda prikaza prometnih podataka koja se ovom disertacijom predlaže kao vizualizacijska metoda i ulazni podatak za metode klasifikacije stanja prometne mreže. STM se zasniva na matrici koja prikazuje vjerojatnost prijelaza (promijene) između ulazne i izlazne brzine na promatranoj tranziciji, gdje tranzicija predstavlja kretanje vozila između dva cestovna segmenta na prometnoj mreži u određenom vremenskom intervalu. Svaka tranzicija sadrži dva povezana cestovna segmenta, označenih kao ulazni i izlazni, ovisno o smjeru kretanja vozila. Vjerojatnosti prijelaza brzine prikazani STM-om vizualno se grupiraju u prometne uzorke koji se mogu koristiti za klasifikaciju stanja prometne mreže procjenom njihove relativne pozicije u STM-u. Stoga je najvažniji parametar za određivanje stanja prometne mreže korištenjem STM prikaza centar mase (engl. Center of Mass - CoM) prikazanog prometnog uzorka. Parametar CoM prikazuje poziciju uzorka na temelju kojeg se klasificira stanje prometne mreže temeljem vjerojatnosti prijelaza brzine na promatranoj tranziciji.

Ova disertacija pisana je prema takozvanom Skandinavskom modelu u sklopu koje je objavljeno tri znanstvena rada koji prikazuju originalni doprinos u području prometa i transporta.

Prvi rad prikazuje razvoj metode detekcije prometnih zagušenja na temelju STM prikaza prometnih podataka . U ovom dijelu istraživanja, atributi koji predstavljaju položaj prometnog uzorka prikazanog STM-om, bit će izdvojeni ekstrakcijom koordinata CoM-a. Izdvojeni atributi će se zatim koristiti kao ulaz u model za klasifikaciju stanja prometne mreže na razini grada. Za ovaj dio istraživanja koristit će se skup podataka koji sadrži GNSS pozicijske podatke. Grad Zagreb je izabran za eksperimentalno područje kao europski grad srednje veličine s velikim brojem prikupljenih podatkovnih točaka. Temeljem relevantne literature i domenskog znanja, svakoj izračunatoj STM je dodijeljeno jedno od stanja prometne mreže i to "slobodni prometni tok", "stabilni prometni tok" ili "prometno zagušenje". Nakon provedene validacije nad stvarnim označenim prometnim podacima i podacima označenim pomoću stručnog priručnika Highway Capacity Manual (HCM), metoda je ostvarila vrijednosti točnosti od 97%, odnosno 91%. U ovom radu metoda je primijenjena za analizu stanja prometne mreže na Zagrebačkim mostovima. Rezultati istraživanja ukazuju na neefikasnu prometnu signalizaciju i upravljanje prometnim tokovima.

Drugi rad prikazuje razvoj metode detekcije neočekivanih događaja (anomalija) u velikom skupu podataka o gradskom prometu. Ovaj dio istraživanja ima za cilj klasificirati prometne uzorke prikazane STM-ovima kao očekivane ili ne, na temelju atributa izdvojenih izračunavanjem CoM-a. Glavna pretpostavka je da se anomalija može otkriti promatranjem promjena u položaju CoM-a unutar STM-a. Povećavanjem Euklidske udaljenost između CoM-a i glavne dijagonale STM-a, raste vjerojatnost da će prometni uzorak biti klasificiran kao neočekivan. Na pozicijama koje su udaljenije od glavne dijagonale, mogu se uočiti dvije vrste prostornih prijelaza: (i) prijelazi s visokih izvorišnih na niske odredišne brzine, koji predstavljaju nagla kočenja vozila (donji lijevi kut STM-a), i (ii) prijelazi s niske izvorišne brzine do visokih odredišnih brzina koje predstavljaju intenzivna ubrzanja (gornji desni kut STM-a). Predložena metoda detekcije anomalija temeljit će se na modelu dekompozicije tenzora. Ulazni tenzori će se formirati pomoću STM-a unutar promatranih vremenskih intervala, koji će rezultirati tenzorom dimenzija $M \times N \times T$, gdje M predstavlja duljinu dimenzije STM-a, N broj promatranih prijelaza i T broj promatranih vremenskih intervala. Zatim će se dekompozicija tenzora koristiti za izdvajanje prometnih uzoraka na razini cijelog gradskog područja. Metoda je uspješno korištena za detektiranje pozicija u gradskom području koja imaju najveći potencijal za stvaranje zagušenja uzrokovanih anomalijama u prometnim tokovima. Također, rezultat metode je pokazao lokaciju i vrijeme kada se navedena zagušenja mogu očekivati.

Treći rad prikazuje razvoj metode detekcije uskih grla na autocestama, koja predstavlja proširenje predložene metode za detekciju anomalija. Ovo je prirodni nastavak jer su uska grla često uzrokovana neočekivanim događajima na autocestama. Glavni cilj je izdvojiti STM atribute zasnovane na CoM-u i koristiti ih kao ulazni parametar metode strojnog učenja za procjenu vjerojatnosti uskog grla. Algoritam zasnovan na primjeni neizrazite logike, dodatno optimiziran genetskim algoritmom, korišten je za procjenu vjerojatnosti nastanka uskog grla. Glavne varijable korištene za algoritam su izdvojene iz STM-a, a predstavljaju udaljenost CoM-a od glavne dijagonale i udaljenost istog od izvorišta STM-a. Za ovaj dio istraživanja korišten je sintetički skup prometnih podataka izdvojen iz mikrosimulacijskog softvera. Metoda je evaluirana na različitim prometnim scenarijima koji uključuju stvaranje uskih grla na autocestama. Scenariji uključuju sudar na autocesti, visok dotok vozila s ulazne rampe koji dovodi do stvaranja uskog grla i takozvano pomično usko grlo uzrokovano velikim brojem vozila teške kategorije vozila na autocesti.

Objavljeni radovi zajednički čine cjelinu i znanstveni doprinos u području prometa i transporta u smislu prijedloga tri metode zasnovane na korištenju nove metode prikaza prometnih podataka naziva STM. Znanstveni doprinosi su sljedeći:

1. Razvoj metode klasifikacije stanja prometne mreže zasnovane na matrici prijelaznih brzina.

2. Razvoj metode detekcije anomalija na prometnoj mreži zasnovane na matrici prijelaznih brzina.
3. Razvoj metode detekcije i procjene vjerojatnosti nastanka uskih grla na autocestama zasnovane na matrici prijelaznih brzina.

Ključne riječi: prijelazna matrica brzina, strojno učenje, detekcija anomalija, procjena stanja prometne mreže, uska grla na autocestama

Contents

Chapter 1

Introduction

Economic, demographic, and technology development act as enablers in support of the increase in the rising need of human mobility, especially in large urban environments. The increased need for mobility inevitably leads to innovations in the traffic management and control domain to cope with the challenges of the increased traffic demand. Aside from the positive effects of the increased mobility, it also results in negative effects mostly related to traffic congestion, like increased travel time and pollution. In a report from the European Commission (European Commission 2017), the authors state that congestion due to the increased mobility accounts for 40% of all CO2 emissions of the European road transport with the 70% of other pollutants from transport. Consequently, the total indirect cost of the congestion, on the scale of the European Union is approximately €100 billion, which is about 1% of the total annual gross domestic product.

The research in the field of the Intelligent Transport Systems (ITS) enable the usage of the large traffic dataset that are collected to cope with the challenges of the increased traffic demand (Nkoro and Vershinin 2014). The first step in dealing with traffic congestion is the traffic state estimation and classification. In this thesis, three methods for the traffic state classification are proposed based on a novel traffic data representation named Speed Transition Matrix (STM). The proposed methods are related to city-wide traffic congestion estimation, anomaly detection and detection of bottlenecks on motorways.

1.1 Speed Transition Matrix

STM is a novel traffic data representation method. It is based on a matrix which shows the probability of change between origin and destination speeds on the observed transition, where the transition represents the movement of vehicles between two road segments on the traffic network in a defined time interval. Each transition contains two consecutive road segments, marked as origin and destination, depending on the vehicle movement direction. This thesis shows the use cases of applying the STM for visualization of traffic data and as an input for road traffic congestion clustering, anomaly detection, and bottleneck detection methods.

As a matrix data representation, the STM can be represented by expression:

$$X^{(ij)}(\Delta t) = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ p_{m1} & \cdots & \cdots & p_{mn} \end{pmatrix}, \tag{1.1}$$

where $X^{(ij)}$ represents the STM which is computed from the origin edge e_i to the destination edge e_j representing the ij-th transition at the observed traffic network, Δt represents data collection time interval, which is used to discretize the time into buckets, ex. 5 min, 15min or 1h. The p_{mn} represents one cell in the STM which represents the probability that vehicle will have origin speed m and destination speed n in ij-th transition during interval of Δt.

Regarding its matrix form, the STM can be represented visually as a heatmap, where ordinate (y coordinate) represents the relative origin speed and abscissa (x coordinate) represents relative destination speed. Figure 1.1 presents two possible STMs computed for traffic road network's transitions. Figure 1.1 (left) shows the congested traffic due to the small origin and destination speeds. On the other hand, Figure 1.1 (right) shows the normal traffic due to high origin and destination speeds.

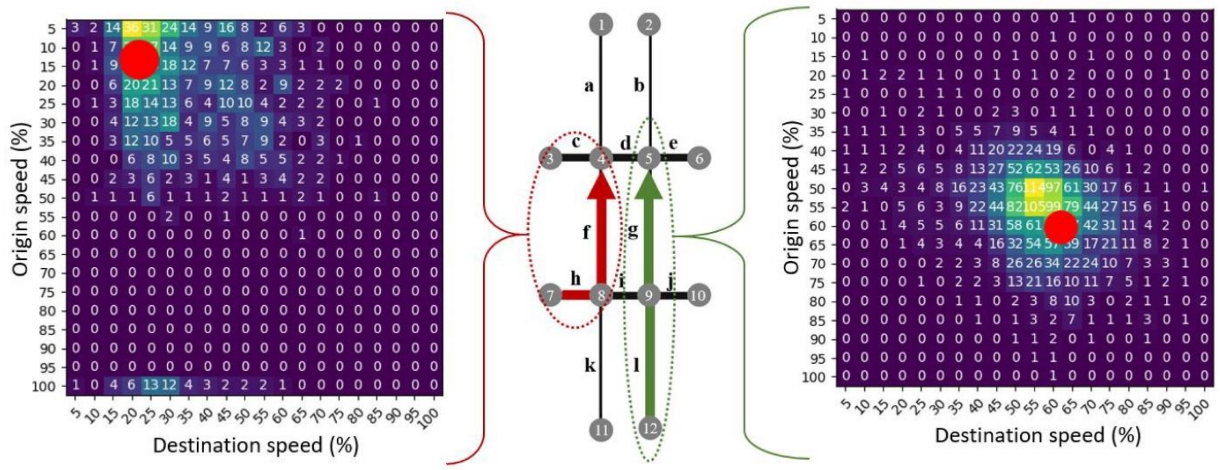


Figure 1.1 Examples of the STMs: (left) congested traffic represented with low origin and destination speeds and (right) normal traffic represented with high origin and destination speeds.

As can be observed in the figure above, transition probability values represented by the STM form a traffic pattern. Its position in the matrix is the most important feature which can be extracted for a traffic state estimation as it directly shows the transitions speeds on the observed road segments. Furthermore, the most important parameter for determining the traffic state using the STM model is the Center of Mass (CoM) (Jordaan 2005) of the presented traffic pattern. The CoM represents the position of the pattern extracted from the STM, which represents traffic state on the network, based on the probability of speed change at the observed transition.

The STM-based traffic data representation captures spatiotemporal correlations between consecutive road segments. This is achieved by capturing speed transition pairs, which form a traffic pattern. Furthermore, this pattern will consequentially show the average speed pairs weighted with the spatial correlations of the observed origin and destination road segments. This approach can be compared with the methods which include more sensor readings of the same parameter to achieve greater accuracy like Kalman filter or Hough transform.

1.2 Literature review

This section reviews the most important literature for every topic that is crucial for the development of the proposed methods. A more detailed review can be found in chapters containing published papers where each paper includes a dedicated literature review section.

1.2.1 Traffic data modeling

When modeling traffic data, the authors most often use models based on aggregating a large amount of collected data on traffic parameters such as speed, density, and traffic flow. Parameters are modeled with vectors that show the traffic parameter change in one day divided into several time intervals of the most common intervals of 5, 15, 30 or 60 minutes, which are called profiles (Erdelić et al. 2021). The problem with this type of modeling arises due to the aggregation of many different values within narrow time intervals, which leads to large deviations in models. Also, the profiles show the change of the traffic parameter in the time domain, which prevents the simultaneous use of space-time data. More complex models, based on matrices, enable the analysis of spatio-temporal data (Ma et al. 2017a; Nguyen, Liu, and Chen 2017). Such matrices most often show a change in the traffic parameter in the matrix cells, where the rows of the matrix show the road segments and the columns of the matrix the time intervals. Another type of matrix models for displaying traffic data are origin-destination matrices (Rao et al. 2018). The cells of such a matrix contain values of the number of vehicles that traveled between the origins shown in the rows and destinations shown in columns of the matrix. The disadvantage of origin-destination matrices is the inability to reconstruct the route used by the vehicle between the origin and the travel destination. To overcome the mentioned shortcomings of vector and matrix models, the use of STM representation is proposed in this thesis. The STM representation does not aggregate data into a single scalar value within narrow time intervals, and it contains all recorded speed transition values in a matrix. In addition to the above advantages, STM can also be used when data is collected from delivery vehicles, which is not the case when using the classic origin-destination matrix.

1.2.2 Traffic congestion classification

The goal of the methods for classifying the traffic network's congestion is to group the road segments based on traffic states into classes that describe the level of congestion. The authors in the known research (Kerner 2002) present the classification of congestion on road segments using the so-called three-phase traffic model where traffic congestion is divided into classes "moving congestion", "synchronized traffic flow" and "free traffic flow". In (Nguyen et al. 2019), the authors use convolutional neural networks to classify traffic images and divide the traffic congestion of the network into "isolated", "low frequency", "high frequency", "homogeneous" and "mixed". In (Jin, Srinivasan, and Cheu 2001; Thianniwet and Phosaard 2009), the authors use speed as the main traffic parameter for classification. The condition of the transport network is divided into three classes "congestion", "high traffic load" and "low traffic load".

In this thesis, the traffic congestion on the network is divided into three classes, "free traffic flow", "unstable traffic flow" and "congestion", respectively. The proposed classification method is based on the calculation of the CoM of each of the STMs showing the traffic pattern on the observed road. By estimating the CoM coordinates, the process of classifying the state of the transport network has been reduced to the problem of classifying points in a two-dimensional coordinate system. This method speeds up and simplifies the process of classification of traffic patterns.

1.2.3 Anomaly detection

The goal of anomaly detection methods is to identify data that deviate from the defined normal state and can thus be classified as unexpected. There are two types of traffic anomalies in the research literature: recurrent traffic anomalies and non-recurrent traffic anomalies. The recurrent traffic anomalies are referred to the ones which can happen in the recurring manner like rush hours. They are classified as anomalies even regarding the fact that recurrent behavior can be modeled and classified and as such do not follow the classic anomaly definition of unexpected events. In the context of road traffic research, the term recurring anomaly is used because rush hours produce extreme congestion (compared to the normal traffic) where its impact cannot be exactly modeled and as such can be defined as anomalies. The second type of anomalies are non-recurrent ones which represent unexpected traffic behavior. This anomalous traffic behavior is often triggered by anomalous events like traffic accidents, large social events or similar.

The first step in implementing anomaly detection methods is to define normal and unexpected behavior based on domain knowledge and data analysis. Review papers in this research topic (Chandola 2009; Schubert, Zimek, and Kriegel 2014) show that most methods for detecting anomalies are based on statistical methods and data mining methods. The authors in (Gupta et al. 2014) present methods for detecting anomalies in time series, while the authors in (Feng and Zhu 2016; Zheng 2015) present methods for detecting anomalies in spatial traffic data. In the cited literature, it is possible to notice the lack of development of methods that include a combination of spatial and temporal data with an emphasis on the detection of traffic anomalies. Anomalies in traffic data often cannot be detected using simple statistical models. Due to the above, this paper proposes a method for the detection of traffic anomalies by modeling data into traffic tensors using STM. The proposed method will include spatial and temporal data and will be able to be used regardless of whether the data were collected on motorways or on urban roads.

Many studies in the field of detecting anomalies in traffic data are commonly used for developing bottlenecks detection methods. This is justified by the fact that bottlenecks are most often caused by unexpected events on the roads. Bottleneck is a phenomenon that manifests itself through a reduction in traffic flow at the downstream traffic flow in relation to upstream traffic flow due to traffic congestion. Traffic congestion can occur due to an unexpected event or daily increases in traffic congestion due to passenger migration. The authors in (Coifman and Kim 2011) analyze the existing methods for bottleneck detection and conclude that traditional models based on the fundamental traffic diagram are not sufficient for bottleneck detection due to sudden changes in traffic parameters. The authors in (Li et al. 2020; Sun et al. 2014) propose methods for the detection of bottlenecks by calculating the congestion index based on traffic parameters. In (Kerner 2007), the authors use Kerner's three-phase traffic model to detect bottlenecks on highways. The bottleneck was detected on the parts of the road where the transition from the traffic state "free traffic flow" to the state "synchronized traffic flow" was observed. Continuation of research (Dülgar et al. 2020) showed that bottlenecks can occur unexpectedly along the road in transitions between different states of the traffic network. These papers have demonstrated the importance of observing the transition between the state of the traffic network to detect unexpected events that may lead to the creation of bottlenecks on highways.

Based on previous research to observe the transient states of the transport network, this dissertation proposes a method for the detection of bottlenecks using the STM representation. The representation is based on observing the vehicle speed transition between two connected highway road segments. Unlike the methods in the revised literature, the method will be able to detect existing bottlenecks on highways, which is the main disadvantage of the proposed methods.

1.3 Research methods and materials

The research requires the appliance of the methods, including computer data analysis, synthesis, simulation, and statistical analysis. The computer workstation with appropriate software was used, including MongoDB database for storing the preprocessed data, Simulation of Urban MObility (SUMO) (Chowdhury, Santen, and Schadschneider 2000) road traffic simulation software, and Python programming language with the corresponding packages for machine learning like Scikit-Learn, Tensor Flow, Pandas, and Numpy.

Other required materials are: (i) Digital traffic map of Croatia and ≈7 billion of GNSS records collected from delivery vehicles between 2009 and 2014, provided by Mireo Inc. as a part of SORDITO project (RC.2.2.08-0022). This dataset will be used to develop traffic congestion estimation and anomaly detection methods in large urban areas. (ii) Vehicles positions data extracted from the motorway simulation developed in the SUMO environment. This dataset will be used to develop a method for bottleneck detection on the motorways, which can be applied in urban and transit motorway contexts.

The research methodology is divided into four parts related to the goals and thesis hypothesis: (i) raw traffic data preprocessing, (ii) traffic congestion estimation method development, (iii) anomaly detection method development and (iv) motorway bottleneck detection method development. The methodology overview is shown in Figure 1.2 where columns are representing the name of the developed method while rows are representing development stages. The color of a component in the diagram defines the affiliation to the developed method except the purple color which represents common stages for the development of every method.

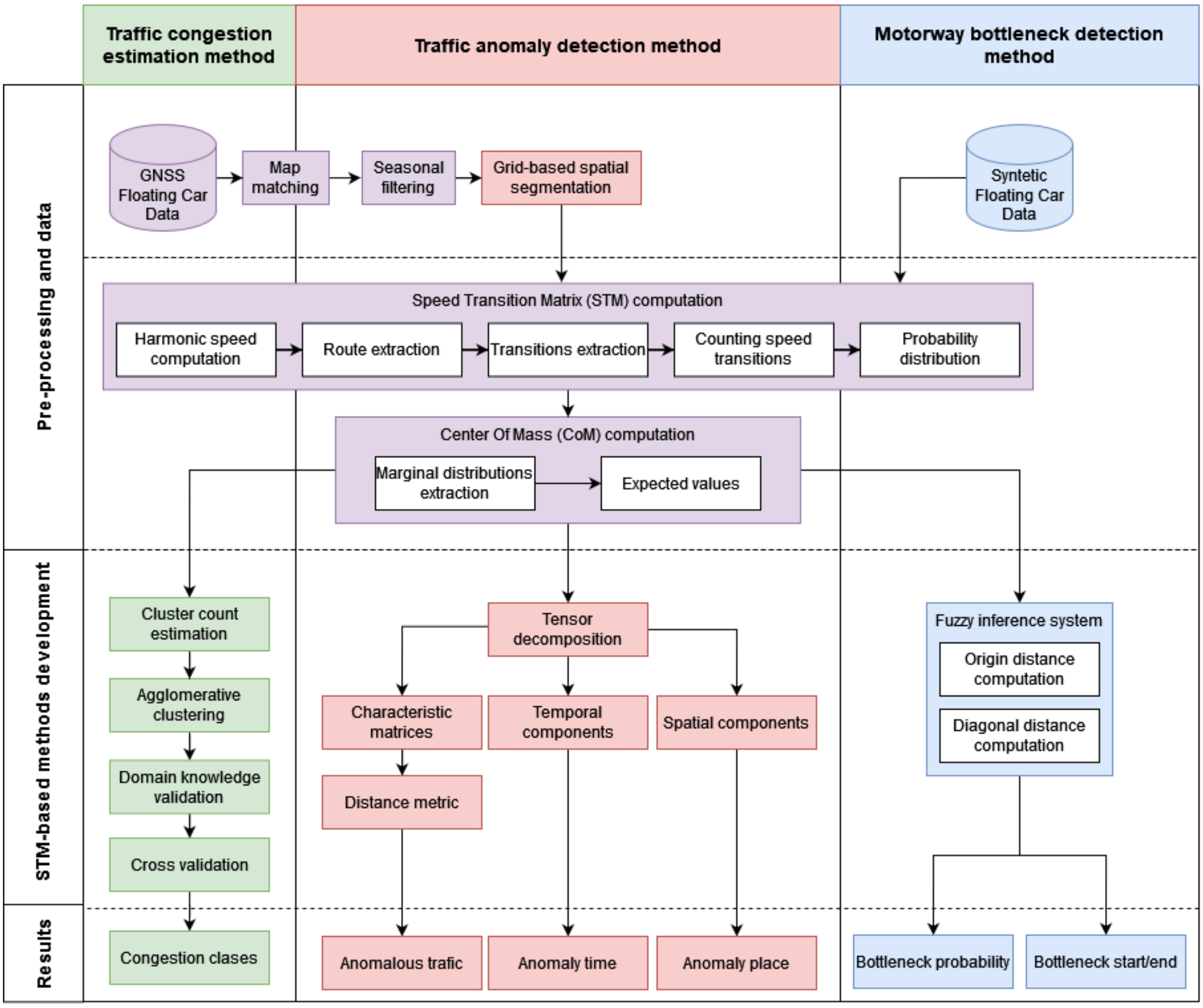


Figure 1.2 Research methodology overview

The main goal of the first part of the research is to preprocess large positional datasets to compute STMs. The first dataset contains ≈7 billion GNSS records collected from delivery vehicles. There are four steps in preprocessing the first dataset: (i) extracting the road segments from a digital map of Croatia, (ii) for every vehicle, routes are extracted on a link-level with computing the vehicle's harmonic mean speed for every link on the route, (iii) construct the STM for every transition in the dataset, and (iv) save the STM data to database. The second dataset is extracted from the SUMO simulation environment to represent motorway traffic data. Preprocessing steps (ii) to (iv) are repeated on the simulation dataset.

The second part of the research is related to developing the traffic congestion estimation method based on the STM representation. In this part of the research, attributes that represent the traffic pattern position in the STM will be extracted by extracting the CoM coordinates. Extracted attributes will then be used as an input for the traffic congestion classification model, which will be used for the classification of traffic states on a city-wide scale. For this part of the research, the first dataset that contains real-life GNSS traces will be used. As an experimental area, the City of Zagreb is chosen as a medium-sized European city with a large number of data points.

The third part of the research is related to developing the anomaly detection method for detecting anomalous events in the large-scale urban traffic dataset. This part of the research aims to classify the traffic patterns represented with STM as anomalous or not based on the attributes extracted by computing the CoM. The main assumption is that anomaly can be detected by observing the changes at the CoM's position inside the STM. If the Euclidian distance between the CoM and the main diagonal of the STM, the anomaly should be higher. In positions more distant from the main diagonal, two kinds of transitions can be observed: (i) transitions from high origin to low destination speeds, which represent sudden breaks (lower-left corner of the STM), and (ii) transitions from low origin speeds to high destination speeds which represent intense accelerations (upper-right corner of the STM). The proposed anomaly detection method will be based on the tensor decomposition model. The input tensors will be formed using STMs within the observed time intervals, which will provide a tensor with the dimension's length of the STM, times number of the observed transitions, times observed time intervals. Then, tensor decomposition will be used to extract the traffic patterns on a city-wide scale. For this part of the research, Zagreb is also chosen as an experimental area representing the first dataset.

The last part of the research includes the development of the motorway bottleneck detection method, which represents the extension of the proposed anomaly detection method. This is a natural extension because the bottlenecks are often caused by unexpected, anomalous events on the motorways. The main goal is to extract the STM attributes based on CoM computation and use them as an input parameter to the machine learning method for bottleneck probability estimation. The second dataset extracted from the SUMO simulation will be used for this part of the research. The method will be evaluated on different traffic scenarios that include a bottleneck formation of the motorway. The scenarios will include the collision site, high on-ramp inflow and moving bottleneck caused by a large number of heavy-duty vehicles.

1.4 Thesis objectives and scientific contributions

The thesis objective is to propose the usage of STM as a novel traffic data representation, by developing three STM-based methods which include traffic congestion classification, anomaly detection, and bottlenecks detection. STM as a data representation and STM-based methods are the main contributions of this thesis which is highlighted within the research objectives and hypotheses.

Development of the STM-based methods resulted in three hypotheses:

1. Speed transition matrix can be used for a development of the traffic congestion classification method on a road traffic network.
2. Speed transition matrix can be used for a development of the tensor-decomposition-based method for anomaly detection on a road network.
3. Speed transition matrix can be used for a development of the method for traffic bottleneck detection on motorways and probability of their occurrences.

Every research objective is followed with a corresponding hypothesis which resulted in following papers which resulted with several scientific contributions:

Paper A: Traffic State Estimation and Classification on Citywide Scale Using Speed Transition Matrices.

Paper B: Spatiotemporal Road Traffic Anomaly Detection: A Tensor-Based Approach.

Paper C: Motorway Bottleneck Probability Estimation in Connected Vehicles Environment Using Speed Transition Matrices.

The first research hypothesis is addressed in **Paper A**, where STM representation is used as an input to clustering method for traffic congestion state estimation by proposing the CoM as a measure for estimating congestion level. The method included three main steps: i) preprocessing large GNSS dataset containing vehicle's geo-location to extract STMs, ii) reduce dimensionality of STM using CoM, and iii) based on domain knowledge, classify STM represented traffic congestion into one of three categories, namely, "Free flow", "Stable flow", and "Congestion". With this paper the traffic congestion classification method based on the STM is produced and validated which represents the first scientific contribution. The advantage of the proposed method can be seen in the reduced computation complexity of the cluster analysis process and the increased interpretability of obtained clusters. The method is validated on real-life dataset with accuracy of 97% and a dataset labeled with the extracted domain knowledge from the HCM manual with the accuracy score of 91%.

The second hypothesis is addressed in **Paper B**, where STM is used as an input for development of the anomaly detection method. In this context traffic anomaly is defined as a recurrent anomaly. From the traffic flow standpoint, there are two types of anomalies, non-recuring (traffic accidents) and recuring (traffic jams). Traffic jams can be identified as a recurring anomaly due to the stochastic traffic nature which is mostly affected by human-driven decisions which cannot be precisely predicted. In this paper, preprocessing step contained dividing the research location (mid-sized European city) into quadratic shaped geographic areas for which tensors are constructed using GNSS data, and anomaly is estimated using proposed anomaly detection method. The anomaly detection method relies on computing the CoM and measuring Euclidean distance between the STM diagonal and the CoM. The method resulted in anomalous traffic patterns that can be identified on a city-level scale with the ability to estimate the location and the time of the anomaly occurrences. With this paper the traffic anomaly detection method based on the STM is produced and validated which represents the second scientific contributions.

The third hypothesis is addressed in **Paper C**, combining the knowledge about the STM patterns to explore the possibility to detect the bottlenecks in connected vehicles environment at motorways. The paper introduces novel STM-based bottleneck detection which can estimate the probability of bottleneck's occurrence which can be used to predict and prevent bottlenecks at motorways. With this paper, the method for motorway bottlenecks detection and probability of their occurrences based on the STM is produced, which represents the third and the final scientific contribution.

Based on the previous description of the thesis hypothesis and the published papers, the major scientific contributions of this thesis are:

1. Development of the traffic state classification method based on the speed transition matrix.
2. Development of the traffic anomaly detection method based on the speed transition matrix.
3. Development of the method for motorway bottlenecks detection and probability of their occurrences based on the speed transition matrix.

1.5 Thesis outline

This thesis is written in the form of a set of published scientific papers, the so-called "Scandinavian model" according to the Ordinance on doctoral studies at the University of Zagreb and the recommendations of the Commission for Postgraduate Studies and Doctorates of the Faculty of Transport and Traffic Sciences.

Within the recommendation scope for forming this kind of thesis, it is divided into three main parts, namely, the introduction (Chapter 1), published papers with original scientific contributions (Chapter 2 – Chapter 4), and conclusions (Chapter 5 and 6).

Chapter 1 provides the introduction to the thesis research topic and presents the most important literature as a basis for topic relevance and scientific contributions. It also provides a list of hypotheses correlated with the published papers and the scientific contributions.

Chapter 2 presents the first published paper entitled *Traffic state estimation and classification on citywide scale using speed transition matrices*. The paper presents the methodology for traffic congestion estimation by estimating the congestion level using STMs and the domain-based knowledge.

Chapter 3 presents the second paper entitled *Spatiotemporal Road Traffic Anomaly Detection: A Tensor-Based Approach*. The paper presents the anomaly detection method based on construction of the traffic data tensor based on the STMs. The proposed anomaly detection method relies on anomalous traffic patterns extraction based on the anomalous pattern shown by the data position inside STMs.

Chapter 4 presents the third paper entitled *Motorway Bottleneck Probability Estimation in Connected Vehicles Environment Using Speed Transition Matrices*. This paper presents the method for detecting bottlenecks at motorways using gained knowledge about the STM pattern extraction and the importance of the STM pattern position.

Chapters 5 and 6 present the joint discussion about the contributions of each paper and the thesis with providing conclusions and the future research directions.

Chapter 2

Traffic state estimation and classification on citywide scale using speed transition matrices

This chapter has been published as: Tišljarić L, Carić T, Abramović B, Fratrović T. Traffic State Estimation and Classification on Citywide Scale Using Speed Transition Matrices. Sustainability. 2020; 12(18):7278. https://doi.org/10.3390/su12187278.

Author Contributions: Conceptualization, L.T. and T.C.; Methodology, L.T.; Software, L.T.; Validation, L.T.; Formal analysis, L.T.; Resources, T.C.; Writing–original draft preparation, L.T.; Writing–review and editing, L.T., T.C., T.F., B.A.; Visualization, L.T.; Supervision, T.C., T.F., B.A.; Funding acquisition, T.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been supported by the European Regional Development Fund under the grant KK.01.1.1.01.0009 (DATACROSS). This research is partially supported by Croatian Science Foundation under project IP-2018-01-8323. Data used for this research is collected during project SORDITO, European Regional Development Fund under contract RC.2.2.08-0022.

2.1 Abstract

The rising need for mobility, especially in large urban centers, consequently, results in congestion, which leads to increased travel times and pollution. Advanced traffic management systems are being developed to take advantage of increased mobility positive effects and minimize the negative ones. The first step dealing with congestion in urban areas is the detection of congested areas and the estimation of the congestion level. This paper presents a a method for a traffic state estimation on a citywide scale using the novel traffic data representation, named Speed Transition Matrix (STM). The proposed method uses traffic data to extract the STMs and to estimate the traffic state based on the Center Of Mass (COM) computation for every STM. The COM-based approach enables the simplification of the clustering process and provides increased interpretability of the resulting clusters. Using the proposed method, traffic data is analyzed, and the traffic state is estimated for the most relevant road segments in the City of Zagreb, which is the capital and the largest city in Croatia. The traffic state classification results are validated using the cross-validation method and the domain knowledge data with the resulting accuracy of 97% and 91%, respectively. The results indicate the possible application of the proposed method for the traffic state estimation on macro- and micro-locations in the city area. In the end, the application of STMs for traffic state estimation, traffic management, and anomaly detection is discussed.

Keywords: speed transition matrix; traffic state estimation; traffic state classification; center of mass; intelligent transport systems; speed probability distribution

2.2 Introduction

Demographic, economic, and technological changes and developments are enablers that support the increase in the human need for mobility, especially in large urban centers. The increase in the need for mobility leads to advanced solutions in the traffic management domain and requires the implementation of Intelligent Transport Systems (ITS) solutions and applications (Gregurić, Mandžuka, and Vidović 2020). Aside from the positive effects, it also has negative effects, such as increased congestion or pollution in urban areas.

Sustainable transport development is often confronted by traffic congestion. The European Commission reports that congestion that is caused by increased mobility accounts for 40% of all CO2.

emissions of road transport and up to 70% of other pollutants from transport, and the total cost of congestion in the EU is nearly 100 billion, which stands for 1% of the annual EU's GDP (European Commission 2017). Traffic congestion can be classified as recurrent, mostly due to a large number of commuters during peak hours, and non-recurrent caused by an unexpected event, such as traffic accidents, extensive weather conditions, or special events. The authors in (Chow et al. 2014) report that recurrent congestion traffic covers almost 85% of all congestion occurring on the urban road networks. On the other hand, different numbers are reported for the highway facilities (United States. Federal Highway Administration 2005), where the authors state that 50% of all traffic congestion is caused by non-recurring congestion and 40% is caused by recurring congestion. The crucial part of ITS supported systems for the decision-making processes are the detection and quantification of traffic state to initiate and implement improvement strategies. Moreover, traffic state estimation is a prerequisite to many other ITS applications, like travel time prediction (Servos et al. 2019), route computation (Pun-Cheng 2012), traffic flow prediction (Ouyang et al. 2020), etc.

This paper presents a method for the traffic state estimation on urban road segments that are based on the clustering of the Center Of Mass' (COMs) of the speed data represented in the Speed Transition Matrices (STMs). The proposed methodology includes three main steps: (i) data preprocessing, (ii) STM computation based on the speed data, and (iii) clustering-based traffic state estimation process. The main part of the preprocessing step is data filtering based on the seasonality that results in summer months and weekend data being excluded from the real-life large Global Navigation Satellite System (GNSS) dataset. The STMs were computed based on the speed data, and they represented the speed probability distribution of vehicles traveling between two road segments (transition). Next, the agglomerative clustering approach is conducted to cluster the traffic data in the form of the STMs. Clustering results in three classes of the traffic state. The results are validated using the cross-validation approach and the specific domain knowledge data, which were extracted from the Highway Capacity Manual (HCM). The validation resulted in the average accuracy of the classification for the cross-validation of 97%, and the domain knowledge data in 91%.

Contributions of this paper are as follows:

- Novel traffic data representation is proposed in the form of the STM which application is shown for the traffic state estimation, routing applications, and anomaly detection.
- The methodology for the traffic state estimation on a city-wide scale is proposed based on the STMs and computed COMs.
- The proposed methodology is applied and validated on the real dataset for the City of Zagreb, Croatia.

The rest of the paper is organized as follows. In Section 2 literature review is presented on recent developments related to GNSS data representation, data modeling techniques, and traffic state estimation approaches. Section 3 presents the methodology that was used to estimate the traffic state using the STMs and computed COMs with clustering, and validation methods. Section 4 describes the used real-life GNSS dataset and the results of traffic state estimation, clustering, and validation. In Section 5, the discussion is given regarding the presented method for the traffic state estimation. The advantages and disadvantages of using the proposed methods are given, and the possible application of the STMs for other traffic-related problems are presented. Section 6 presents the conclusion and future work suggestions.

2.3 Literature Review

2.3.1 GNSS Data for Traffic Representation

The crucial part of the ITS are data-driven services (Mandžuka et al. 2016; Škrinjar et al. 2020), which are supported by the advancements in technologies that enabled the lower cost of data collection systems. There are roughly two groups of data collection systems: (i) dedicated infrastructures which mostly consists of point detectors such as loop detectors, radar or lidar counters, and traffic cameras, and (ii) Floating Car Data (FCD) with devices mounted inside vehicle or carried by the driver, like GNSS devices or cellular data. The FCD from GNSS devices is often used to collect traffic data, because it provides a low cost, high accuracy, less delay, and wide coverage. One of the main FCD advantage is the ability of route construction and analysis.

The authors in (Herrera et al. 2010) used GNSS data to conduct a field experiment to validate collected traffic data. The result suggests more reliability when using the GNSS data than loop detector data. In (Herrera and Bayen 2010), the authors conducted an experiment with mobile phones onboard GNSS devices for traffic density and volume estimation. The results show that the proposed models successfully incorporate GNSS data to estimate the traffic parameters. In (Kong et al. 2013), the authors used GNSS probe data for the traffic state estimation based on the speed by incorporating the curve fitting method. In (Erdelić et al. 2019), the authors used streaming GNSS data to classify the travel modes based on the characteristic distribution of velocity and acceleration for different travel modes.

The GNSS datasets enable the use of dynamic routing applications. Dynamic routing can be defined as a process of changing the original (pre-computed) route based on the current traffic state on the road traffic network (Rakha and Tawfik 2009). The traffic network's dynamic nature is manifested in both temporal and spatial changes that can be captured by GNSS data. The authors in (Juhász 2017; Wahle et al. 2001) conclude that drivers with available real-time traffic information can significantly decrease travel time if compared to the drivers using offline navigation tools.

In this paper, FCD data that were based on the GNSS tracks were used. The data were preprocessed to compute the STMs that present the speed probability distributions when vehicles are moving between consecutive road traffic segments. Computed STMs were used to represent the traffic state on the observed road network segments. The STMs can be used in both offline and online route planning scenarios.

2.3.2 Data Modeling Techniques

When estimating traffic states from sparse GNSS datasets, most of the authors use aggregation-based methods to determine the observed traffic parameter value. Traffic data, like speed, volume, or density, are mostly aggregated in profiles that represent the change of observed parameter over the defined time period, i.e., one day. Data is usually aggregated in a narrow time interval (common values are 5-, 15-, 30-, 60-min.) as an average or median of all values recorded in the observed time interval. Because of the data aggregation, profiles could include large deviations in some time intervals that raises the question of the reliability of obtained results. One more challenge in data aggregation is related to the missing data, as it can extremely influence the average or median values.

Vector representation of traffic data in the form of a time series is one of the most common data modeling techniques (Erdelić, Ravlić, and Carić 2016). The change of traffic parameter under observation with dimensions $1 \times n$ is examined through a daily profile in defined n time intervals. The shortcomings of such approaches are reflected in the inability to represent spatial components of the observed parameter. In contrast, some authors, (Ma et al. 2017b; Nguyen et al. 2017), used matrix models in order to represent the traffic data. Matrix models can model more complex data, with the ability to represent more dimensional data. Then, both spatial and temporal information can be analyzed simultaneously. In most cases, matrix dimensions are represented with $m \times n$ where m represents the number of spatial segmentations (often road segment) and n number of time intervals. This kind of modeling can be used to extract spatial and temporal dependencies between the observed traffic parameters. For example, in the case of the commonly used data representation form of Origin-Destination (OD) matrices, one more dimension must be added to analyze the temporal component. OD matrices represent the number of vehicles traveling between defined points in the traffic network. While the provided information is useful for mobility pattern research, the patterns could indicate false information due to the predefined delivery routes if the data consists of delivery vehicles.

The novel data representation in the form of STM is proposed to overcome the mentioned limitations regarding the sparse GNSS data analysis. STM does not suffer from aggregation or missing data limitations, as data are not aggregated in such a way, and all the recorded speed data is shown in one matrix. The origin and destination vehicle movement are limited to consecutive links, enabling the usage of delivery vehicles in the traffic analysis process.

2.3.3 Traffic State Estimation Approaches

Many traffic state estimation approaches are developed to overcome the challenge of quantifying the congestion on the road networks (Afrin and Yodo 2020). Measures are related to the available traffic parameters, such as speed, travel time, congestion indices, delay, volume, and level of service. In this paper, the speed is chosen as a traffic parameter for traffic state estimation. The authors in (Rao and Rao 2012) report that speed is a good measure for traffic state estimation, because the congestion is a function of speed reduction, which is related to increase in time travel, vehicle operating cost, fuel consumption, and emissions. When using the GNSS data, the speed is relatively easy to compute and does not require complex processing if compared to extracting the traffic volume from the same dataset (Liu et al. 2016).

One of the main research goals with a topic that is related to traffic state estimation is to discern different states and provide a certain threshold for the classification of the traffic states. The well-known traffic study (Kerner 2002) presents the three-phase traffic theory. It includes classification to three states, namely "wide moving jam", "synchronized", and "free flow". The additional research was presented in (Kerner et al. 2004), where the authors presented a method for tracking synchronized flows and propagating moving jams. The authors in (Nguyen et al. 2019) classified the congestion based on the traffic images that represent speed on the observed road segment, resulting in five classes, namely "isolated", "low frequency", "high frequency", "homogeneous", and "mixed". In (Kim and Mahmassani 2015), the authors used a clustering technique to estimate the traffic state based on the trajectory data. The authors in (Thianniwet and Phosaard 2009) used artificial neural networks and decision tree algorithm to classify road traffic congestion levels based on the extracted speed traffic patterns. The research resulted in three classes, namely "jam", "heavy traffic", and "light traffic". In (Jin et al. 2001), the authors used neural network to classify traffic patterns with the aim of incident detection. The authors report three classes of traffic state described with speed, volume, and occupancy values. The authors in (Kan et al. 2019) estimate the congestion of the trajectory segments with three different intensities. Subsequently, congestion events are identified in the traffic network on each turning direction through multiple clustering approaches based on the speed, distance, and time of day. The authors in (Keler, Ding, and Krisp 2016) aim to visualize the traffic conditions on the urban road network. Traffic conditions are classified based on the average speed and grouped into five different classes.

In this paper, three classes of traffic state were used, namely "Free flow", "Stable flow", and "Congestion". This approach is used to characterize the congested traffic state and describe all the states that can occur on the traffic network. The COM estimation process is used to simplify the clustering process of traffic data that is represented by STMs. The three-dimensional STM representation is transformed into the two-dimensional representation using the x and y coordinates of COM. With this transformation, the classical distance metric, such as euclidean distance, is more interpretable than the classification of raw matrices, where distance is measured between every cell in two matrices. The proposed approach presents a simple but effective way of clustering the traffic data represented by the STMs. The advantage of the proposed method can be seen in the reduced computation complexity of the cluster analysis process and the increased interpretability of obtained clusters.

2.4 Methodology

Given the large FCD dataset, this paper aims to describe a traffic state estimation and classification method. Figure 2.1 presents the proposed methodology. In this Section, the main steps are briefly described: (i) STMs computation, (ii) traffic state estimation process, and (iii) clustering and validation methods.

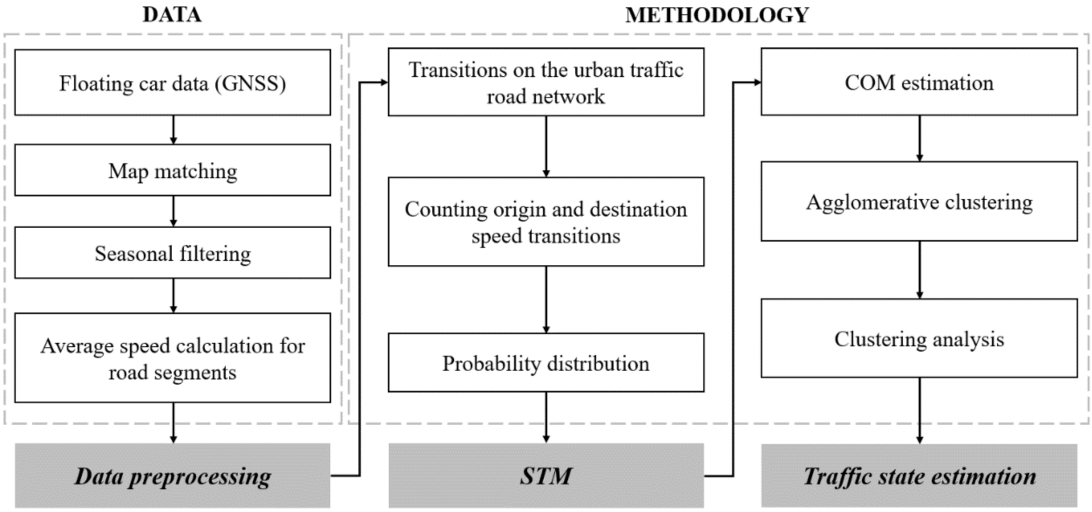


Figure 2.1 The methodology for the traffic state estimation and classification based on the Center Of Mass (COM) estimation of the speed data represented using Speed Transition Matrices (STMs).

2.4.1 Speed Transition Matrix

Most of the authors represent traffic data as a time series vector $v \in \mathbb{R}^{1\times n}$ (Erdelić et al. 2016) or a two-dimensional matrix $M \in \mathbb{R}^{m\times n}$ (Liu et al. 2016). Dimensions m and n refer to the numbers of the road network segments (the spatial component) and the number of time intervals (the temporal component) of the observed road network. The STM concept is proposed on the Markov chain theory, where the transition matrix shows the probability of transition from one state to another. The STM is used to represent the probability of speed change when traveling between two consecutive road network segments. In this paper, the road network is represented as a directed graph $G = (V, E)$, where V is a set of vertices representing intersections, and E is a set of edges representing road segments that connect two adjacent intersections. The transition is defined as a spatial change in vehicle trajectory when traveling from edge e_i to edge e_j in time interval t. As a traffic parameter under observation, the average speed is used. The average speed computed on e_i labeled as the origin speed s_o, and the average speed on the e_j segment is labeled as destination speed s_d. Two examples of the transition are visually represented in Figure 2.2 a) with red and blue colors. The transitions describe the vehicles traveling between edges c and f, and edges b and g. Subsequently, the STM matrix X is constructed as follows. First, all speed transitions from s_o to $s_d$between e_i and e_j are discretized and then counted within the particular time interval t. Each obtained value represents the count of transitions between $s_o$and s_d. The speed counts are further transformed into the speed transition probability distribution to obtain the probability for every transition. Values are put into the matrix X, which dimensions depend on the chosen resolution (sensitivity) of the speed change and the maximal observed speed. In this paper, 5 km/h is chosen as the speed discretization value and 100 km/h for the maximal possible speed, which resulted in matrix dimensions of 20×20. The specific maximal speed value is chosen, because experiments are conducted on the road segments with a speed limit between 50 and 80 km/h. Equation ((2.1) presents the STM, where every value p_{ij} represents the probability that the vehicle had origin speed s_o and destination speed s_d in the observed transition at time interval t.

$$X^{(ij)}(\Delta t) = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ p_{m1} & \cdots & \cdots & p_{mn} \end{pmatrix} \tag{2.1}$$

Figure 2.2 b) and c) show two examples of the STM. Example (b) represents normal traffic flow as the high-speed values are present on both origin and destination segments, and (c) represents congested traffic flow, as speed is speed much lower when compared to the free flow speed. It can be noted that the position of the captured pattern is important information for the traffic state estimation.

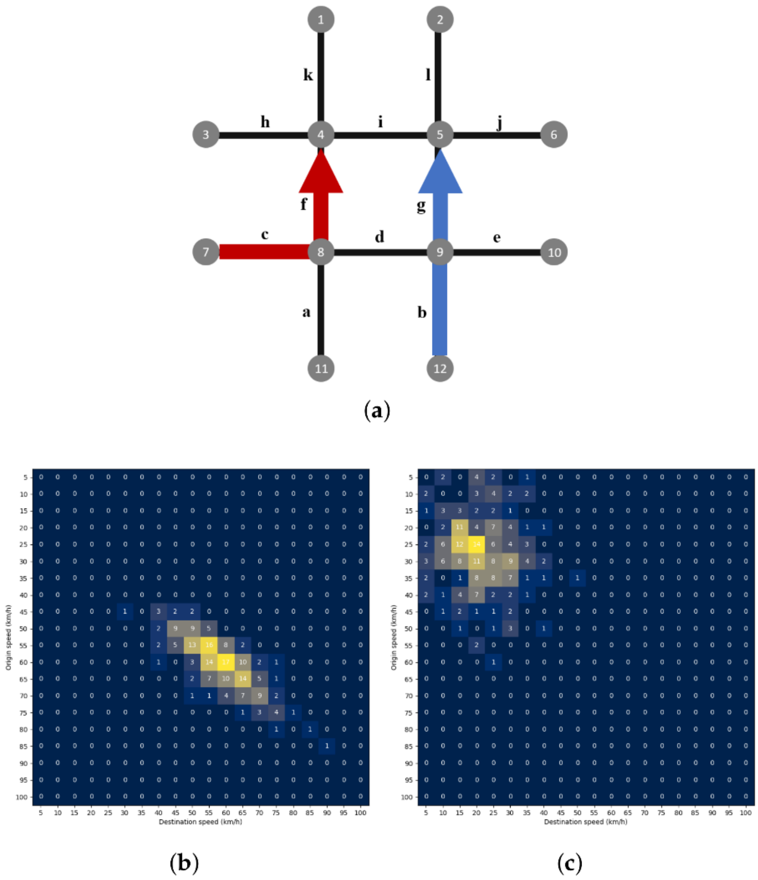


Figure 2.2 (a) Example of transitions on a simple road network, (b) STM representing the normal traffic flow, and (c) STM representing the congestion.

2.4.2 Center of Mass Estimation

This method presents a simple but effective feature extraction process. The feature, in this case, is the position of the traffic pattern extracted from the STM. The COM's position is the single most important information that is useful for the traffic state estimation problem when using the STMs, as the position can indicate different traffic conditions. If placed in the upper left corner, the position indicates that the average speed is very low if compared to the speed limit and traffic state can be declared as heavily congested. For the lower left corner, the position indicates very high speed on origin road segments while, at the same time, the speed is extremely low on the destination road segment. The same behavior can be noticed if COM is in the upper right corner, where speed values are low on the origin, and extremely high on the destination road segment. If the COM coordinates are positioned in the center of the matrix or in the lower right corner, it indicates that the speed values on both origin and destination road segments are relatively close or higher than the speed limit. This behavior can be interpreted as normal traffic behavior, as the speed value points to traffic flow without congestion. If COM's position is located in-between mentioned traffic states, the traffic state could be declared unstable.

STMs are transformed to extract the COM to simplify the classification process. As the result, the 20×20 STM is represented by COM coordinates, c_x, andc_y. Subsequently, all the points with coordinates c_x and c_y can be plotted in a two-dimensional space and clustered by the position in the coordinate system. Figure 2.3 represents the transformation from STMs and simplification to COMs plotted all in one coordinate system.

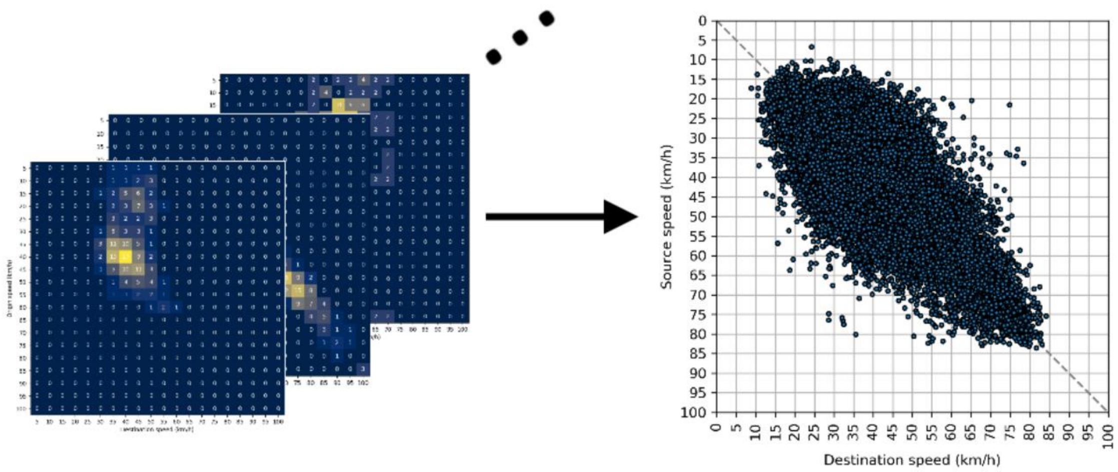


Figure 2.3 Representing the three dimensional STM data with the two dimensional COM coordinates.

For the COM estimation, a method that is based on the computation of the expected coordinate values is used (Jordaan 2005). First, marginal distributions for the x and y coordinates are computed while using:

$$p_x(x_j) = \sum_{i=1}^{m} p_{ij}\,; j = 1,2, \dots, n \tag{2.2}$$

$$p_y(y_i) = \sum_{j=1}^{n} p_{ij}\,; i = 1,2, \dots, m \tag{2.3}$$

where p_x is a marginal distribution of the x coordinates of the STM, and p_y is a marginal distribution of the y coordinates of the STM. Afterwards, x and y coordinates of the COM are computed as the expected values:

$$c_x = \sum_{j=1}^{n} p_x(x_j) \cdot j \tag{2.4}$$

$$c_y = \sum_{i=1}^{m} p_y(y_j) \cdot i \tag{2.5}$$

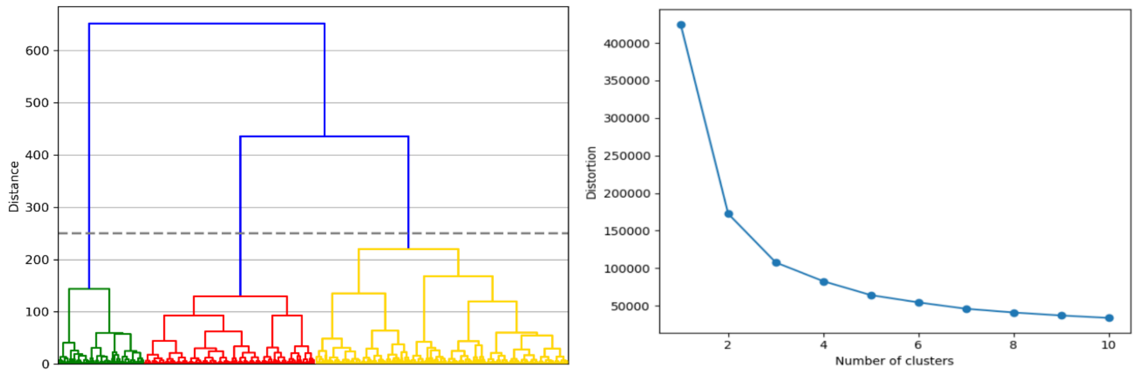
where c_x is a x coordinate of the COM, and c_y is a y coordinate of the COM.

2.4.3 Clustering

In this paper, clustering aims to find groups of traffic patterns that are represented as STMs used to represent the current traffic state on the observed road segment. Three classes of traffic state were used, namely “Free flow”, “Stable flow”, and “Congestion”. This approach is used to characterize not only the congested traffic state but to describe all states that can occur on the traffic network. The “Free flow” class describes the traffic conditions when a vehicle travels on an empty road or with speed close to the speed limit. The "Stable flow” class describes traffic conditions that most drivers feel as “normal”, when drivers experience speed reductions, but the traffic is flowing smoothly most of the time. The “Congestion” class indicates traffic conditions with a strong decrease in travel speed and increased travel times on the traffic road network.

2.4.3.1 *Agglomerative Clustering*

Hierarchical clustering is chosen for the clustering method. This approach constructs a hierarchical representation of a dataset, which presents an overview of the distribution of existing COMs extracted from STMs. This approach’s advantage is in providing the ability of reproducibility of resulting clusters and it provides more explanatory results (Nguyen et al. 2019). There are two types of hierarchical clustering: (i) agglomerative and (ii) divisive. The approaches differ by way of constructing binary tree representation. The agglomerative approach uses a top-down, and the divisive approach uses the bottom-up strategy. In this paper, the agglomerative approach is used. It initiates each pattern as a single cluster and measures the distance between patterns and intermediate clusters. Subsequently, in every iteration, it combines the two closest patterns into a new cluster. The process is repeated until only one cluster remains. Figure 2.4 a) shows the results of an agglomerative clustering presented by dendrogram plot.



(a) (b)

Figure 2.4 (a) Result of the agglomerative clustering approach, and (b) the elbow method results.

2.4.3.2 *Clustering Validation*

The elbow curve is presented for the observed data to confirm the number of clusters. In the cluster analysis process, the elbow curve is a heuristic used to determine the number of clusters in the dataset. Figure 2.4 b) presents the elbow curve with distortion plotted against the different number of clusters. The "elbow" can be detected in the part of the curve where the number of clusters is 3. At this point, the further increase of the number of clusters would not significantly contribute to the clustering quality. This value is used for the number of clusters for further experiments.

As the first validation technique of the classification process, cross-validation is adopted from (Kan et al. 2019; Nguyen et al. 2016). For the cross-validation process, the 1000 data instances (COMs) from every class are randomly selected and labeled based on the visual inspection. The selected dataset is then separated into the training and test sets with a ratio of 80% for training and 20% for testing. The labeled dataset is then used as a ground truth value and compared to the agglomerative clustering results.

The second validation process is related to comparing the resulting classes with the domain knowledge data. The well-known HCM values of the Level of Service (LoS) are used to represent the specific domain knowledge data for the traffic state estimation process. HCM defines six levels of service for road segments that are based on driving speed values, from A to F, with LoS A representing the best driving conditions and LoS F the worst. Label A represents the best traffic conditions, with vehicle speeds larger than 80% of the free-flow speed, while label F represents the most extreme congestion, where vehicle speeds are less than 30% of the free-flow speed (Anon 2010). LoS quantifies the increase in travel time due to the conditions on the road segments and it is also a measure of driver discomfort, and fuel consumption. In this paper, LoS values are used to validate the traffic state estimation process. Firstly, the LoS values are merged in three classes in following way: (i) free-flow traffic conditions represented by the LoS labeled as A and B, (ii) traffic conditions represented by the LoS labeled with C and D are labeled as stable, and (iii) congested traffic conditions that are represented with LoS are labeled with E and F. Then, the same test dataset for the cross-validation is labeled with three classes. The labeled dataset is then used as a ground truth value and compared to the agglomerative clustering results.

2.5 Results

2.5.1 Data

The large real-life FCD acquired from vehicles equipped with tracking devices is used. Each record contains a timestamp, geographical longitude and latitude, speed, and heading. Due to the storage limitation, most of the data are sampled in the following way: 100 m for vehicles in driving mode and every 5 min. for turned off vehicles. Raw data are map matched to the road segments in a digital map based on the measured latitude, longitude, and heading. Data that could not be matched to appropriate road-segment due to the errors caused by tunnels, high building concentration, or other causes were filtered out. GNSS data for Croatia's road network were recorded for five years between August 2009 and October 2014 by approximately 4200 tracked vehicles. The tracked vehicle fleet is versatile and mostly consists of delivery vehicles (vans, caddies, small trucks) and taxi cars. The historically tracked data, which consist of 6,55 billion records, was provided by Mireo Inc. as a part of the SORDITO project (Carić et al. 2016). In this paper, we analyze the data and estimate traffic state using the proposed method for some of the most relevant road segments and intersections in the City of Zagreb, the capital, and the largest city in Croatia.

The seasonality of the traffic flow is considered to lower the deviation. Summer months, July, and August are excluded from the experiment. They significantly influence the results on the road network of Zagreb, due to the different, and lower traffic flows that are caused by vacations (Żochowska and Karoń 2016). Data is further divided into two groups: working days and weekend days. The working day data, Monday to Friday, show different traffic conditions when compared to the weekend data (Saturday and Sunday), mostly due to the daily commuters and, therefore, the weekend data are also excluded.

Results of a traffic state estimation are shown for the eight time intervals throughout a day. Time intervals are defined by (Carić et al. 2016; Carić and Fosin 2020) for the congestion estimation problem based on the same real-life FCD dataset that refers to speed data for the City of Zagreb. Intervals are defined, as follows: (i) 05:30–06:45 as morning interval with very small traffic volume, (ii) 06:45–07:25 as interval before the morning rush hour, (iii) 07:25–08:20 as morning rush hour, (iv) 08:20–15:30 as interval between morning and evening rush hour, (v) 15:30–17:05 as evening rush hour, (vi) 17:05–19:00 as interval after evening rush hour, (vii) 19:00–22:00 as interval that represent evening traffic conditions, and (viii) 22:00–05:30 as the night interval.

Table 2.1 represents the results for the traffic state estimation grouped into three classes. The results are shown while using the ratios between the number of classified transitions and the total number of transitions in the observed time interval. The rows show the distribution of classified transitions in the observed time interval. Rush hours are highlighted and as expected, have the largest values of the transitions labeled as congested. The time interval between rush hours shows the largest value of the ratio for the congested transitions. This indicates that congestion that started in the morning rush hour is prolonged to the next time interval. This also could indicate that congestion starts at the interval between rush hours and is prolonged to the evening rush hour.

Table 2.1 Results of traffic state estimation represented using three classes.

	Free Flow	**Stable Flow**	**Congestion**
22:00–05:30	28.31%	54.15%	17.54%
05:30–06:45	27.38%	53.55%	19.08%
06:45–07:25	19.08%	42.83%	38.10%
07:25–08:20	13.97%	32.53%	53.50%
08:20–15:30	12.09%	33.86%	54.05%
15:30–17:05	13.05%	35.46%	51.04%
17:05–19:00	15.53%	40.59%	43.88%
19:00–22:00	19.77%	46.33%	33.90%

This kind of behavior could indicate inefficient traffic regulations on observed transitions. The ratio of the congested roads in the time interval of 17:05–19:00, shows a large portion of the transitions classified as congested, although it is not the rush hour interval. This behavior can be addressed to the city attractions placed in the city center and people visiting such locations in their free time. This fact can be confirmed by the classes' spatial distribution that is presented in Figure 2.5 g).

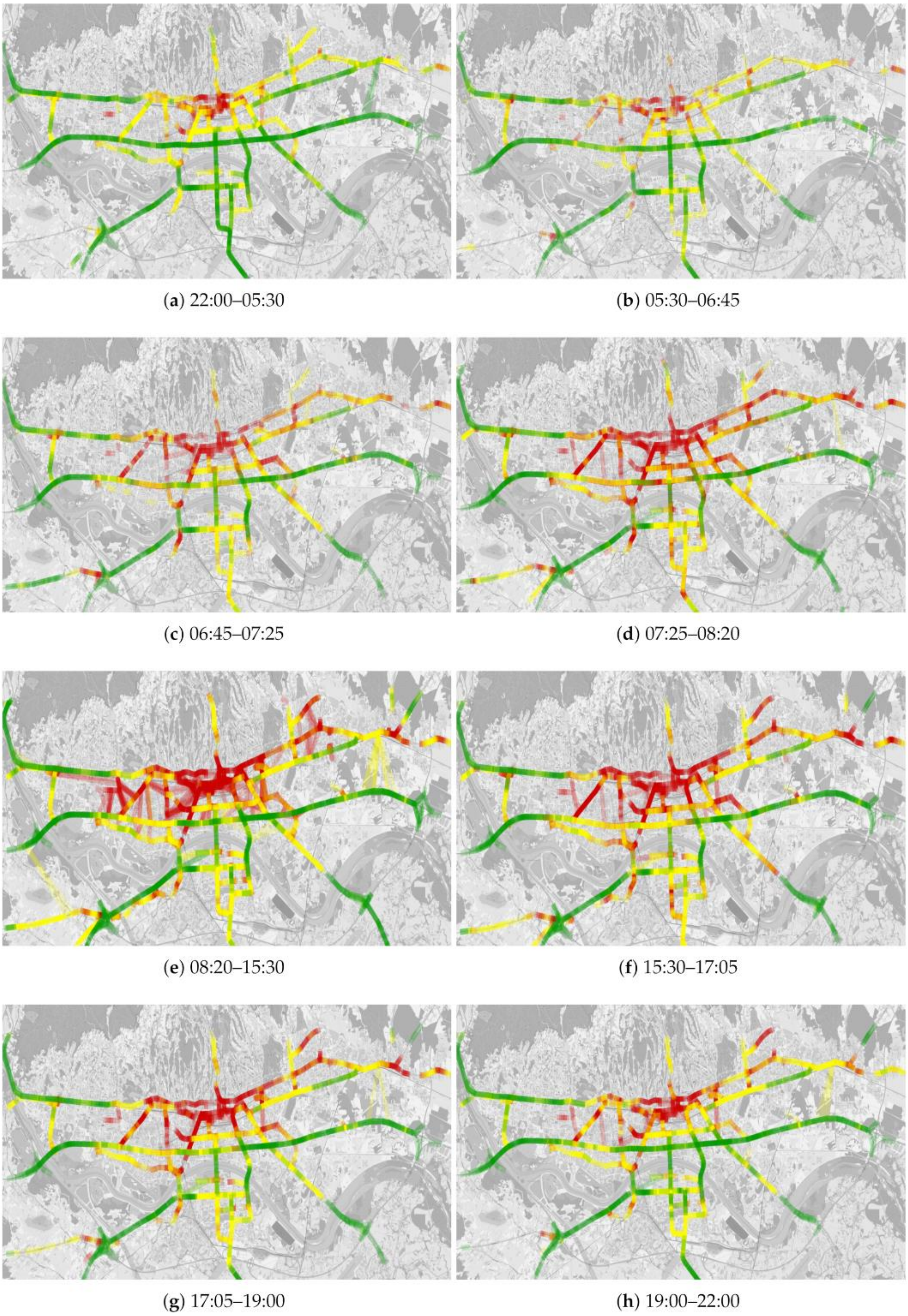


Figure 2.5 Results of traffic state estimation visualized on the map of the City of Zagreb.

Figure 2.5 shows the spatial distribution of the classified traffic patterns for every observed time interval. The colors used for the visualizations are: (i) red— "Congestion", (ii) yellow— "Stable flow", and (iii) green— "Free flow". It can be noticed that the congestion level is the highest in the city center and the west part of the city. Some congestion patterns can be extracted by analyzing all time intervals. For example, the southern part of the city represents the business area. The congestion occurs only in the morning and evening rush hours due to the daily commuters. Visualization of the congestion patterns can be used for more in-depth and more granular analysis of the traffic state. For the case study, Zagreb's three most important bridges across Sava River are chosen, which divide the southern and north part of the city. The bridge Jadranski most is detected as the most congested bridge. Figure 2.6 a) shows the enlarged image of the traffic congestion estimation results for the Jadranski most in the rush hours. The results show the most congested approaches to the bridge and the roundabout at the southern approach. The figure shows that STMs can estimate traffic state at micro-locations and consider the direction of the traffic flow. It can be observed that the direction from south to north is more congested than the opposite one. The same behavior can be noticed in the afternoon rush hour.

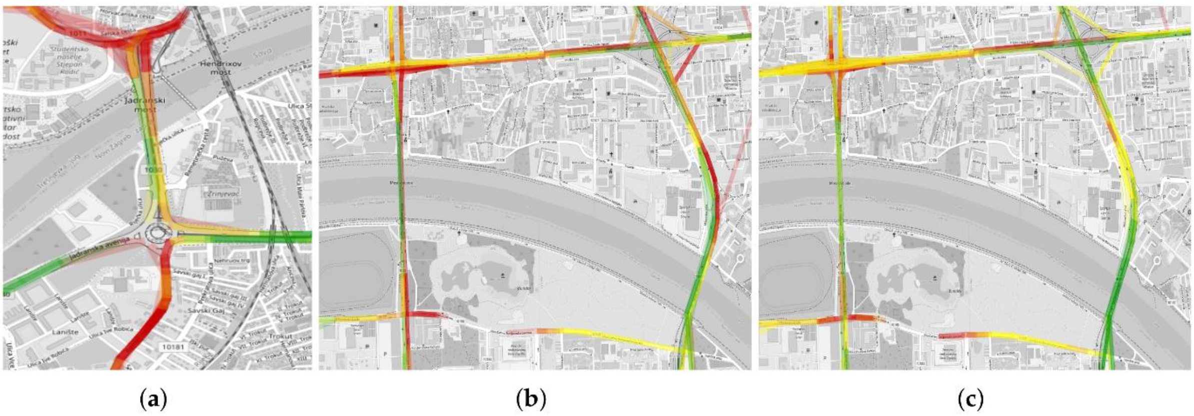


Figure 2.6 (a) Traffic state on the Jadranski most in both peak hours, (b) Traffic state on the bridges Most slobode and Most mladosti in the morning peak hours, and (c) Traffic state on bridges Most slobode and Most mladosti in the afternoon peak hours.

On the other hand, different behavior can be observed on the other two bridges, namely Most slobode (left) and Most mladosti (right), as shown in Figure 2.6 b) and c). The traffic state estimation results show different behavior in the morning and afternoon peak hours. In the morning peak hours, the traffic congestion is increased in a direction towards the city (south to north), which indicates the increase in traffic demand due to the commuters, while the other direction (north to south) represents the normal or free-flow conditions. In the afternoon peak hours, both bridges' flow indicates the normal or free-flow conditions, and congestion occurs at the intersections due to inadequate traffic lights signalization.

2.5.2 Validation Results

The results for both validation processes are reported while using the confusion matrices and the classification report that shows the total accuracy of the model, precision, recall, and F1 scores for every class. The confusion matrix reports the performance of the classification in a visual manner. Each row of the matrix represents the ground truth values, while the columns present the predicted class labels. The values in the matrix represent the accuracy of the prediction computed as a number of data instances that are correctly classified divided by all of the data provided for the considered class.

In classification problems with more than two classes, the precision is computed as the sum of the true positive values, divided by the sum of true positive and false positive values computed across all classes. The F1 scores are computed as the harmonic mean of precision and recall. Accuracy is the measure for the accuracy of the model computed across all classes by averaging the total true positive, false negative, and false positive values.

Table 2.2 presents the classification report for the cross-validation method. The validation achieved the average prediction accuracy of 97%. The recall of the class labeled as “Free flow” with a value 84% shows that, even if the precision shows the perfect score, there are some values that are not classified correctly.

Table 2.2 Results of the cross validation.

	Precision	**Recall**	**F1 Score**
Free flow	1.00	0.84	0.91
Stable flow	0.93	1.00	0.97
Congestion	1.00	1.00	1.00
Accuracy			0.97

If Table 2.3 (confusion matrix) is observed, 15.9% of the values of the class “Free flow” are labeled as “Stable flow”. The results of the cross-validation method indicate that the classes labeled with “Stable flow” and “Congestion” are well separated and they can be classified with high accuracy, while classes that are labeled as “Free flow” and “Stable flow” to some extents are harder to separate.

Table 2.3 Confusion matrix of the cross validation.

Predicted / **Known**	**Free Flow**	**Stable Flow**	**Congestion**
Free flow	0.841	0.159	0
Stable flow	0	1.000	0
Congestion	0	0.003	0.997

Table 2.4 presents the classification report for the validation that is based on the domain knowledge extracted from the HCM. The validation achieved the average prediction accuracy of 91%. The lower prediction accuracy can be accounted for by the strict boundaries of defined LoS values. The lowest precision value can be noticed in the class labeled as unstable operations.

Table 2.4 Results of validation using domain knowledge data.

	Precision	**Recall**	**F1 Score**
Free flow	0.99	0.95	0.97
Stable flow	0.83	1.00	0.90
Congestion	1.00	0.81	0.89
Accuracy			0.91

If the corresponding confusion matrix is observed in Table 2.5, 19.1% of unstable traffic instances are predicted as congested traffic and 4.6% are predicted as normal traffic.

Table 2.5 Confusion matrix of validation using domain knowledge data.

Predicted / Known	Free Flow	Stable Flow	Congestion
Free flow	0.954	0.046	0
Stable flow	0.004	0.996	0
Congestion	0	0.191	0.809

2.6 Discussion

In this paper, novel traffic data representation in the form of the STM is proposed. This section describes the potential application of the STM in traffic-related research and applications and highlights some drawbacks that need to be addressed in further research. This section also emphasizes expected impact for academia, such as using STMs for visualization, quantifying, and classifying traffic state, capturing the changes in the speeds on a road network, identifying the anomalous behavior, and using COM's movement for presenting the probability of traffic state change or the capturing complex traffic patterns that can be used to identify potentially dangerous traffic situations.

Alongside the presented application of the method, it has some drawbacks that need to be addressed. Traffic state value is only based on the speed values. This could be a problem on short road segments bounded by unsynchronized traffic lights because vehicles would have very low speeds due to the traffic light's signal plan. Regarding this property, the traffic state could be wrongly estimated as very high. Secondly, sparse GNSS data is used, which entails wide time intervals for the experiment. Commonly used, shorter time intervals like 5-, 15-, 30-, or 60-min could give better insight on the traffic state and, thus, on the traffic state on the observed road segments. In this paper, the dataset used for the experiment is data that only includes working days to only capture the most extreme congestion conditions in the urban road network. A possible improvement would be to include the weekend data to analyze the differences between traffic in working and weekend traffic flow fluctuations.

2.6.1 Traffic State Estimation

In this paper, the use of the STM is proposed to estimate and classify the traffic state on the urban road network. It is shown that the STM can be useful for the visualization, quantifying, and classifying traffic state. The STM is a possible traffic data modeling approach for traffic state prediction. The proposed data model can be used as a set of images for training some machine learning model to predict the future state of traffic. The full potential of the STMs can be utilized in (near) real-time analysis when the position of the COM for every STM is changing over the observed time period. The position itself and the movement of the COM (positions in the past observed intervals) could provide usable and actionable information for traffic management systems. The COM's movement indicates the change in traffic state and can present the probability of traffic state change, which is an important factor in the traffic state prediction problem.

2.6.2 Routing Applications

The routing applications benefit the most from the traffic state estimation and prediction. Every route planner must include current, and possibly future, traffic state information to enable fast and secure delivery. Traffic state estimation based on the STM can provide useful information regarding congestion, and therefore routing through the less congested roads. The framework for solving the well-known routing problem Time-Dependent Vehicle Routing Problem, is presented in (Carić and Fosin 2020). The authors used speed profiles to extract the congestion zones and quantified the congestion by computing the slowdown coefficients using the travel times. The STM can be used in both steps. The congestion zones can be identified based on the position of the COM, while the same point represents the slowdown probability on the observed road segments.

2.6.3 Anomaly Detection

Anomaly detection is a crucial part of ITS, especially in the incident management domain. The detection of recurrent anomalies like heavy congestion could improve reaction time and give some actionable information for the traffic management authorities. On the other side, fast detection of non-recurrent anomalies, like traffic incidents, could even save a human life. STM presents an opportunity to capture the changes in the speeds on the observed road network and identify potential anomalous behavior. In contrast to the speed profiles, the STM provides a two-dimensional distribution of speed on consecutive road segments in an observed time interval. It enables the capture of more complex traffic patterns that can be used to identify potentially dangerous traffic situations. As one of the applications for anomaly detection, traffic bottleneck detection and propagation can be represented while using STMs. The STM presents a tool to capture, visualize, and analyze the bottlenecks' impacts on the road network. The bottlenecks result in traffic congestion on one part of the network caused by the traffic accident, badly timed traffic lights, or slow vehicles that disrupt the traffic flow. The STM can capture such scenarios. The COM's position in the upper right corner, or in the lower-left corner, could indicate a serious accident, as it shows very high-speed values on one road segment if compared to its consecutive one with very low-speed values.

2.7 Conclusions and Further Research

This paper presents a novel traffic data representation of the GNSS dataset by using the STMs. The methodology is presented for the traffic state estimation and classification on a citywide scale. The COM for every matrix was extracted to classify the STMs. This approach resulted in simplification of the classification process and higher interpretability of the resulting classes. The results show that STMs can be used to estimate the traffic state on a citywide scale and on micro-locations. The results are validated using the cross-validation method, and specific domain knowledge, which resulted in an accuracy of 97% and 91%, respectively.

As presented in the discussion section, the STM is a traffic data representation model that shows multiple possible implementation possibilities in different traffic and transport-related research and applications. Some of the applications are: (i) real-time traffic state estimation, (ii) routing applications, and (iii) anomaly detection in traffic data by identifying unusual traffic patterns that are captured by the STM.

There are multiple possible further research directions for the academic community. The first one could include training a deep learning model that is based on the Convolutional Neural Network (CNN) as a traffic state classifier. The STM is a data model that is formed as a traffic image that can be used as input data for training the CNN. The second one would include a tensor-based analysis. Traffic tensor could be created as multiple STMs placed in the tensor-based on the time interval, in which the STM is collected. Subsequently, a tensor-based analysis could give more spatiotemporal insight into traffic conditions.

Chapter 3

Spatiotemporal Road Traffic Anomaly Detection: A Tensor-Based Approach

This chapter has been published as: Tišljarić L, Fernandes S, Carić T, Gama J. Spatiotemporal Road Traffic Anomaly Detection: A Tensor-Based Approach. Applied Sciences. 2021; 11(24):12017. https://doi.org/10.3390/app112412017

Author Contributions: Conceptualization, L.T. and T.C.; methodology, L.T., S.F. and J.G.; software, L.T.; validation, L.T. and T.C.; formal analysis, L.T.; investigation, L.T.; resources, T.C.; data curation, L.T. and T.C.; writing—original draft preparation, L.T. and S.F.; writing—review and editing, L.T., S.F., T.C. and J.G.; visualization, L.T.; supervision, T.C. and J.G.; project administration, T.C.; funding acquisition, T.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by European Regional Development Fund under the grant KK.01.1.1.01.0009 (DATACROSS). Sofia Fernandes acknowledges the support of FCT (Fundação para a Ciência e a Tecnologia) via the Ph.D. scholarship PD/BD/114189/2016.

3.1 Abstract

The increased development of urban areas results in a larger number of vehicles on the road network, leading to traffic congestion, which often leads to potentially dangerous situations that can be described as anomalies. The tensor-based methods emerged only recently in applications related to traffic anomaly detection. They outperform other models regarding simultaneously capturing spatial and temporal components, which are of immense importance in traffic dataset analysis. This paper presents a tensor-based method for extracting the spatiotemporal road traffic patterns represented with the speed transition matrices, with the goal of anomaly detection. A novel anomaly detection approach is presented, which relies on computing the center of mass of the observed traffic patterns. The method was evaluated on a large road traffic dataset and was able to detect the most anomalous parts of the urban road network. By analyzing spatial and temporal components of the most anomalous traffic patterns, sources of anomalies can be identified. Results were validated using the extracted domain knowledge from the Highway Capacity Manual. The anomaly detection model achieved a precision score of 92.88%. Therefore, this method finds its usage for safety experts in detecting potentially dangerous road segments, urban traffic planners, and routing applications.

Keywords: anomaly detection; tensor-based approach; traffic data; speed transition matrix; Intelligent Transport Systems

3.2 Introduction

The increased development of urban areas results in a larger number of vehicles on the road network, leading to traffic congestion, especially in rush hours. Intelligent Transport System (ITS) solutions present applications that can be useful in detecting and dealing with problems that are related to congestion like increased pollution (Li et al. 2021). In this context, anomaly detection represents an attractive research topic in the ITS field because it is one of the crucial parts in detecting dangerous and potentially life threatening situations on the road traffic network. Anomaly detection, in general terms, is a process that aims to find unexpected or significantly different behaviors of some data instances in the observed dataset. Its importance, combined with the analysis of the anomalous events, lies in potentially useful, actionable information for road traffic information providers and authorities to identify severe traffic accidents, traffic congestion, or a violation of the regulations.

This paper presents a tensor-based method for the extraction of the spatiotemporal road traffic patterns, with the aim of detecting anomalies on the urban road network. To distinguish between the recurrent congestion and anomalous events, this method is focused on two types of anomalies: The first one is sudden braking in transition which can be described as a bottleneck start, and the second type is intense acceleration in transition where vehicles are achieving unexpectedly high speeds when leaving the congested area.

The proposed method differs from other proposed methods in this research field because it presents a novel traffic anomaly paradigm based on Center of Mass (CoM) computation of the observed traffic pattern represented by the Speed Transition Matrix (STM). Compared to the traffic flow-based approaches, the main advantage of such an approach is its property that congestion cannot be wrongly detected as an anomaly.

In the context of mentioned disadvantages, the contributions of this paper are as follows:

- proposed method for the spatiotemporal road traffic patterns extraction which includes STM computation,
- the usage of the tensor composed of STMs to model the traffic patterns to address the spatiotemporal nature of the traffic data,
- proposed anomaly detection paradigm for the road networks based on the center of mass computation which addresses the problem of averaging many speed records into one value,

- results of the anomaly detection are evaluated on the urban road network segments in a medium-sized European city.

The proposed method consists of the three main steps: (i) Data preprocessing, (ii) grid-based map segmentation with STMs computation, and (iii) anomaly detection. The anomaly detection results are validated using the domain knowledge, extracted from the Highway capacity Manual (HCM) level of service values, with the achieved precision score of 92.88%.

In this context, we revise and extend our previous work (L. Tišljarić et al. 2020) by (i) more detailed problem and methodology description, (ii) introducing the STMs computing the harmonic vehicle speed, defined as relative values to be comparable with any road segment, (iii) improved tensor construction, introducing the grid-based map segmentation of the city area, and (iv) novel paradigm for the traffic anomaly definition on the road networks based on the computation of the STM.

This paper is organized as follows. Section 2 presents related work on road traffic anomaly detection methods, emerging tensor-based traffic data modeling techniques, and general tensor-based models for anomaly detection. In Section 3, the background, definitions, and preliminary concepts are presented. Section 4 presents the proposed methodology used for anomaly detection. Section 5 presents the method's results, including data processing, validation, analysis of the anomalous spatiotemporal patterns, and comparison to other approaches. Finally, Section 6 concludes the paper with a summary and future work directions.

3.3 Related Work

3.3.1 Traffic Data Modeling

Most traffic data like speed, density, or traffic flow profiles were represented by vectors, which consist of time series data (Erdelić et al. 2016). Each value in the vector represents the observed traffic parameter, which is averaged within a defined time interval. The limitations of such an approach are reflected in the impossibility of representing spatial components like spatial correlations between consecutive road segments. On the other hand, matrix-based models are used to model more complex traffic data and are often represented as traffic images (Ma et al. 2017b; Nguyen et al. 2017). Authors in (Zhang et al. 2022) generated traffic images as an input to proposed spatiotemporal generative adversarial network with a goal to represent urban mobility dynamics. In (Zhang et al. 2020), authors modeled traffic data using matrix that represent counts of origin and destination trips of a car-hailing service. Indexes of the matrix cells are often labeled as $m \times n$.

where m represents the road traffic segments and n time intervals. These models are used for spatiotemporal dependencies extraction between the observed traffic parameters, but only if the matrix is constructed to represent both spatial and temporal components. Such cases can be observed when using common mobility data representation, Origin-Destination (O-D) matrices. To extract temporal components of the O-D matrices, one more dimension must be introduced. One of the most used matrix decomposition methods, Principal Component Analysis (PCA), is used here. The PCA method is suitable for data interpretation with a smaller number of components and detects anomalies. On the other hand, the authors in (Wang et al. 2012) report that the PCA was not a suitable method when analyzing the traffic data because of large deviations in data due to many outliers in sensor readings. As the PCA relaxes three-dimensional data to the bi-dimensional form, authors in (Fanaee-T and Gama 2016) claim that it cannot be used for spatiotemporal patterns extraction.

Commonly, researchers extract speed profiles from different large traffic datasets. Values in speed profiles are extracted mainly by aggregating a large amount of Global Navigation Satellite System (GNSS) data recorded in defined time intervals into a single value. This process could result in significant deviations. Similarly, if O-D matrices represent data, there could be a large number of missing values in some data intervals. In most cases, large traffic data includes many delivery vehicles that significantly influence O-D matrices due to predefined delivery routes. The proposed STM traffic data representation model can avoid this behavior.

As a traffic data modeling technique, tensor-based models emerged only recently. The main advantage is that those models do not suffer from mentioned limitations regarding spatiotemporal data representation because of their property to model multi-dimensional data. The proposed method in this paper incorporates a tensor-based approach that is constructed using STMs. The model does not suffer large deviations as data is not aggregated from narrow time intervals. Secondly, the method can be used regardless of delivery vehicles because speed is the main observed traffic parameter.

3.3.2 Tensor-Based Anomaly Detection Approaches

Tensor-based approaches can be divided into three classes: (i) Supervised, (ii) semi-supervised, and (iii) unsupervised approaches. Supervised approaches are based on prediction (Xu et al. 2018), classification (Rendle 2012), and the dimensionality reduction (Prada et al. 2012). Semi-supervised approaches use normal data for a tensor construction, and for the baseline, decomposition results are used. The anomaly is estimated by observing the examples that do not pass the null hypothesis test (Tian et al. 2009) or fails to align with the baseline by comparing the eigenvectors and eigenvalues of the factor matrices (Fanaee-T and Gama 2015). Most of the unsupervised approaches rely on the manual anomaly detection performed by the field expert after the decomposition (Gauvin, Panisson, and Cattuto 2014).

Tensor decomposition methods fins their usages in traffic-related research especially in modeling of time-evolving traffic networks modeling (Fernandes et al. 2021; Wang et al. 2014), traffic data anomaly detection (Wang et al. 2019), road segments travel time estimation (Tang, Chen, and Liu 2018), correlation analysis of spatiotemporal traffic data (Tan et al. 2013), traffic parameters prediction (Pan et al. 2018; Tan et al. 2016), and missing data imputation (Chen et al. 2019).

This paper presents the tensor-based method for extracting the road traffic patterns on a city-wide scale represented by the graph-based map segmentation. Using the method for detecting the anomalous road segments expands the efforts to use tensor-based methods in road traffic-related studies. The unsupervised method is proposed, which is validated using the expert's knowledge extracted from the HCM.

3.3.3 Road Traffic Anomaly Detection Approaches

Many anomaly detection methods are developed for specific application domains, while there are some more generic methods. Review papers on anomaly detection focus only on some outlier detection categories like statistical or pattern mining methods. Most of the review papers present anomaly detection methods not explicitly designed for some area (Chandola 2009). Schubert et al. (Schubert et al. 2014) presented the review on local outlier detection with an application on spatial data, video, and network outlier detection methods. In (Feng and Zhu 2016; Zheng 2015) the authors presented an overview of the methods related to the trajectory of data mining techniques. Methods for various tasks related to the mining of trajectory data are presented, like trajectory pattern mining, anomaly detection, movement behavioral analysis, and trajectory classification. Gupta et al. (Gupta et al. 2014) presented the review paper on the detection of temporal anomalies. It gives an overview of various data types like time series data, data streams, distributed data, spatiotemporal data, and network data. Methods for anomaly detection are presented for each data type. The most recent survey paper (Djenouri et al. 2019) gives a comprehensive review of traffic anomaly detection methods in an urban area context. It divides the anomaly detection methods into two main categories: Trajectory and traffic-related anomalies.

There are three general approaches in anomaly detection: (i) Model-based, (ii) proximity-based and (iii) density-based methods (Tan, Steinbach, and Kumar 2006). Model-based methods include statistical models based on the assumption that normal observation has a much higher probability of occurrence in the model than the outlier occurrence. Capturing data is fitted to the statistical model, and a statistical interference test is applied to determine if the data behaves according to that model or not (Guo, Huang, and Williams 2015). Proximity-based approaches are distance-based anomaly detection methods. Anomalous observations are those values that are the most distant from all the other values (Pan et al. 2013). Density-based methods estimate the density of observations, and the anomaly is detected as the observation with low density when compared to its local neighbors (Chen, Wang, and van Zuylen 2010).

This paper presents the novel paradigm for the anomaly detection of the traffic networks using the STM. The proposed measure explains an anomalous traffic state as unexpected traffic flow behavior and avoids detecting recurrent congestion as an anomaly as the anomaly is defined as sudden breaks and intense accelerations events. Many road traffic anomaly detection methods are based on the detection of the large deviations within the traffic parameter observed in a defined time period. When analyzing the hour-by-hour data, recurrent traffic congestion could be wrongly detected as an anomaly because it represents the peak traffic load in rush hours that do not occur simultaneously with the same intensity on the whole city-wide area. For example, many anomaly detection methods based on the computation of the traffic volume cannot detect the anomaly in some time interval if the daily average traffic volume is not changed. Based on the STM, the proposed approach does not suffer from false anomaly detection, and it is adaptable for near real-time and real-time anomaly detection applications.

3.4 Background

3.4.1 Road Network Elements and Anomaly Definitions

Definition 1. Road network: *A road network is represented as a directed graph* $G = (V, E)$ *where* V *is a set of vertices representing the points of connection between two edges and* E *is a* *set of edges of the graph representing the road segments. Every edge $e_i \in E$ from a graph G represents a road network segment with the starting vertex v_i and the ending vertex v_j.*

Definition 2. Transition: *A transition is a movement of one vehicle between two consecutive road network segments e_i and e_{i+1}. Where origin edge of transition is e_i and destination edge is e_{i+1}.*

Definition 3. Speed transition: *A speed transition is a change in obtained speed when a vehicle is traveling through one transition. Then, speed on the origin edge e_i is named origin speed s_O and speed on the destination edge e_{i+1} is named destination speed s_D. Both speeds are computed as harmonic mean speeds of all obtained speed values on the origin and destination edges.*

Definition 4. Traffic anomaly: *This method is focused on two types of road traffic anomalies: The first one is sudden braking in transition which can be described as a bottleneck start, and the second type is intense acceleration in transition where vehicles are achieving unexpectedly high speeds when leaving the congested area. The anomaly is defined as a distance-based approach by computing the distance between the CoM and the closest point at the diagonal of the STM, d_{CoM}. The main goal is to find the anomaly distance d_A, which is used as a threshold value for the anomaly detection. Then, if distance from the observed CoM from the diagonal d_{CoM} is larger or equal than d_A anomaly is detected. In Figure 3.1 two types of anomalies are represented: (a) Sudden breaks and (b) intense accelerations.*

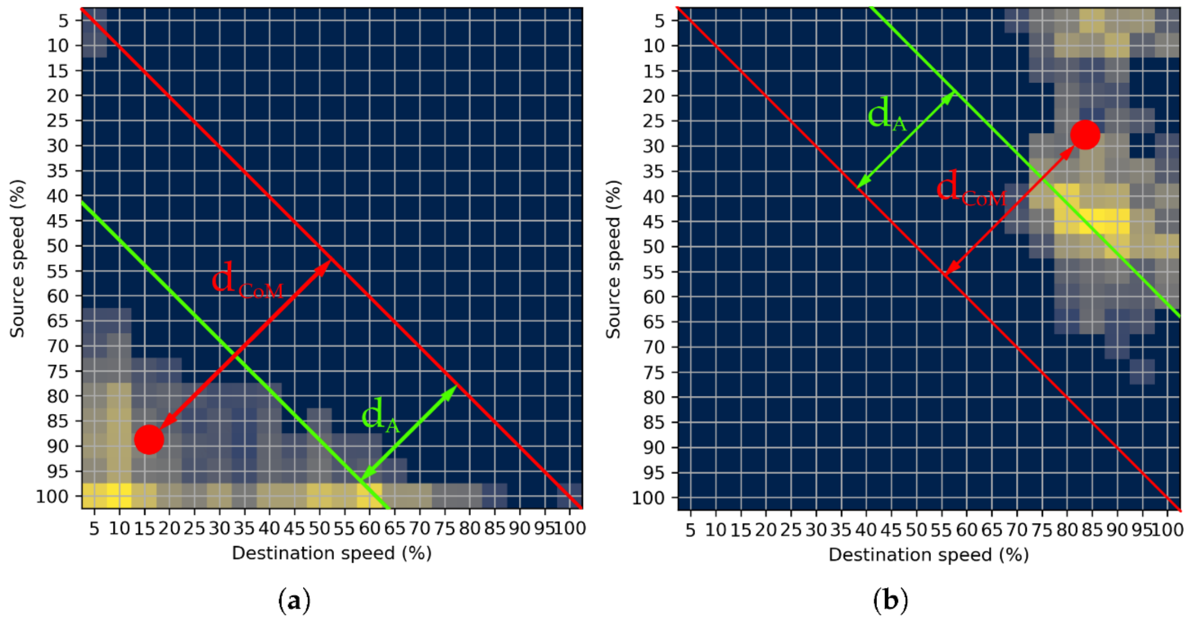


Figure 3.1 Example of two possible anomaly types: (a) Sudden breaks and (b) intense accelerations.

3.4.2 Speed Transition Matrix

The STM is a novel traffic data representation and modeling technique that captures the vehicle's speed at the movement between two consecutive road segments called transition (Tišljaric et al. 2021). It is used to represent the speed probability change, and therefore it represents the speed probability distribution at one transition in one time interval. The transition is defined as a spatial change in vehicle trajectory when traveling from edge e_i to edge e_{i+1} in time interval Δt. As a traffic parameter under observation, the relative harmonic speed is used. The speed is relative to the speed limit on the observed edge. Two examples of the transition are visually represented in Figure 3.2 with red and green colors. Transitions describe the vehicles that are traveling between edges $h-f$, and $l-g$ with corresponding STMs. The STMs represent a very different traffic pattern: (i) On the left-hand side, the traffic congestion with very low origin and destination speeds, and (ii) on the right-hand side, stable traffic flow with origin and destination speeds around 60% of the speed limit. The red circles show the CoM for represented traffic patterns. It can be observed that the position of the CoM is one of the most important parameters when estimating the traffic state, and the next chapters will debate how to use the position of the CoM for anomaly detection.

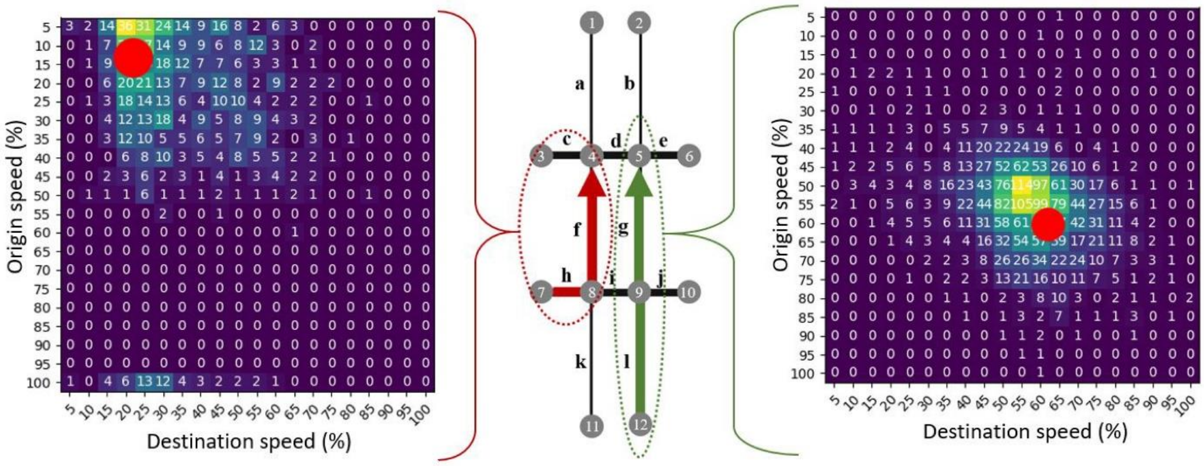


Figure 3.2 Transitions (center), congested STM (left), and normal traffic STM (right) examples on a simple road network.

In this paper, 5% is chosen as the discretization period, and 100% is the maximal possible speed, which resulted in matrix dimensions of 20×20. The STM can be represented using equation (2.1).

3.4.3 Tensors

Tensor τ is defined as multi-dimensional array $\tau \in \mathbb{R}^{N_1 \times N_2 \cdots \times N_M}$, where M represents the order of the tensor (number of dimensions). A vector is then represented with the first-order tensor, matrix with the second-order tensor, and three or more order tensors are called higher-order tensors (Kolda and Bader 2009). For the analysis of the road traffic spatiotemporal data, most authors use a third-order tensor composed using origin×destination×time and profile×roadsegments×time where profile represents the speed or volume time series on the observed road network segment. Notations and abbreviations are adopted from Kolda and Bader (Kolda and Bader 2009).

The decomposition method used in this article is the CANDECOMP/PARAFAC (CP) in its non-negative form. The CP decomposition factorizes a tensor into a sum of component rank-one tensors. For tensor τ CP is the following:

$$\tau = \sum_{r=1}^{R} a_r \circ b_r \circ c_r \tag{3.1}$$

where R is a positive integer that represents the decomposition rank. Then, rank one components can be expressed as factor matrices $A \in (a^{(1)}\, a^{(2)} \ldots a^{(R)})$, $B \in (b^{(1)}\, b^{(2)} \ldots b^{(R)})$, and $C \in (c^{(1)}\, c^{(2)} \ldots c^{(R)})$. Most of the authors predefined a tensor rank based on the underlying knowledge of the phenomena that is observed (Wang et al. 2014). The Core Consistency Diagnostic (CORCONDIA) (Bro and Kiers 2003) method is used in this paper.

3.5 Methodology

This paper aims to propose a tensor-based road traffic pattern extraction method for the purpose of spatiotemporal anomaly detection. The proposed methodology is presented in Figure 3.3 Proposed methodology for the anomaly detection. and encompasses the main steps: (i) Data preprocessing, (ii) grid-based map segmentation with STMs computation, and (iii) anomaly detection based on the CoM estimation.

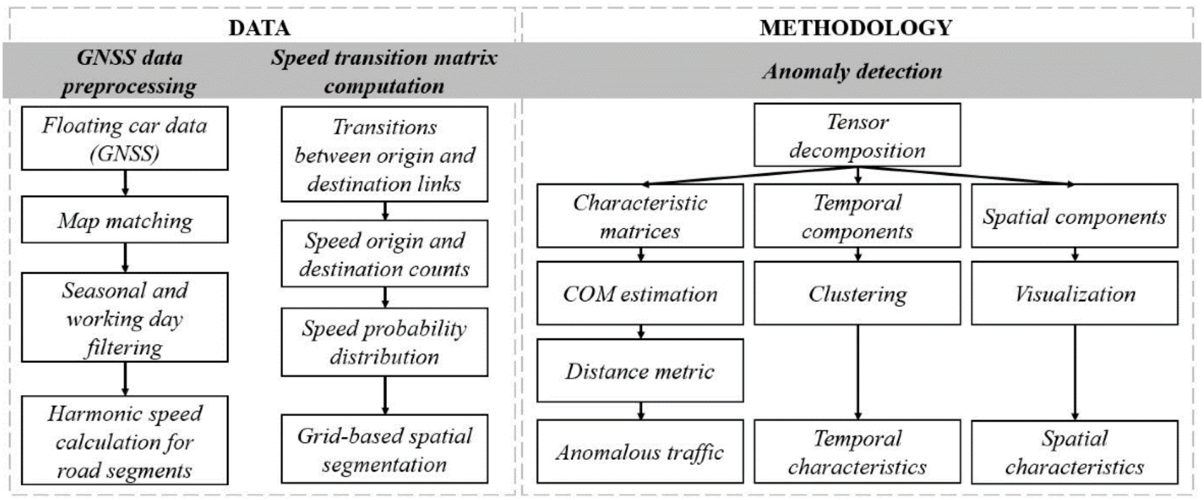


Figure 3.3 Proposed methodology for the anomaly detection.

3.5.1 Grid-Based Map Segmentation

When constructing a tensor, many researchers are using one tensor to model the spatiotemporal dataset. In this paper, the goal was to extract many different traffic flow patterns to capture more diverse patterns that are different for many parts of the city. The grid-based map segmentation approach is used to divide the city into many smaller cells. Then, for every cell, all transitions were extracted. The cell's size was fixed to 500 × 500 m. According to (Carić and Fosin 2020), this cell size is sufficient to capture the most important traffic patterns. Transitions were further filtered by discarding every road segment with a speed limit smaller than 50 kmph. This filter was used to avoid any possible false anomaly detection regarding the observed road segments' low speed or parking lots. There are many different approaches for map segmentation. This approach is used because this paper aims to find and analyze the anomalies on a city-wide scale and give an overview of the city's most dangerous road segments and possible problems with inadequate traffic signalization.

3.5.2 Tensor Construction

In this paper, the spatiotemporal tensor composed from the STMs is proposed as a traffic data modeling method. The tensor is constructed by flattening STMs into matrix as a frontal slice, placing transitions as spatial components, and adding time intervals as temporal components, represented in Figure 3.4. Tensor $\tau \in \mathbb{R}^{m \times n \times t}$ is constructed, where m represents the flattened size of the STM, n represents the number of observed transitions in the road network, and t represents eight time intervals. Frontal slices of tensor τ can be represented with matrix $\tau_{::t} \in \mathbb{R}^{m \times n}$, where every STM matrix X is flattened into a vector $x \in \mathbb{R}^{m \times 1}$ and placed into the matrix $\tau_{::t}$ as column. Dimension m had the value of 400 as STM size is 20×20. Instead of using one tensor with all the data, data is divided into several smaller tensors using the grid-based map segmentation, where n represents the number of the transitions inside one grid cell. Then, the final form of the tensors is $\tau^{(1)}, \tau^{(2)}, \dots, \tau^{(N)}$ where $\tau^{(i)} \in \mathbb{R}^{400 \times n \times 8}$. With the proposed approach, anomalies can be captured from different parts of the road network, while smaller spatial dimensions of the cells allow capturing more diverse traffic patterns.

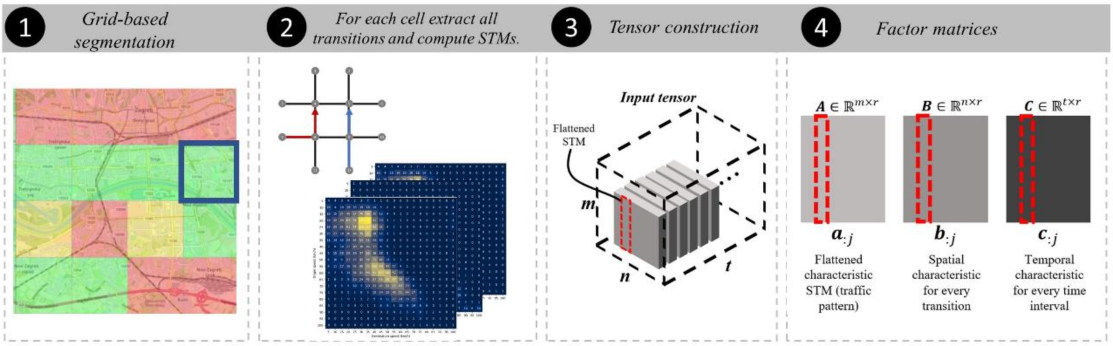


Figure 3.4 Steps that are describing the tensor construction method using the STMs: (1) Grid-based map segmentation, (2) STM extraction, (3) tensor construction, and (4) factor matrices.

3.5.2.1 *Tensor Rank Estimation*

In this paper, CORCONDIA is applied as the tensor rank estimation method using the AutoTen algorithm (Papalexakis 2016). It is essential to mention that tensor rank estimation methods are used to get recommendations more than the rank's exact actual value. The algorithm was run five times on randomly chosen tensor $\tau^{(i)}$, and the average estimated rank resulted in a value of $R = 10$, which is the rank used for the experiments.

3.5.2.2 *Factor Matrix Discussion*

The tensor decomposition resulted in three factor matrices $A \in \mathbb{R}^{400\times10}$, $B \in \mathbb{R}^{n\times10}$, and $C \in \mathbb{R}^{8\times10}$ as presented in Figure 3.4. Factor matrix A consists of extracted characteristic traffic patterns on the road network. If the column $a_{:j} \in \mathbb{R}^{400\times1}$ of the factor matrix A is reshaped into the matrix 20×20 it represents the characteristic STM (traffic patterns). The goal of anomaly detection is to find the anomalous traffic patterns and link them to the corresponding values in spatial and temporal factor matrices. The matrix B represents the spatial factor matrix, and the values in the rows

$b_{i:}$ represent how well each of the characteristic STM represents the traffic flow on the corresponding transition on index i. The values in the columns $b_{:j}$ show how well each characteristic matrix describes each of the transitions (spatial components) in the observed road network. The matrix C represents the temporal factor matrix. The values in the rows $c_{i:}$ represent how well each of the characteristic STM represents the corresponding time interval on index i, and the values in the columns $c_{:j}$ show how well each characteristic matrix describes each of the time interval (temporal components). The larger values in the factor matrices B, and C suggest the greater impact of the spatial or temporal components on the corresponding factor (Qi et al. 2019).

3.5.3 Anomaly Detection

When working with traffic data represented by the STM, the anomalous traffic can be represented by large deviations between origin and destination speeds, which highly depends on the represented pattern's position. These dangerous traffic situations can be identified by the vehicle's sudden braking, or a very high acceleration, with the corresponding positions in the STM, lower left, and the upper right corner. Then, normal traffic behavior can be represented if the position of the pattern is close to the diagonal of the STM. Those values represent normal traffic behavior, which extends from the congested (upper left corner) to the free traffic flow (lower right corner). All scenarios are illustrated in Figure 3.5.

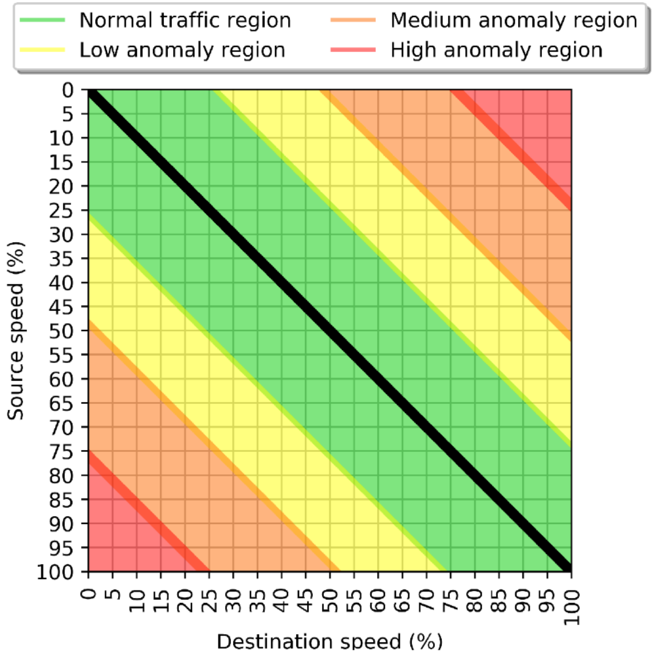
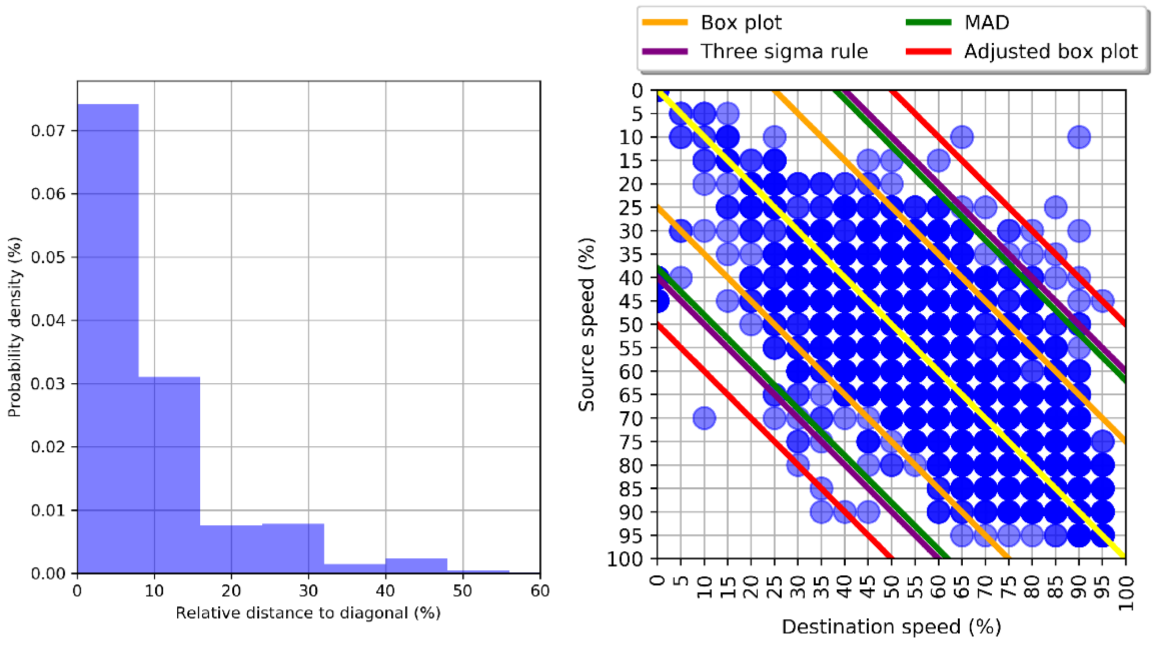


Figure 3.5 Regions in the STM that shows pattern location importance for anomaly detection.

To amplify the importance of the location of the patterns, the method for the anomaly detection is based on (i) CoM estimation for the pattern represented by the characteristic STM, and (ii) measuring the relative distance between CoM and the diagonal of the STM. CoM is computed for every characteristic STM extracted from the tensor decomposition method. With this, extracted CoM represents the most probable speed transition in the characteristic STM. This approach is used because the position of the pattern represented by the STM is crucial for the traffic state estimation and the anomaly detection (Leo Tišljarić et al. 2020). CoMs are computed based on the computation of the expected value, adopted from (Jordaan 2005). Firstly, marginal distribution for the coordinates (origin and destination speed) are computed using Equations (2.2) and (2.3). Then, the CoM coordinates are computed using (2.4) and (2.5).

After the CoM estimation, the relative distance between the CoM and the diagonal is measured using the Euclidean distance. The most suitable anomaly detection method is chosen by comparison of the most used methods represented in Table 3.1, that reports the name of the anomaly detection method, number of anomalies detected, and the lower and upper bounds. Relative distance from the diagonal values that are placed outside of the computed bounds are considered anomalous ones. The box plot method resulted in detecting the most anomalies, three sigma rules, and MAD resulted in detecting the same number of anomalies, while, adjusted box plot detected eight anomalous characteristic matrices. After examining the relative distance distribution (Figure 3.6a), the adjusted box plot is chosen as an anomaly detection method. It is a method that does not take any parametric assumptions and uses med couple as a robust skewness estimator (Hubert and Vandervieren 2008). Other methods assume the normal distribution of the data and cannot be used in this case. The results of applying all the methods can be observed in Figure 3.6b, where the plotted lines show the upper bound of the anomaly detection methods with plotted CoMs for every calculated characteristic STM resulted in tensor decomposition. It can be observed that the adjusted box plot resulted in detecting only the most anomalous transitions regarding the anomaly definition presented in Section 3.4.1



(a) (b)

Figure 3.6 Choosing the anomaly detection method: (a) Distribution of the relative distances to the diagonal of the STM, and (b) CoMs of the characteristic matrices with labeled anomaly measures results.

Table 3.1 Comparison of the multiple anomaly detection methods.

Method	N. Anomalies Detected	Bounds
Box plot	261	[−15.00,25.00]
Three sigma rule	58	[−20.12,39.13]
MAD	58	[−24.38,38.52]
Adjusted Box plot	8	[−4.65,46.13]

The tensor-based anomaly detection method is presented in Algorithm 3.1. The algorithm begins with the empty lists initialization, namely, list of tensors τ and the list of anomalous characteristic matrices M. Every tensor $\tau^{(i)}$ is constructed by flattening every STM recorded inside the spatial cell $g^{(i)} \in G$, and constructing a frontal slice $\tau_{::t}$ explained in detail in Section 3.5.2.2. Then, on every tensor, the Non-negative Tensor Decomposition (NTD) is applied to compute three factor matrices. A_i represents extracted traffic patterns, where every row $a_{:i}$ represent flattened extracted characteristic matrix, B_i spatial factor matrix, and C_i temporal factor matrix. Then, every $a_{:i}$ is reshaped into two dimensional STM $X_{ch}^{(i)}$ and CoM coordinates are computed using Equations (2.2)-(2.5). The final step is the anomaly detection, which is estimated by using two distances d_{CoM} and d_A, where d_{CoM} represents the distance from the CoM to the diagonal of the STM, and d_A threshold distance for the anomaly detection visually shown in Figure 6. The anomaly is detected if the d_{CoM} is larger or equal than the d_A and placed into list M.

Algorithm 3.1 Tensor-based anomaly detection pseudo code

Input: Spatial cells G, STMs

Initialize empty list of tensors τ

Initialize empty list of anomalous characteristic matrices M

for each spatial cell $g^{(i)}$ in G **do**

  Construct a new tensor $\tau^{(i)} \in \mathbb{R}^{m \times n \times t}$ using STMs for cell $g^{(i)}$

  Add new tensor $\tau^{(i)}$ to list τ

end for

for each tensor $\tau^{(i)}$ in list τ **do**

  Apply Non-negative Tensor Decomposition on $\tau^{(i)}$ and store the result in matrices A_i, B_i, and C_i

  for each flattened characteristic matrix $a_{:i}$ in A_i **do**

    Reshape matrix $a_{:i}$ to 20×20 matrix and set it as $X_{ch}^{(i)}$

    Compute CoM coordinates $c_x^{(i)}$ and $c_y^{(i)}$ from $X_{ch}^{(i)}$

    Compute distance d_{CoM} between CoM coordinates and the diagonal of $X_{ch}^{(i)}$

    if $d_{CoM} \geq d_A$ then

      Add $X_{ch}^{(i)}$ to list M

    end if

  end for

end for

3.6 Results

3.6.1 Data

A real-life dataset was provided by the Mireo Inc. from Zagreb, Croatia (Erdelić and Ravlić 2016). It consists of large GNSS data collected between 2009 and 2014 by vehicle fleet with the size of approximately 5000 vehicles (Table 3.2). The dataset includes around 6.55 billion GNSS records driven across all Croatia. For this paper, the dataset is filtered to represent data for the City of Zagreb, as a mid-size city in the European context with a population of around 800,000 people. To lower deviations due to the road traffic seasonality issue (Capparuccini et al. 2008), weekend days, the summer months, July, and August are excluded from the dataset. The grid-based map segmentation and the filtering were applied to the dataset, and the results are shown in Figure 3.7, where green represents the cell with the data, and red cells are excluded from this research.

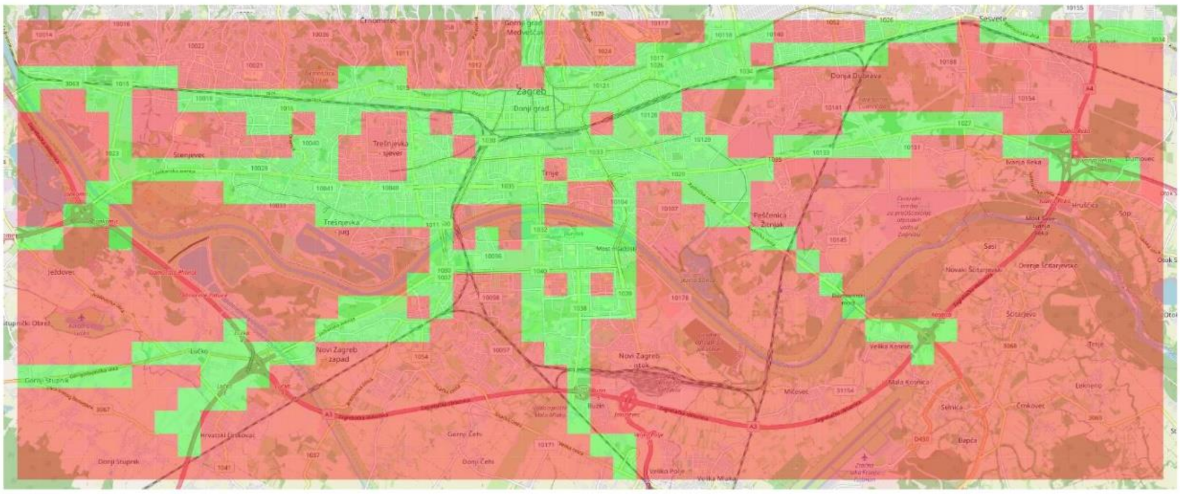


Figure 3.7 Result of the grid-based map segmentation and the data filtering process.

Table 3.2 Data summary.

Number of GNSS traces	6.55 billion
Sampling rate	100 m/5 min
Time-span	August 2008–October 2014
Number of vehicles	4200
Number of road segments (Croatia)	2,000,000
Number of road segments (Zagreb)	86,900

3.6.2 Anomalous Traffic Patterns

This section shows the evaluation results using the real-life dataset. Figure 3.8 presents eight extracted characteristic matrices, with corresponding temporal components, in which the morning (07:25–08:20) and evening (15:30–17:05) rush hours are labeled with striped, green lines. Figure 3.9 represents the spatial placement of the extracted anomalous cells, where every anomaly event is labeled with a letter that corresponds to the Figure 3.8 labels.

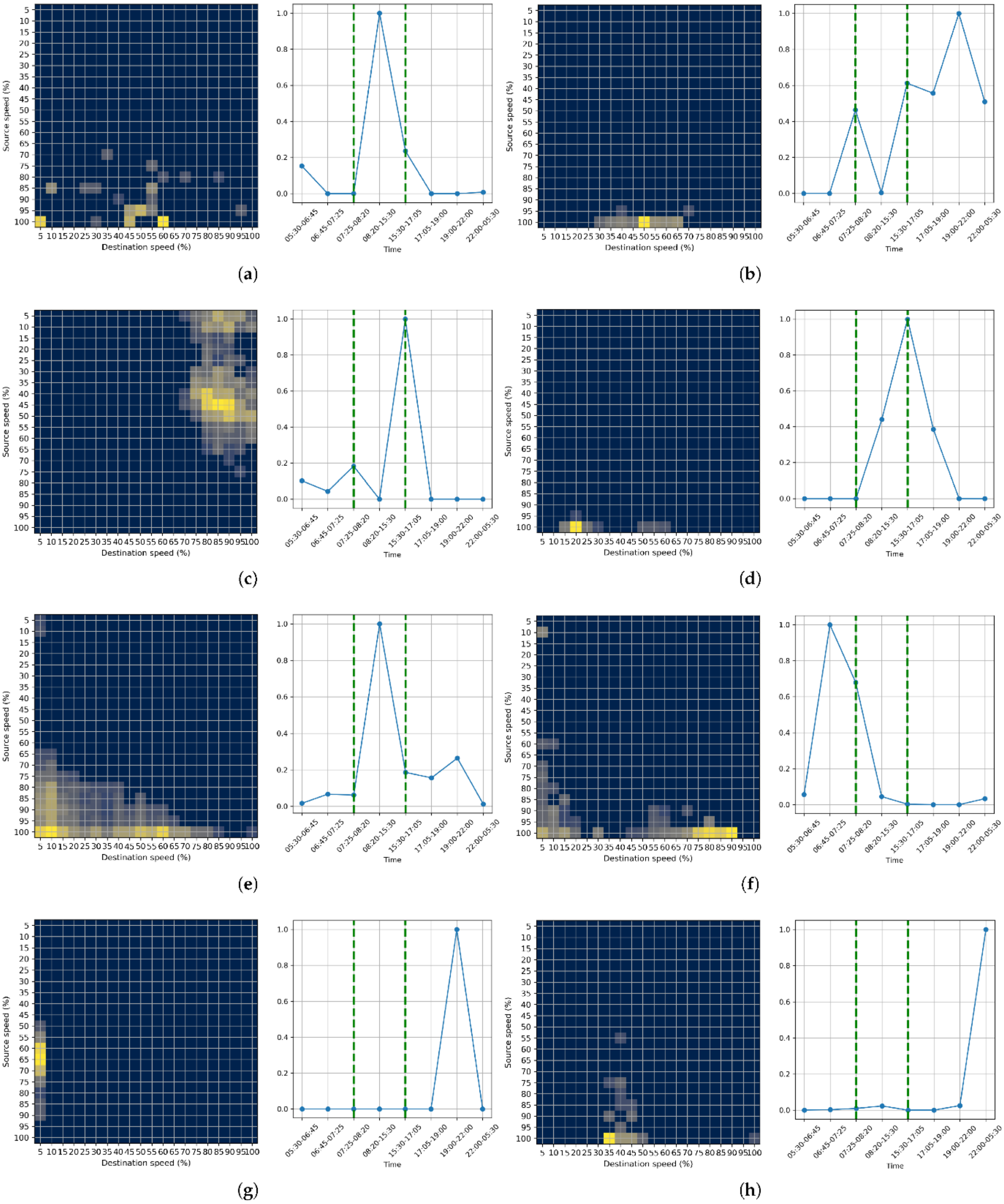


Figure 3.8 Results of the anomaly detection; (a–h) represent characteristic matrices which represent anomalous patterns (left) with corresponding temporal components (right).

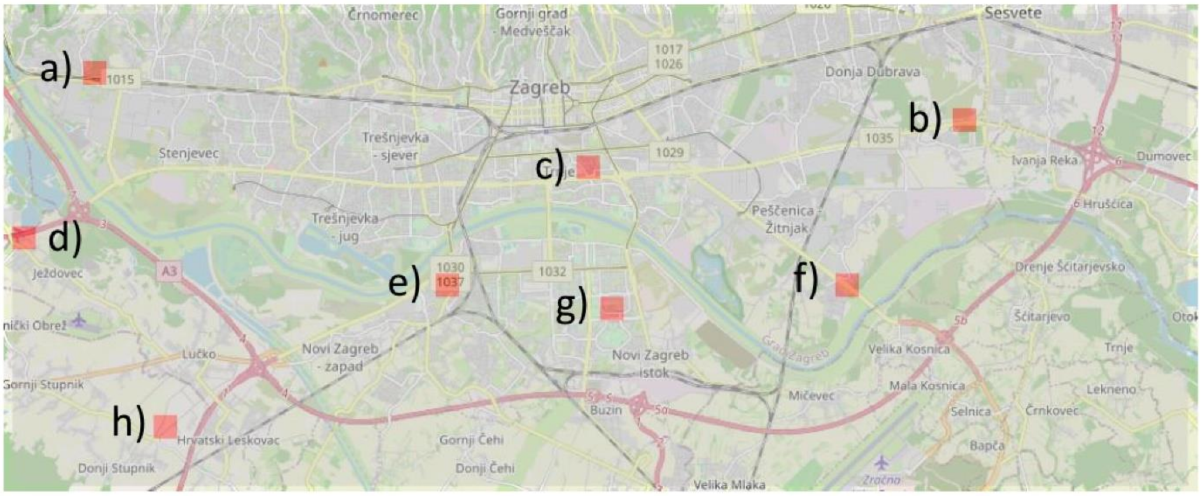


Figure 3.9 Positions of the anomalous cells on the map (a–h) represent most anomalous parts of the traffic network in the City of Zagreb.

All characteristic matrices, except for (c) example, represent the same anomaly type as the CoM placed at the matrix's lower-left corner. This traffic situation is characterized by sudden brakes and large speed decrease when traveling from origin to destination links. These situations occur when a vehicle is facing congestion ahead and represent a potentially serious safety threat. The (c) example is the opposite situation, where the CoM is placed in the upper right corner. Here, a different but potentially dangerous event occurs, where vehicles are accelerating from low to very high speeds.

Most of the temporal components indicate that anomalous events occur at rush hours. It can be observed that the period between rush hours (08:20–15:30) and the evening rush hour is the most represented case. This claim is justified because in rush hours, many vehicles are on the roads, and the anomaly probability arises.

Figure 3.9 represents the spatial placement of the extracted abnormal cells, where every anomaly event is labeled with a letter that corresponds to the labels in Figure 3.8. The spatial placement of the anomalies indicates two spatial clusters: (i) Edges of the city represented with the examples (a), (b), (d), (f), (h), and (ii) city center represented with the examples (c), (e), and (g). The cluster of transitions placed at the edges of the city points to the congestion related to daily commuters traveling to work from outside of the city center. By considering temporal components, it can be observed that the evening rush hour contributes mostly to the congestion and the anomalous events in the city, except for example (f), where the anomalous events mostly occur during the morning rush hour. The city center cluster is characterized by the temporal components that point to the intervals between rush hours, evening rush hour, and the interval later in the day, after the evening rush hour. This behavior is mostly attributed to the inefficient traffic signalization, which leads to the prolongation of the anomaly events mostly caused by the rush-hour congestion. Similar behavior can be caused by the tourist attractions and other entertainment facilities provided at the city center. During the transition in example (e), the most congested bridge in the city was captured. This information indicates the possible usable information for the urban planners by suggesting the need to build a new bridge that will connect the north and the south parts of the city.

3.6.3 Domain Knowledge Validation

The HCM provides methods for computing relevant traffic parameters to estimate the capacity and the level of service for different road types (Anon 2010). Level of service is defined using a relative traffic flow speed on the observed road segments, labeled with letters from A to F, where A represents the best traffic conditions, with vehicle speeds larger than 80% of the free-flow speed, and F represents the most extreme congestion, where vehicle speeds are less than 30% of the free-flow speed. The 2000 STMs were labeled using the HCM data for the level of service. STMs were labeled as anomalous only if the transition contains a significant change in the level of service, i.e., from A to F, or from F to A. With this setup, the recurrent congestion was not detected as an anomaly. In other cases, STMs were labeled as normal. With this setting, only the most extreme anomalies were labeled as abnormal. Firstly, 500 anomalous and 500 STMs without the anomaly were selected randomly from the labeled data as a training dataset. Then the results of our approach were compared to the HCM classification as the ground truth. We report the precision calculated as *true positive / (true positive + false positive)*, recall as *true positive / (true positive + false negative)*, and F1 score in Table 3.3.

Table 3.3 Validation results of the proposed method by using the domain knowledge data.

Anomalous STMs	Normal STMs	Precision	Recall	F-1
500	500	92.88%	87.55%	90.14%

3.6.4 Comparison to Other Approaches

This section compares the proposed approach to other approaches for road traffic anomaly detection. While there are many tensor-based approaches focused on congestion estimation, missing data imputation, and event detection, there are only a few specialized in anomaly detection on urban roads. Table 3.4 shows several tensor-based approaches for the anomaly detection of road traffic networks. Most of the authors are using O-D matrices that show the number of the vehicle (volume) traveling between two points in the traffic network (Fanaee-T and Gama 2016; Lin et al. 2018; Wang et al. 2019). Regarding the potentially large spatial distance between the O-D pairs, those approaches extract the patterns and detect global anomalies related to traffic fluctuations (Lykov and Asakura 2020). Therefore, most of the authors focus on the detection of the general events that are related to traffic movements like tourist attractions or other social events (Chen et al. 2017; Lin et al. 2018).

Table 3.4 Comparison of the proposed approach to other approaches for the traffic anomaly detection.

Literature	Data Type	Traffic Parameter	Anomaly Detection
Fanaee et al. (Fanaee-T and Gama 2016)	O-D matrices (car)	Traffic volume	Traffic volume
Wang et al. (Wang et al. 2019)	O-D matrices (car)	Traffic volume	Traffic Flow
Lin et al. (Lin et al. 2018)	O-D matrices (car)	Traffic volume	Event detection
Chen et al. (Chen et al. 2017)	GNSS (bicycle)	Traffic volume	Event detection
Lykov et al. (Lykov and Asakura 2020)	Simulation (car)	Traffic speed	Traffic patterns

This paper focuses on detecting the anomalies that affect the traffic flow on micro-locations (transitions). The proposed approach is more suitable for detecting anomalies that could potentially lead to traffic accidents like sudden braking or fast acceleration. Therefore, this approach can be applied in real-time traffic accident detection and prediction.

Alongside the possible applications of the proposed methodology, some drawbacks of the method must be addressed in further research. The anomaly detection method is based on the speed, which can lead to false anomaly detection on a short road segment bounded with the non-synchronized intersections. This is the reason why the short road segments with the speed limit less than 50 kmph were excluded from this research. Secondly, more narrow time intervals like 5-, 15-, or 30 min are used in most of the road traffic related research. Narrower time intervals could provide more informative results with the possibilities of implementing the method on a real-time case study especially in the environment with the mixed traffic flows (Vrbanić et al. 2021). Further improvements of the method could also include the analysis of the interactions between the consecutive cells. The modeling process of the STM should also be addressed. The important parameters of the STM like size, discretization period, and cell size should be analyzed for optimization purposes.

3.7 Conclusions

For the development of more secure, cleaner, and overall, more sustainable cities, traffic congestion and corresponding anomalies must be addressed. This paper presents a novel method for the extraction of road traffic patterns and anomaly detection using tensor-based method. It integrates a tensor decomposition with the anomaly detection approach based on estimating the CoM of the observed traffic pattern represented by the STM. This method is evaluated on a large real-life GNSS road traffic dataset and validated using the domain knowledge data. The result presents valuable traffic insights useful for the routing application, responsible urban planners, and road infrastructure maintenance authorities. It can be used as valuable traffic information about the need for infrastructure expansion, additional improvement strategies, or to analyze the traffic influence of the new road infrastructure.

Compared to other approaches related to road traffic anomaly detection, the proposed method is more focused on detecting the anomalies that affect the traffic flow and could lead to dangerous situations and, consequently, to traffic accidents. Furthermore, anomaly detection will include the expansion of the proposed method for the real-time anomaly detection framework.

Chapter 4

Motorway Bottleneck Probability Estimation in Connected Vehicles Environment Using Speed Transition Matrices

This chapter has been published as: Tišljarić, L.; Vrbanić, F.; Ivanjko, E.; Carć, T. Motorway Bottleneck Probability Estimation in Connected Vehicles Environment Using Speed Transition Matrices. Sensors 2022, 22, 2807. https://doi.org/10.3390/s22072807

Author Contributions: Conceptualization, L.T.; methodology, L.T. and F.V.; software, L.T. and F.V.; validation, L.T.; formal analysis, L.T. and F.V.; investigation, L.T. and F.V.; resources, E.I. and T.C.; data curation, L.T. and F.V.; writing—original draft preparation, L.T. and F.V.; writing—review and editing, L.T., F.V., E.I. and T.C.; visualization, L.T. and F.V.; supervision, E.I. and T.C.; project administration, E.I. and T.C.; funding acquisition, E.I. and T.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partly funded by European Regional Development Fund under the grant KK.01.1.1.01.0009 (DATACROSS), and by the Croatian Science Foundation under the project IP-2020-02-5042.

4.1 Abstract

Increased development of the urban areas leads to intensive transport service demand, especially on urban motorways. To increase traffic flow and reduce congestion, motorway traffic bottlenecks caused by high traffic demand need to be efficiently resolved using Intelligent Transport Systems services. Communication technology development that supports Connected Vehicles (CVs), which act as an active mobile sensor for collecting traffic data, provides an opportunity to harness the large datasets to develop novel methods regarding traffic bottlenecks detection. This paper presents a speed transition matrix-based model for bottleneck probability estimation on motorways. The method is based on the computation of the speed at the vehicle transition point between consecutive motorway segments, which forms a traffic pattern that is represented using transition matrices. The main feature extracted from the traffic patterns was the center of mass, whose position is used as an input to the fuzzy-based system for bottleneck probability estimation. The proposed method is evaluated on four different simulated motorway traffic scenarios: (i) traffic collision site, (ii) short recurring bottleneck, (iii) long recurring bottleneck, and (iv) moving bottleneck. The method achieves comparable bottleneck detection results on every scenario, with a total accuracy of 92% on the validation dataset. The results indicate possible implementation of the method in the motorway traffic environment with a high CVs penetration rate using them as the sensory input data for the control systems based on the machine learning algorithms.

Keywords: motorway bottleneck; connected vehicles; bottleneck detection; bottleneck probability; speed transition matrix; fuzzy-based bottleneck probability; traffic simulation

4.2 Introduction

Bottleneck detection on motorways is a known and investigated research topic owing to the increased development of urban areas and intense transport service increase due to globalization (Gong and Yang 2009). The result of bottleneck occurrence is increased travel time, decreased traffic safety, and increased pollution due to stop-and-go driving.

Connected Vehicles (CVs) have emerged recently due to the development of communication technologies and Intelligent Transport Systems (ITS) services. Recently, many researchers are covering the CVs topic with the research regarding collaborative machine learning (Barbieri et al. 2022), CVs in the mixed traffic flow environment (Zhang et al. 2021), CVs security (Pascale et al. 2021), and upcoming challenges (Lu et al. 2014).

The traffic bottleneck is a phenomenon that can occur in urban and motorway roads. According to (Iordanidou et al. 2015), the most common reasons for the bottleneck activation are merging of the on-ramp vehicles, lane drop, intense braking, fixed speed limits, and traffic incidents. All these events can be classified as unexpected driver behavior on the motorway. In this paper, the proposed method for bottleneck detection is focused on detecting the unexpected behavior that could lead to bottleneck start or the events that occur when vehicles are already in the bottleneck state. Bottleneck detection is conducted by observing the vehicle transitions between consecutive motorway segments. The bottleneck probability will rise when events like sudden breaks, stopping of the vehicle, or intense accelerations are detected.

As discussed in (Wegerle et al. 2020), standard traffic flow models are not able to represent traffic flow in a way to detect traffic breakdowns manifested in bottleneck occurrences. The CV technology enables the data collection in real-time from every vehicle on the traffic segments. Those data can be incorporated into emerging traffic models to detect bottlenecks efficiently and more accurately. In the context of mentioned research possibilities, this paper proposes a novel method for the road traffic bottleneck detection on motorways using the emerging traffic data modelling technique Speed Transition Matrix (STM) discussed further in (Leo Tišljarić et al. 2020; Tišljarić et al. 2021). The method is based on the CV data collection on the observed motorway, focusing on traffic patterns that emerge between consecutive motorway segments. Traffic patterns are represented with the STMs, while the bottleneck probability occurrence is estimated by applying the Fuzzy Inference System (FIS) with the position of the traffic pattern as the input variables. The main advantage of the proposed method is the ability to detect bottlenecks regardless of the current state of the traffic flow, and bottleneck type (moving, recurrent, etc.).

In this context, the contributions of this paper are as follows:

- Proposed method for the spatiotemporal motorway traffic patterns extraction which includes STM-based model.
- Proposed bottleneck detection method based on computation of parameters extracted from the STM.
- Bottleneck detection results evaluated using different motorway traffic scenarios that include collision site, recurrent, and moving bottlenecks prove the effectiveness of the proposed method.

This paper is organized as follows. Section 2 presents the literature overview of the methods and parameters used for road traffic bottlenecks detection. In Section 3, the background, definitions, and concepts used in this paper are explained. Section 4 presents the methodology overview for the proposed motorway bottleneck detection method. In Section 5, the simulation scenarios and used parameters are explained. Section 6 presents the results of the proposed method with the validation on the simulated dataset. Section 7 presents the advantages and disadvantages of the proposed method usage with the possible application directions. In Section 8, the conclusion and future work directions are given.

4.3 Literature Review

4.3.1 Traffic Parameters

The first step in bottleneck detection is choosing the traffic parameter that will be used. Many researchers use traditional traffic parameters to extract the traffic congestion areas and identify the bottleneck. The most common traffic parameters used for traffic data analysis and the road traffic scenario creation are speed, traffic flow, and density (Vrbanic, Miletic, et al. 2021). The observed parameters are commonly modelled as profiles, which are represented with vectors $A \in \mathbb{R}^m$, where $A = (a^{(1)}, a^{(2)}, \dots, a^{(m)})$ and $a^{(i)}$ represents the observed traffic parameter in time interval $i \in m$ (Erdelić et al. 2021). The main disadvantages of the vector representation are its inability to represent spatiotemporal relations between multiple road segments, and values of the observed parameters are aggregated into narrow time intervals, leading to large deviations. On the other hand, spatiotemporal traffic data are often represented with the traffic matrix $B \in \mathbb{R}^{m \times n}$, called traffic image due to a grid-based representation, commonly visualized with heatmaps (Gregurić, Vujić, et al. 2020). Here, a respective traffic parameter is represented using matrix $B = (b^{(1)}, b^{(2)}, \dots, b^{(m)})$, where m represents time intervals and n motorway segments. Matrix data representation can represent spatiotemporal data but requires more computing power and advanced analysis techniques. Speed and density are the most common parameters used in the research for bottleneck detection. In (Gong and Yang 2009), authors propose computation of the congestion index based on the headway and density. Then, the bottleneck is identified by comparing the changes in the index in the spatial and temporal domain. In (Li et al. 2020), authors use a speed threshold to detect congestion that is defined as n% of the average vehicle speeds on the observed road segment, where n varies between 10 and 90.

Transition models can be seen in 90's papers within Daganzo's research (Daganzo 1994). The author presented a cell transmission model with a density as the main parameter for traffic state estimation. The main goal of the transition cell model was to observe the traffic parameter at the transition between two consecutive cells. The parameter difference due to the transitioning is then used as a measure for the representation of the motorway traffic flow.

In this paper, the speed transition model is proposed by using the STMs. STM does not suffer from data loss due to averaging the values into narrow intervals because speed data are not averaged in such a way. Average values of the speed at single road segments are computed using all collected speed data on the observed road segments. Thus, there are two challenges related to classic averaging that need to be addressed: (i) averaging thousands of scalar values often results in wide confidence ranges, and (ii) values represent the speed only on one road segment, with no data related to the interactions with the consecutive road segment. The STM averages the speed values not on one road segment, but on the transition point between two consecutive segments, which includes pair of two speeds that show the interaction between two adjacent road segments. By using the Center of Mass (CoM) measure, pairs of origin and destination speed values are averaged using the harmonic mean. Thus, the speed for one vehicle is computed using harmonic mean, on origin and destination road segments for one transition, and placed into the STM. With this procedure, we weighted the speed average with the CoM, and spatiotemporal relations of the transition were captured. Thus, the change of the traffic pattern position in the matrix is observed during the transition, which provides a more visually interpretable way of representing and detecting the bottlenecks.

4.3.2 Bottleneck Detection

The authors of (Coifman and Kim 2011) analyzed existing bottleneck modelling techniques on the motorways. They concluded that models using fundamental traffic diagrams do not adequately describe the bottleneck phenomenon due to the decreased traffic flow and speed reduction observed in downstream traffic flows.

Consequentially, there is a need to develop novel modelling methods that include interactions between consecutive road segments on motorways.

In (Li et al. 2020), the authors proposed congestion bottleneck definition using computed congestion level cost and contagion cost related to congestion propagation to other road segments. The authors of (Sun et al. 2014) analyzed different urban network topologies and concluded that a strong community structure could improve the network performance to resist the propagation of the bottlenecks. The bottleneck was modelled using the cell transmission model to propagate the congestion through the downstream traffic flow from one road segment to another. In (Kerner 2007), the authors used the three-phase traffic theory to identify bottlenecks that started due to the high on-ramp inflow. The bottleneck is identified in the phase when vehicles are transitioning between free-flow and synchronized flow states. The authors distinguished two types of congestion patterns: synchronized flow patterns and general congested patterns. The authors of (Dülgar et al. 2020) analyzed empirical random phase transitions when vehicles were transitioning between the free traffic flow and synchronized traffic flow. The study showed that those transitions can occur randomly and can be used for bottleneck detection and analysis. In (Wegerle et al. 2020), the authors proposed a method for predicting moving bottlenecks by using probe vehicles data. The method is based on recognizing the phase transitions between the free flow and synchronized flow defined in Kerner's three-phase traffic flow model to detect the bottleneck.

The main limitation of the mentioned motorway bottleneck detection approaches is the inability to detect an already existing bottleneck on the observed road segments. The flexibility of the proposed STM-based approach in this paper fills the gap, and it can detect several different bottleneck scenarios, including an already existing bottleneck, moving bottleneck, and recurrent bottleneck. The second limitation of the observed approaches relates to the flexibility of the methods to consider a bottleneck's length and duration. With the usage of the STM-based approach, this is achievable by counting the motorway segments affected by a bottleneck.

4.4 Background

4.4.1 Motorway Network Elements and Bottleneck Definitions

Definition 1. *Motorway road network: Formally, road networks are represented as directed graphs $G = (V, E)$, where E is a set of edges representing road network segments, and the intersections or connections between edges are represented with the set of vertices V. In the defined network, every edge $e^{(i)} \in E$ has a start vertex $v^{(i)}$ and end vertex $v^{(i+1)}$.*

Definition 2. *Spatial transition: Movement of one vehicle between two edges $e^{(i)}$ and $e^{(i+1)}$ in time interval Δt is defined as a transition. There are two types of edges in the transition: origin edge $e^{(i)}$ and destination edge $e^{(i+1)}$.*

Definition 3. *Speed transition: Vehicle's travel throughout a transition in interval Δt, where two speeds are measured to compute the speed transition: speed on the origin edge v_o and speed on the destination edge v_d. Speeds are computed as a harmonic mean speed to emphasize the importance of the smaller values to capture low-speed values.*

Definition 4. *Road traffic states: In this paper, three road traffic states categories are defined based on* (Elefteriadou 2016; Leo Tišljarić et al. 2020)*: (i) free flow, (ii) unstable traffic flow, and (iii) congested traffic flow. Free flow is defined as traffic conditions with no interactions between vehicles due to low traffic density. Thus, speeds at spatial transitions are close to speed limits. The unstable traffic flow is represented by the traffic conditions with some interaction between vehicles due to increased traffic density. The speeds at spatial transitions are approximately 50–80% of the defined speed limit. The congested traffic flow is characterized by traffic jams with speeds at spatial transitions close to zero.*

Definition 5. *Motorway bottleneck: Normal traffic conditions can be described with traffic flow $q_{in} \approx q_{out}$, where q_{in}stands for upstream traffic flow and q_{out} represents downstream traffic flow on the observed part of the motorway. The bottleneck is manifested by decreasing the downstream traffic flow due to the congestion. In this paper, three distinct traffic situations are defined as bottlenecks: (i) sudden breaks, as the start of a bottleneck where vehicles are approaching the congested area, (ii) heavy congested area, where vehicles are slowed down or* *not moving, and (iii) sudden acceleration area, where vehicles are leaving the congested area. With those events, three bottleneck scenarios can be described: the start of the bottleneck, vehicles in the bottleneck, and the bottleneck clearance. In this paper, the bottlenecks are described using the proposed bottleneck probability metric p_b. The p_b will show the probability of bottleneck occurrence in the observed motorway segments with the values [0,1]. Here, value of 0 represent the traffic state with no interactions between vehicles, while the value of 1 represents the occurrence of one of the mentioned traffic bottleneck scenarios. Thus, p_b will not show the distinction between three scenarios but show the bottleneck probability if the scenario occurs.*

4.4.2 Speed Transition Matrix

Definition 6. *The STM captures the vehicle's speed at the spatial transition to represent the speed probability change, and therefore it represents the speed probability distribution at one spatial transition in Δt. The harmonic speed is chosen as the traffic parameter regarding its property of favouring the lower values during the aggregation compared to average speed. This property enables recognition of potential bottleneck generation even with the low amount of measured vehicle speeds with high deviations. Measure vehicle speed is represented relative to the speed limit on the motorway road segments that are observed. The STM can be represented with the expression (2.1).*

There are five examples of the characteristic STMs on the motorway visually represented in Figure 4.1. Colors in the figures represent the speed change probability at the observed transition between two consecutive road segments labelled as origin and destination segment. Dark colors represent low probability, while light colors represent high probability of speed transition. The first example in Figure 4.1a shows that vehicles on the observed transitions had both origin and destination speeds, close to 100% of the speed limit. Thus, the traffic state can be defined as free flow with no congestion. The second example in Figure 4.1a shows the more unstable traffic state because vehicles have speeds close to 60% of the speed limit. This event can indicate the start of the congestion, but it cannot represent the start of the bottleneck regarding the relatively high speeds. Figure 4.1b, c represents the start and the end of the bottleneck, respectively. The bottleneck's beginning is characterized by the transitions from high-speed values to low values because vehicles are transitioning from free flow (or unstable flow) to congested flow. On the other hand, the bottleneck's end is characterized by the opposite event. Here, vehicles are transitioning from congested traffic flow to free flow. The last example in Figure 1e shows the congested traffic state that occurs "inside" the bottleneck and is characterized by very low speeds on one transition's origin and destination segments.

From the examples in Figure 4.1 it can be observed that the position of the observed traffic pattern, represented by the STM, has a crucial role in determining the traffic state on the observed motorway traffic segments. Therefore, CoM is chosen as a method for extracting the traffic state in this paper. The CoM estimation method is covered in more detail in the next section.

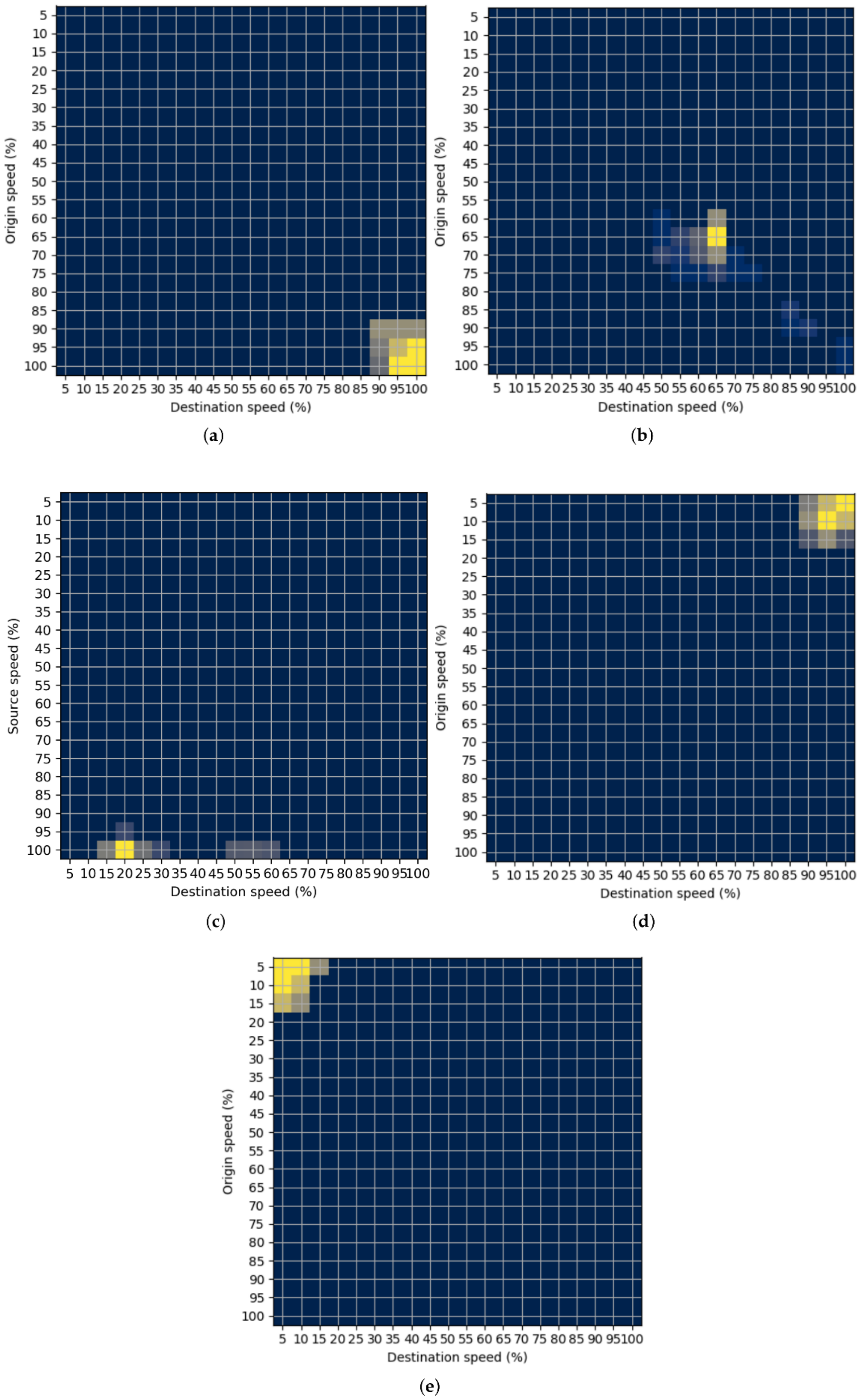


Figure 4.1 Examples of the characteristic STMs. (a) Free flow; (b) Unstable flow; (c) Bottleneck start; (d) Bottleneck end; (e) Heavy congestion.

4.5 Methodology

This paper aims to propose a methodology for motorway bottleneck probability estimation using traffic patterns extracted from the STMs. The overview of the methodology is presented in Figure 4.2 with three main steps. The first step is data preprocessing, which includes simulated vehicle position data preparation as input for STM computation. For every vehicle, its route is extracted with corresponding relative harmonic speed values, and it is matched to the corresponding edge. The second step includes the computation of the STMs, which represent the speed probability distribution for every vehicle travelling between two consecutive edges. Then, the last step includes FIS for bottleneck probability estimation with input variables computed as CoM distances from the origin and diagonal of the STM, respectively.

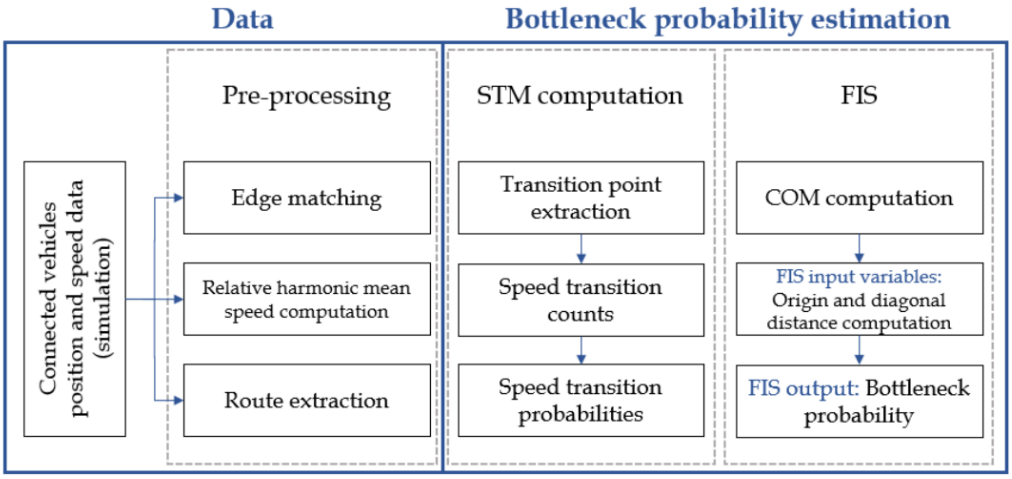


Figure 4.2 Overview of the methodology for the bottleneck probability estimation.

The bottleneck probability estimation is further explained in depth with the example in Figure 4.3. This example shows a simple motorway divided into seven road segments labelled as edges $e^{(i)}$ where $i \in 1,2, \dots 7$. Every STM represents the traffic state on the transition between two edges. The input and the output traffic flows are labelled with q_{in} and q_{out}. The traffic bottleneck begins at the edge $e^{(4)}$ and spans to $e^{(5)}$, which is caused by congestion.

To amplify the importance of the location of the patterns, the method for the motorway bottleneck detection consists of three parts: (i) estimation of the CoM for every traffic pattern represented with STM, (ii) computation of the distance between CoM and the diagonal of the STM, and (iii) computation of the distance between CoM and the origin of the STM. The next subsections explain each of the parts in detail.

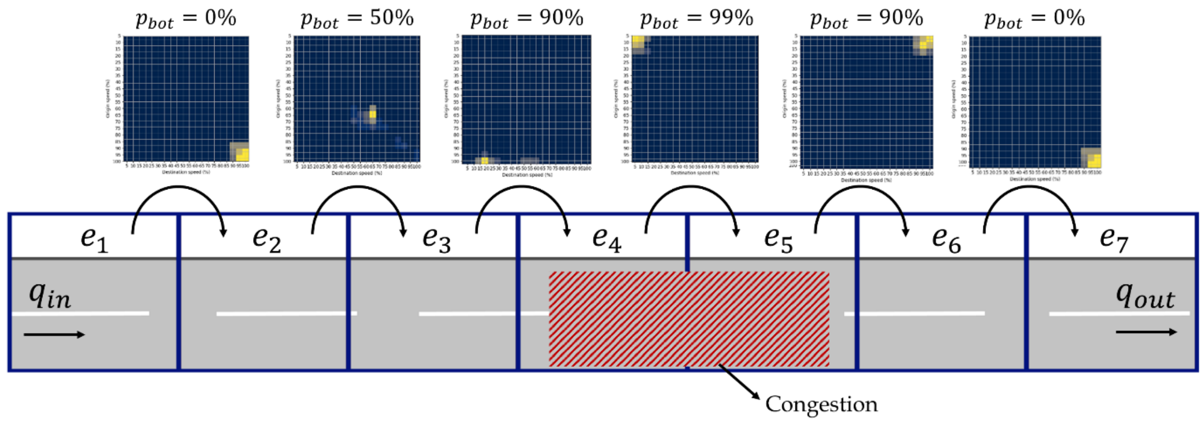


Figure 4.3 Overview of the proposed method for the bottleneck probability estimation.

4.5.1 Center of Mass Estimation

To successfully detect the motorway bottleneck caused by the congestion using the STM, the position of the traffic pattern represented by the STM is crucial. The position of the traffic pattern is used as the main feature in estimating the traffic state on the motorway. In this paper, the CoM, based on the computation of the expected value, adopted from (Jordaan 2005), is chosen as a method for detecting the position of the traffic pattern. To compute CoMs, expected values of the coordinates (origin and destination speed) are computed using Equations (2.2)-(2.5). Further explanation of the method can be found in our previous works (Tišljarić et al. 2021; Tišljaric et al. 2021).

4.5.2 Fuzzy Inference System

When the CoM is computed, appropriate features must be extracted to quantify the bottleneck probability. In this paper, two features are extracted from the computed CoM: (i) distance from the origin of the STM labelled as d_S, and (ii) distance from the diagonal of the STM labelled as d_D. Two features are shown in the example in Figure 4.4a.

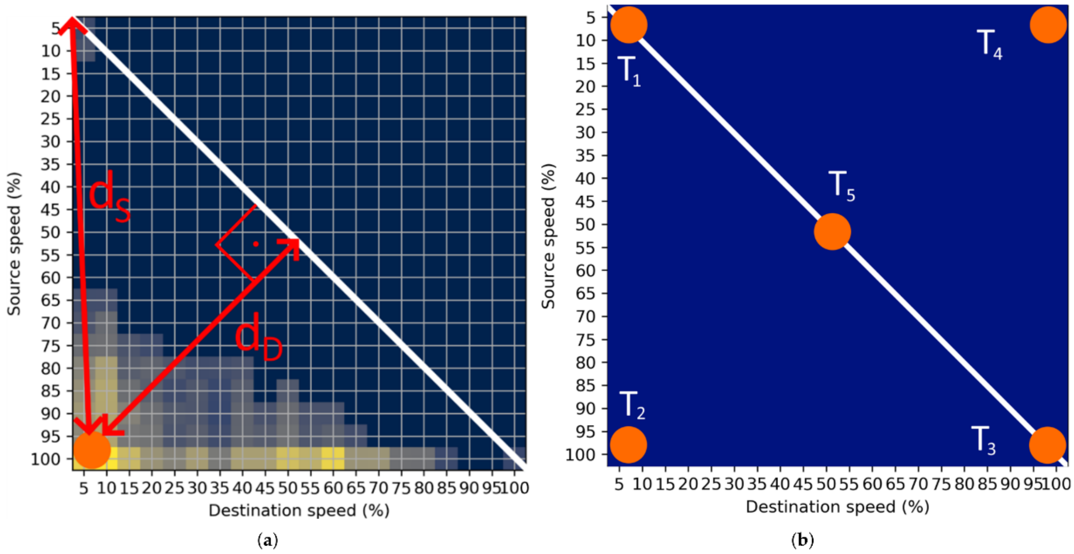


Figure 4.4 Method for FIS input variables computation. (a) Example of the d_S and d_D computation; (b) CoM positions of characteristic STMs.

The first feature, d_S, is important for estimating the traffic state on the observed transition. Let us examine some extreme points represented in the Figure 4.4b. At point T1, d_S is small. Then, the STM represents the transition of vehicles that had very low speed during the transition in the origin and destination segments, which can be declared as heavy congestion. On the other hand, at point T3, d_S has its largest value. Then, represented speeds will be very high (relative to the speed limit) in the origin and destination segments, which can be declared as free-flow conditions. At point T5, feature d_S is at 50% of the maximal value. This event can be declared as unstable traffic flow, as described in Figure 4.1b. It can be concluded that if the CoM is at the diagonal of the STM, d_S can be effectively used for the estimation of the traffic state.

However, the second feature must be introduced to represent other possible positions of the CoMs, for example, at points T2 and T4. At these points, feature d_S is at 70% of its maximal value, leading to a false conclusion that the traffic state is close to free flow. Additionally, the second feature d_D is at its maximal value and gives crucial information about the traffic state. At point T2, the start of the bottleneck can be observed because of transitions from very large origin speeds to very low destination speeds. On the other hand, at T4, clearance of the bottleneck can be observed as the origin speeds of the transition are low, and destination speeds are very high.

As features d_S and d_D can be represented as linguistic variables, an appropriately set FIS is used to detect the probability of the bottleneck occurrence. Fuzzy rules set for the bottleneck probability estimation is presented in Table 4.1. The rules were created using expert knowledge about the bottleneck definition extracted from (Elefteriadou 2016; Leo Tišljarić et al. 2020). An important consideration for rule setting was to address the behavior of the two input parameters d_S and d_D with the corresponding correlations.

Table 4.1 Set of fuzzy rules used for bottleneck probability estimation.

	d_D		d_S		p_b
IF	d_D is "small"	AND	d_S is "small"	THEN	p_b is "large"
IF	d_D is "small"	AND	d_S is "medium"	THEN	p_b is "medium"
IF	d_D is "small"	AND	d_S is "large"	THEN	p_b is "small"
IF	d_D is "medium"	AND	d_S is "small"	THEN	p_b is "medium"
IF	d_D is "medium"	AND	d_S is "medium"	THEN	p_b is "medium"
IF	d_D is "medium"	AND	d_S is "large"	THEN	p_b is "small"
IF	d_D is "large"	AND	d_S is "small"	THEN	p_b is "large"
IF	d_D is "large"	AND	d_S is "medium"	THEN	p_b is "medium"
IF	d_D is "large"	AND	d_S is "large"	THEN	p_b is "large"

Figure 4.5 represents the setup of the proposed FIS for bottleneck detection. It consists of two input variables d_S and d_D with corresponding output p_b that represents bottleneck probability. All variables are represented with range [0,1], relative to their

maximal values. The maximal value of the d_S is the length of the STM diagonal that can be computed as $20\sqrt{2}$, while the maximal value of d_D can be computed as $d_S/2$.

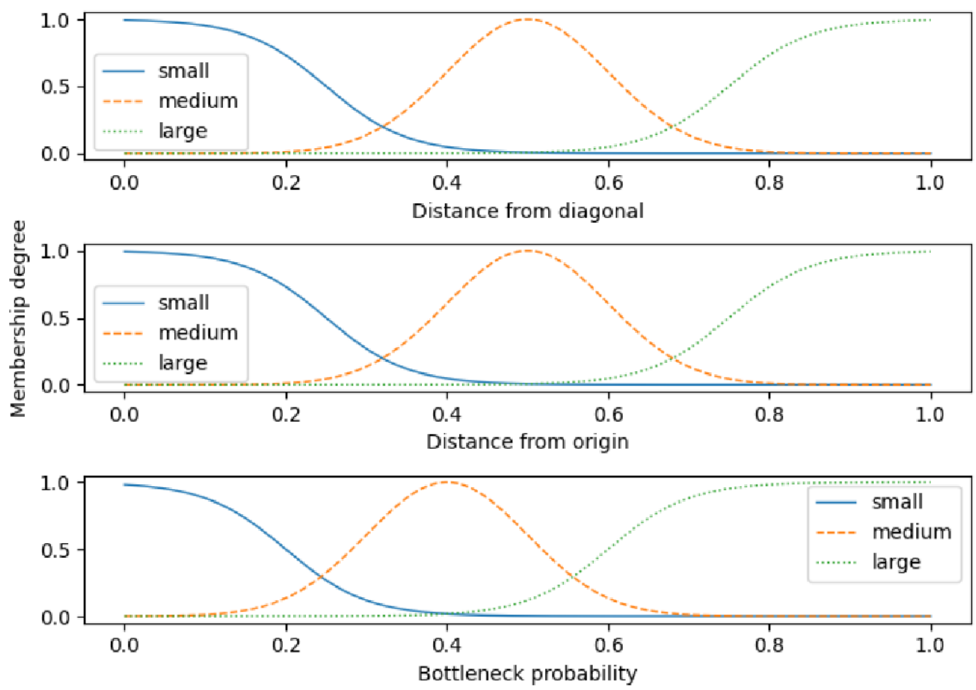


Figure 4.5 Initial FIS setup for the bottleneck probability estimation.

Every variable is represented using three linguistic expressions, “small”, “medium”, and “large”, representing the possible states of the variables. The term “small” is represented using the Z-type membership function, “medium” is represented by Gaussian function, and “large” is represented by S-type membership function. The bottleneck probability variable has an offset of 0.1 to the left to allow the system to be more sensitive to speed changes.

4.6 Simulation of Motorway Traffic

4.6.1 Simulation Setup

The framework for motorway bottleneck estimation was made in SUMO traffic simulator (Lopez et al. 2018) coupled with Python programming language script via

TraCI interface (Wegener et al. 2008). The simulated motorway model mainstream is 8 km long and segmented into 500 m road segments for testing and 50 m length road segments for validation, totaling 16 and 160 road segments, respectively. Different road segments' lengths are chosen to test the method's ability to adapt to different scenarios of collecting traffic data, regardless of the segment's length. The motorway model has one on-ramp and one off-ramp. The simulation lasts for two hours of traffic simulation, totaling 24 of the 5 min intervals. Data for each vehicle and road segment is collected cumulatively every second for each 5 min interval. Harmonic mean speed and density are computed for each road segment for every interval.

4.6.2 Traffic Scenarios

To evaluate the proposed method, four traffic scenarios for simulation were created as shown in Figure 4.6. It has to be noted that two cases were created using the increased on-ramp inflow scenario given in Figure 4.6b. The first scenario (Figure 4.6a) simulates the collision in one lane on the motorway. The collision starts at the 1-hour mark and lasts for 30 min. The collision is simulated in the right-most lane. The mainstream flow was set to 2400 veh/h with the increase to 2800 veh/h between the 15th and 55th minute to simulate the peak inflow of vehicles. The on-ramp flow was set to 600 veh/h with the increase to 1100 veh/h between the 15th and 55th minute to simulate the peak inflow of vehicles to the mainstream flow. The goal of this scenario is to test the method's ability to detect a bottleneck when a traffic incident occurs.

The second and third scenarios (Figure 4.6b) were created to simulate the increased inflow of on-ramp vehicles. The second scenario had the same mainstream flow as the first scenario, while the on-ramp flow was set to 600 veh/h and was increased to 1200 veh/h between the 15th and 55th minute to simulate the increased peak inflow of vehicles to the mainstream flow. The third scenario had the same mainstream flow as the first scenario, while the on-ramp flow was set to 600 veh/h and was increased to 1400 veh/h between the 15th and 55th minute to simulate a very high inflow of vehicles to the mainstream flow. The second and third scenarios show the method’s ability to detect recurrent bottlenecks due to daily commuters or similar recurrent events.

The fourth scenario simulates a high inflow of Heavy-Duty Vehicles (HDVs). The inflow of vehicles on the mainstream and on-ramp was set to be the same as for the first scenario. With this scenario, the proposed method is tested against the detection of moving bottleneck caused by slow HDVs.

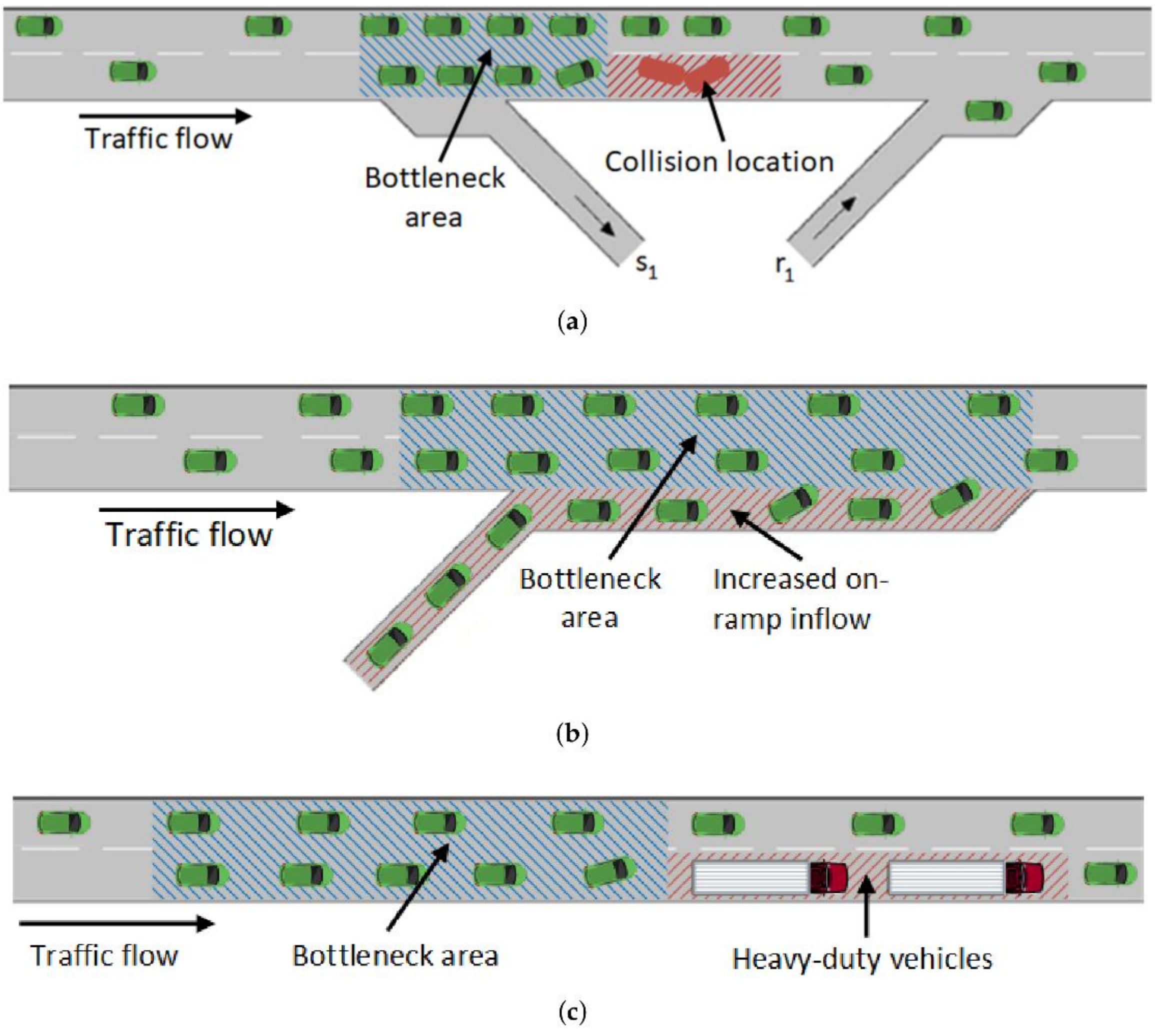


Figure 4.6 Analyzed simulation scenarios. (a) Collision scenario; (b) Increased on-ramp inflow scenario; (c) Heavy-duty vehicles scenario.

4.7 Results

The proposed bottleneck probability estimation method was evaluated on a synthetic dataset extracted from the SUMO traffic simulation framework. The method was evaluated on four different possible motorway congestion scenarios that originate from different sources of congestion, namely: traffic accident, short recurrent congestion due to high on-ramp inflow, long recurrent congestion due to high on-ramp inflow and moving bottleneck originating from a substantial amount of slow HDVs on the motorway.

4.7.1 Data

The most important parameters extracted from the simulation are vehicle speed, traffic density on the observed link, and the location of the vehicles on the motorway. Table 4.2 presents recorded parameters extracted from the simulation.

Table 4.2 Exported data from simulation scenarios.

Parameters	Speed, Density, Location
Collection frequency	1 s
Time interval length	5 min
N. Intervals	24
Simulation time	120 min
N. Vehicle routes	15,541
Motorway length	8000 m
N. Segments (test)	16 (500 m)
N. Segments (validation)	160 (50 m)

Two types of datasets were extracted from the simulation: edge data and vehicle routes data. Traffic parameters for every observed motorway road segment in the edge's dataset are collected in 5 min time intervals. For every edge, harmonic vehicle speed and density is computed. The density is computed by using the number of vehicles by an edge. STMs cannot be computed from this dataset because there is no vehicle routes information. Therefore, this dataset is used for testing and validation of the method. The second data set presents vehicles' routes recorded and collected in 5 min time intervals. The vehicle route is extracted for every vehicle on the observed motorway containing speed and location parameters. From this dataset, STMs are computed by counting the speed transitions between consecutive motorway segments.

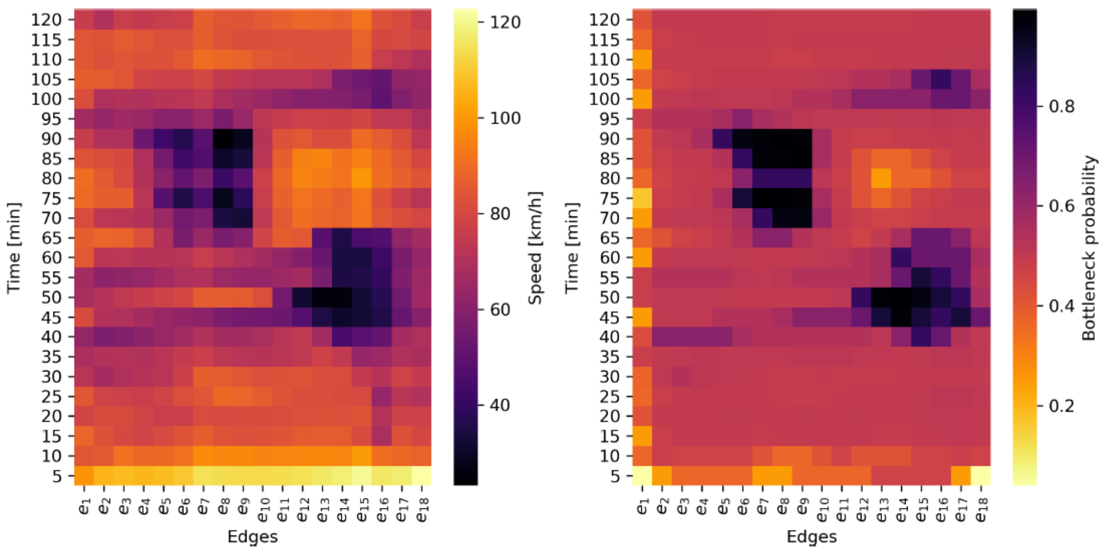
4.7.2 Bottleneck Probability Estimation

The results of the bottleneck probability estimation were compared to measured average mean speed values across every scenario. The comparison was made in the spatiotemporal domain presented in Figure 4.7. From Figure 4.7a–d, every example consists of two images: the image on the left shows the harmonic mean speed values, and the image on the right show values of computed bottleneck probability. In the temporal domain, every cell represents one 5 min time interval, and spatially, the motorway is divided into 500 m edges.

In the first example in Figure 4.7a, the bottleneck is caused by two traffic events. The first is high on-ramp inflow which started at the 15th minute and lasted until the 55th minute. The second event is a traffic accident that occurred at the 60th minute. The left image shows that the speed drops at the 15th and 65th minutes and two bottlenecks occurred at the 45th and 65th minutes. The proposed method was able to detect both congested situations that resulted in bottleneck formation.

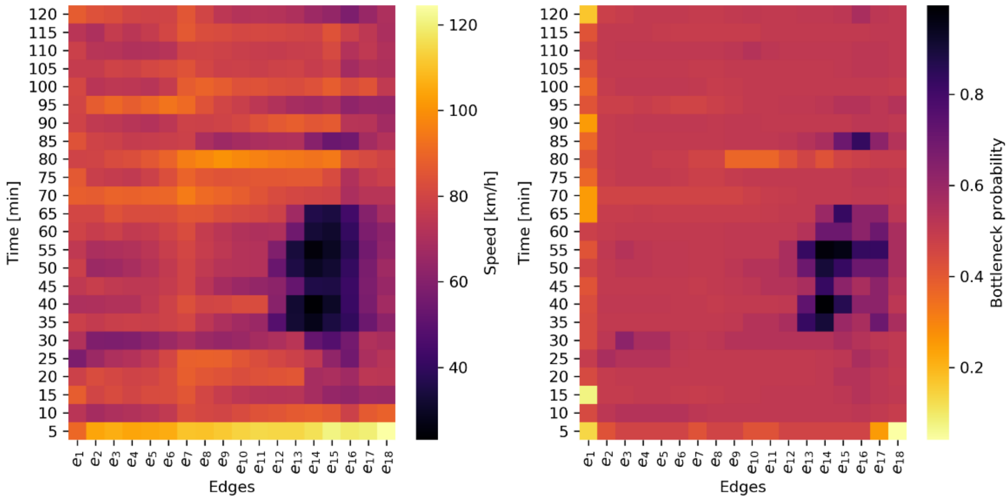
The second example in Figure 4.7b presents the bottleneck due to the short recurrent congestion caused by an increased inflow of vehicles to mainstream flow from an on-ramp. The bottleneck starts to form at the $e^{(10)}$ to $e^{(13)}$ at the 35th minute and lasts until the 65th minute of the simulation time. The speed values indicate the speed decrements even between the 10th and 15th minutes, which could be misleading for the control algorithm at the motorway. On the other hand, p_b values start to increase between the 30th and 35th minutes, which provide a more accurate time for the response of the motorway control algorithm.

The third example in Figure 4.7c presents the bottleneck caused by the very high recurrent congestion due to an increased inflow of vehicles to mainstream flow from an on-ramp. The bottleneck spans from the $e^{(9)}$ to the $e^{(17)}$ within the time interval from the 35th to the 70th minute. The pb values start to increase at the 30th minute of the simulation, especially at the $e^{(16)}$, at which point the start of the bottleneck formation can be identified.

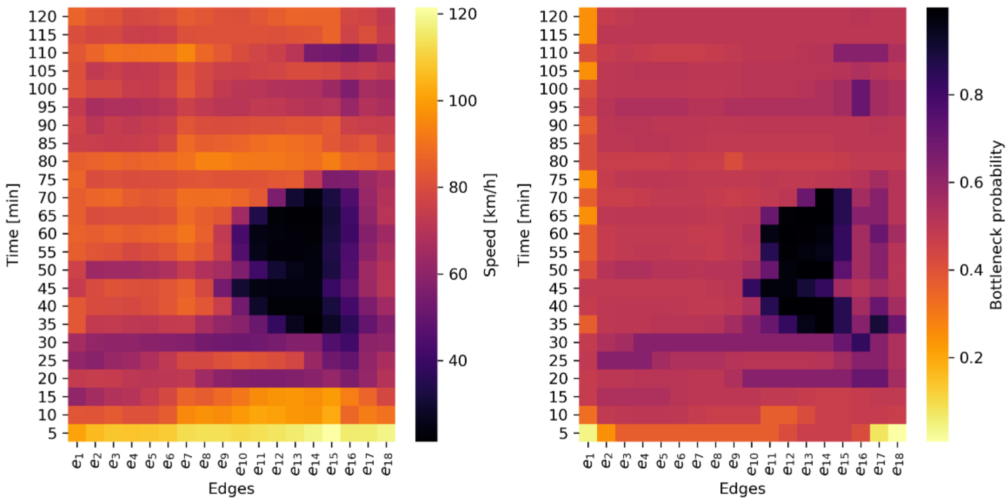


(a)

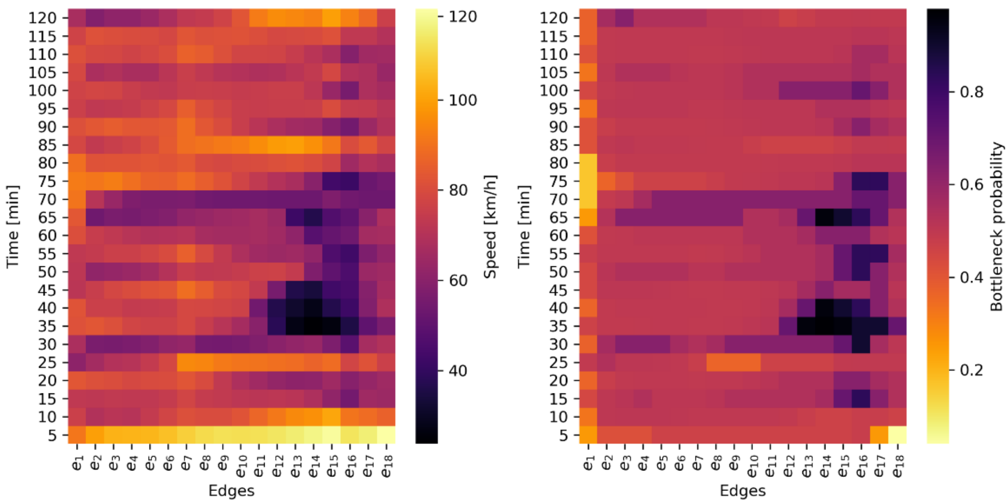
Figure 4.7 Cont.



(**b**)



(**c**)



(**d**)

Figure 4.7 Results of the comparison between absolute harmonic speed measurements (left column) and proposed bottleneck probability estimation method (right column). (a) Scenario 1—collision site; (b) Scenario 2—recurring short bottleneck; (c) Scenario 3—recurring

The fourth example in Figure 4.7d presents the moving bottleneck caused by the slow HDVs on the motorway. In the observed scenario, a very large number of HDVs entered the motorway, which resulted in a moving bottleneck that started at the 65th simulation minute. The consequences of this type of bottleneck are reflected with the increased p_b values on a large number of edges, especially at the 70th minute. The speed values confirm the claim with the speed decrease in the whole motorway after the entrance of the HDVs.

4.7.3 Validation

The first step in the validation process was to create the ground truth dataset to compare it with the proposed method. The main goal was to estimate the bottleneck probability using common traffic parameters like speed and density and find a corresponding threshold for the bottleneck detection. For the validation process, the motorway scenario representing the collision site is chosen (Figure 4.7a).

Critical density ρ_c was estimated according to (Elefteriadou 2016), where the value of 28 veh/km/lane was reported as a critical value. When the vehicles on the motorway operate under ρ_c condition, the traffic flow is maximal. If the current density is above the critical one $\rho > \rho_c$, congestion occurs, which consequentially leads to a bottleneck start. Another used traffic parameter was the critical speed v_c. Authors in (Elefteriadou 2016; Inoue, Miyashita, and Sugita 2016; Li et al. 2020; Lipan and Groza 2010; Wang et al. 2016), reports critical speed values from 50–60% of the free flow speed. In this paper, the value of 55% of the free flow speed is chosen as the critical speed on the motorway. According to the Highway Capacity Manual (Elefteriadou 2016), at this speed, serious congestion with the lowest level of service is certain.

Speed and density were measured on all observed 50 m motorway segments for the creation of the ground truth dataset. The harmonic mean speed and density were collected for each road segment. The exact values of parameters are shown in Figure 4.8a, c for the observed time period of 125 min divided into 5 min intervals where the dark color represents the low values of the speed and the high values of the density. The corresponding critical values are shown in Figure 4.8b and the binary image containing critical speed values can be represented with matrix $V \in \mathbb{R}^{\mathrm{m}\times\mathrm{n}}$. The elements of the matrix $v^{(ij)}$ are binary values expressed as:

$$V^{(ij)} = \begin{cases} 1, v^{(ij)} \leq v_c \\ 0, v^{(ij)} > v_c \end{cases} \tag{4.1}$$

where m represents the number of time intervals, and n represents the number of motorway segments.

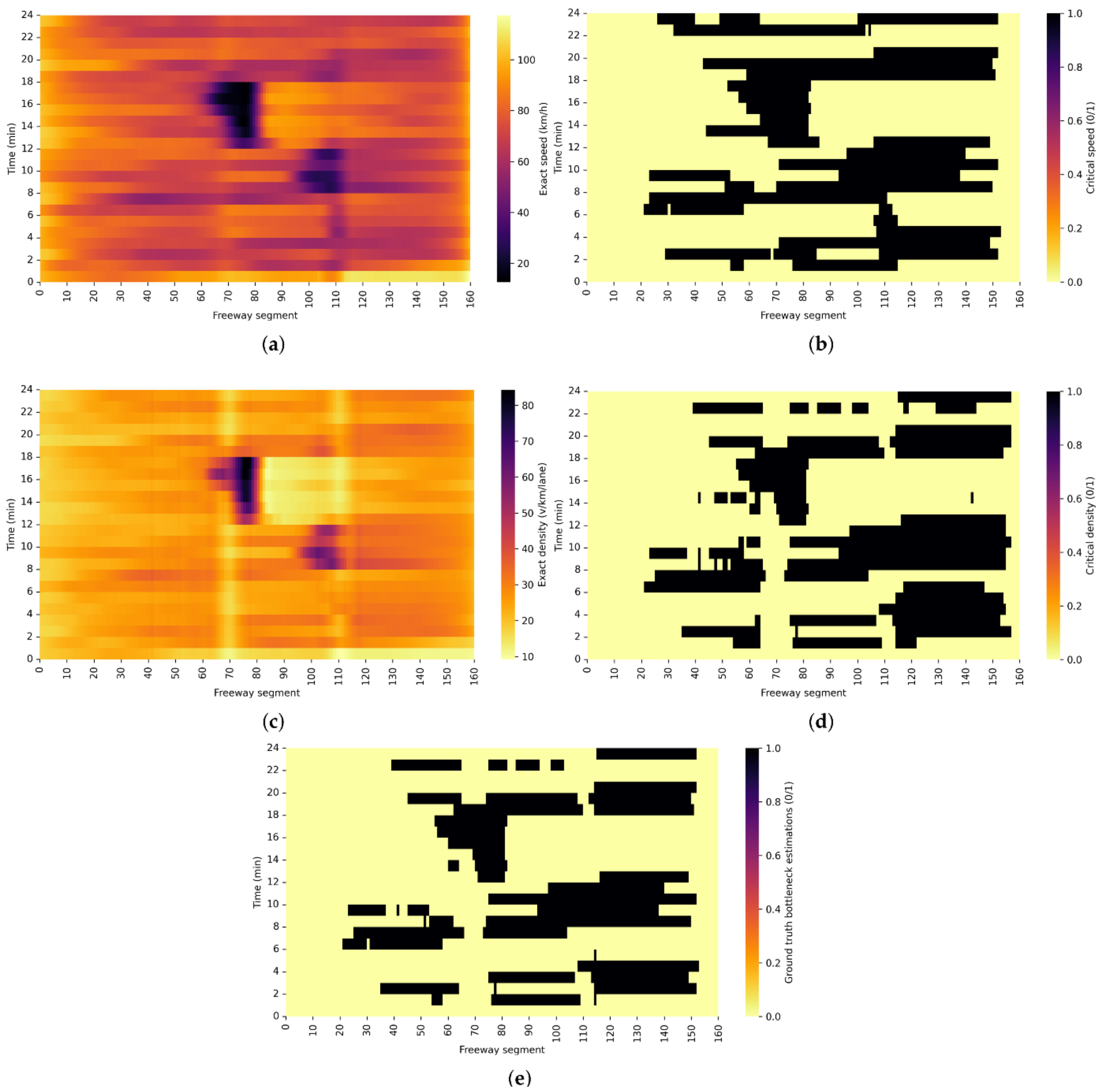


Figure 4.8 Ground truth data creation process for validation of the proposed method. (a) Exact values of the speed measurement; (b) Binary image where 1 represents critical speed; (c) Exact values of the density measurement; (d) Binary image where 1 represents critical density; (e) Intersection of critical speed and density values.

On the other hand, the binary image containing critical densities (Figure 4.8d) can be represented with matrix $D \in \mathbb{R}^{m \times n}$, where elements of the matrix $\rho^{(ij)}$ are also binary values expressed as:

$$D^{(ij)} = \begin{cases} 1, \rho^{(ij)} \leq \rho_c \\ 0, \rho^{(ij)} > \rho_c \end{cases} \tag{4.2}$$

Finally, Figure 4.8e represents the intersection of the critical values $P_{eval} = V \cap D = \{x : x \in V, x \in D\}$. The intersection of critical values of speed and density P_{eval} represent the bottleneck occurrence on the observed motorways. There, variable x will contain values 0 if a bottleneck is not detected, and 1 if a bottleneck is detected. The matrix P_{eval} is further used as a ground truth data.

The next challenge was to adopt the results of the proposed method for the comparison with the ground truth data. As the result of a bottleneck probability estimation is a decimal number in the range [0,1]. The probability threshold for the bottleneck detection must be defined to discretize the values to 1 if the bottleneck was detected and 0 if not. Table 4.3 shows the result of the threshold estimation. Estimating the probability threshold was conducted by changing the threshold values α from 10% to 90% and reporting the precision, recall, and F1-score between ground truth data and discretized the proposed method. It can be observed that the threshold of 50% gives the best results with an accuracy score of 0.92. Formally, the result of the proposed method can be represented by the matrix $P_{prop} \in \mathbb{R}^{m \times n}$ with the values expressed as:

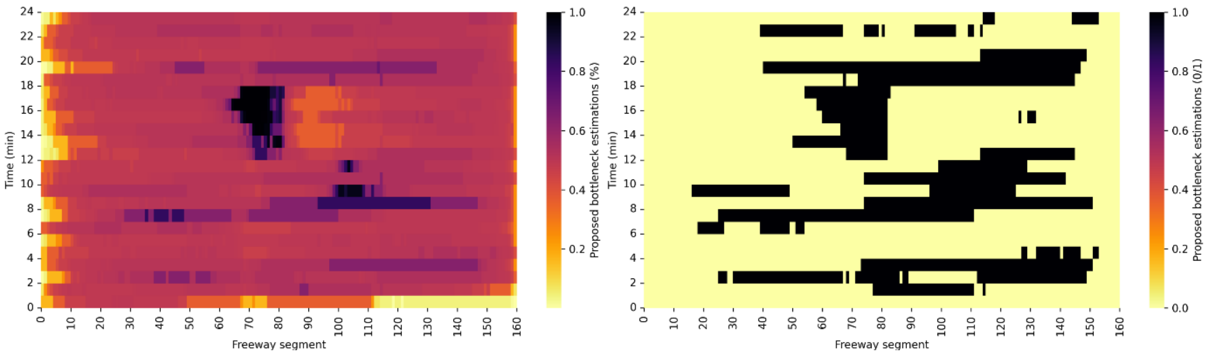
$$P_{prop}^{(ij)} = \begin{cases} 1, p_{bot}^{(ij)} \leq \alpha \\ 0, p_{bot}^{(ij)} > \alpha \end{cases} \tag{4.3}$$

where $p_{bot}^{(ij)}$ represents the bottleneck probability and α represents defined threshold value for the discretization.

Table 4.3 Validation of the threshold for the proposed bottleneck probability estimation method.

Threshold	Class	Precision	Recall	F1-Score	Accuracy
10%	Normal	1.00	0.03	0.06	0.29
	Bottleneck	0.28	1.00	0.44	
20%	Normal	1.00	0.05	0.10	0.31
	Bottleneck	0.28	1.00	0.44	
30%	Normal	1.00	0.06	0.11	0.32
	Bottleneck	0.28	1.00	0.44	
40%	Normal	1.00	0.14	0.24	0.37
	Bottleneck	0.30	1.00	0.46	
50%	Normal	0.94	0.95	0.94	**0.92**
	Bottleneck	0.86	0.84	0.85	
60%	Normal	0.80	0.99	0.89	0.82
	Bottleneck	0.96	0.36	0.52	
70%	Normal	0.75	1.00	0.86	0.76
	Bottleneck	1.00	0.13	0.23	
80%	Normal	0.75	1.00	0.86	0.76
	Bottleneck	1.00	0.12	0.22	
90%	Normal	0.74	1.00	0.85	0.74
	Bottleneck	1.00	0.05	0.09	

With the defined threshold, any value of the bottleneck probability greater than 50% will be considered a detected bottleneck and represented with the value of 1. Bottleneck probabilities for the observed validation scenario are presented in Figure 4.9a with the corresponding binary values in Figure 4.9b.



(a) (b)

Figure 4.9 Proposed method for the bottleneck probability. (a) Estimated bottleneck probability exact values; (b) Binary image where 1 represent bottleneck.

4.8 Discussion

This paper presents a novel motorway bottleneck detection approach based on the STM-based traffic data model. This section aims to describe the method's application possibilities and address the disadvantages and considerations for using the method. The emphasis of the proposed method is given on the impact of future academic research related to traffic control strategies on urban motorways and the usage of artificial intelligence methods.

Alongside the advantages and presented application of the proposed method, some considerations must be addressed. The method is evaluated on a dataset produced with simulated CVs on the motorway. Currently, all vehicles are not connected and integrated into one communication system. Accordingly, the method could be evaluated on a dataset, which contains mixed traffic flow (containing CVs and not connected vehicles) to be applicable to the current motorway scenarios. On the other hand, speed is used as the only traffic parameter for bottleneck detection. Introduction of several more transition matrices containing parameters like density or traffic flow could increase the accuracy and reliability of the method's result in more versatile motorway traffic scenarios. Final consideration when using STMs is the trade-of between the time interval of data collection for one matrix Δt and the traffic pattern accuracy, represented by the STM. The Δt depends on the type of the research that is conducted. For macroscopic considerations, like traffic state estimation or finding the anomalies in large datasets, intervals can be set to wider range, with intervals of one hour or larger. In this case, many different traffic patterns are captured by the STM, but the most dominant can be extracted. On the other hand, if the research is related to micro scale like traffic control on highway or intersections, Δt should be set to a minute scale. In this paper, the interval was set to 5 min, which was used for STM data collection, and is appropriate for bottleneck detection on motorways.

4.8.1 Features for Bottleneck Probability Estimation

The STM represents the speed change in one spatial transition observed on the motorway. In the process of recording the speed transition, a spatiotemporal correlation between consecutive road segments in the form of the speed pair is captured. In this way, the speed transition will show the weighted average of the speed pairs rather than a simple averaging technique. The proposed STM-based approach can be compared with the methods that rely on a combination of multiple parameters to get more correct values by combining two or more parameters or sensor values like Kalman filter or Hough transform.

Extracted attributes d_S and d_D represent the positional information for the pattern extracted from the STM with the additional information related to two important research topics in the field of ITS, traffic state estimation and anomaly detection. The value of the d_S represents the traffic state with values in the range [0,1], where 0 represents congested traffic state, and 1 represents the free-flow traffic. On the other hand, d_D addresses the anomaly measure in the traffic data with the values range in [0,1] where the 0 represents normal traffic behaviour with d_S as an only important attribute, and 1 represents the anomaly in the observed traffic pattern. Combining those two attributes provides complete insight into the traffic state. It allows one to analyze the traffic behavior with one or both features, depending on the considered use case.

4.8.2 Motorway Traffic Control Strategies

There are two most researched motorway traffic control systems, namely, Variable Speed Limit (VSL) (Zhang et al. 2013) and Ramp Metering (RM) (Papageorgiou, Hadj-Salem, and Middelham 1997). The goal of both control systems is to harmonize traffic flow on the main flow at the motorway to prevent bottleneck occurrence, reduce congestion, decrease travel time, and decrease pollution. VSL consists of traffic signs positioned along the motorway that show the current speed limit to the drivers encountering some traffic problem at the downstream traffic flow. On the other hand, RM consists of a traffic light at the motorway on-ramp and regulates the number of vehicles that can access the main traffic flow.

Recent studies (Gregurić, Mandžuka, and Vujić 2020; Pan et al. 2021) are proposing the usage of a virtual VSL suited for CVs that will receive speed limit information directly to the car dashboard. The STM-based traffic data modelling approach provides a more insightful data model as an input to both motorway traffic control systems. The bottleneck detection method proposed in this paper presents insights into the traffic state at the downstream traffic flow. It could provide actionable information to adapt the speed limits to harmonize traffic flow. The second important information extracted by the proposed bottleneck detection approach is the bottleneck length and duration propagation through spatial and temporal components. The propagation can be simply captured and implemented by counting the number of cells inflicted by the bottleneck in the spatial and temporal domain, respectively.

4.8.3 Reinforcement Learning Methods

The authors in (Vrbanić et al. 2021) summarized different reinforcement learning methods to VSL on the motorways in mixed traffic flows that were used to improve the performance of the control system. The emphasis of research in (Müller, Carlson, and Kraus 2016; Vinitsky et al. 2018; Vrbanic, Ivanjko, et al. 2021) is using Connected Autonomous Vehicles (CAVs) as the mobile sensors that provide data for cooperative VSL, which is used as a speed control system on a motorway. The main goal was to use CAVs to maximize the mainstream traffic flow for reducing the delay time of vehicles by adjusting motorway speed limits with the appliance of appropriate speed limits. Apart from using the CAVs as actuators, they can be used for state estimation when dealing with Q-learning (Vrbanic, Ivanjko, et al. 2021). By doing so, the classical traffic measurements such as speed, flow, and density can be replaced with the state estimation of bottleneck probability on a sequence of small segments on a motorway.

Thus, using the bottleneck probability as an input to the Q-learning could potentially increase the performance of VSL and allow finer calibration of the VSL system and improve performance.

4.9 Conclusions

Emerging technologies like CVs present challenges for developing new methods for traffic data analysis, traffic state estimation, and control strategies. Thus, CVs can be mobile sensors and actuators. In this paper, the method for motorway traffic bottlenecks probability estimation is proposed. The method is based on the STM computation and extraction of the represented traffic parameter position as an input to the fuzzy-based algorithm. The result of the method was the bottleneck probability estimated on the simulated motorway traffic scenarios. The method showed comparable results on four different traffic scenarios including traffic incident, recurrent traffic, and the moving bottleneck, with a validation accuracy of over 92%. The proposed method can be used in an environment with a high penetration rate of CVs and as an input to the motorway traffic control algorithms.

Future work relates to the implementation of the STM-based bottleneck detection method as an input to classical motorway traffic control systems like VSL. The bottleneck probability and the bottleneck length will be used as a reward function for training the reinforcement learning algorithm for controlling the speed limits at VSL signs. Here, the agent will be rewarded if the bottleneck length decreases, or the bottleneck probability drops due to a more efficient control strategy.

Chapter 5

Joint discussion

Increased traffic demand inevitably leads to more innovations in the traffic management and control domain to cope with the challenges of the increased need for human and goods mobility. The main driver of the research and innovations is coping with congestion as the greatest challenge related to increased traffic demand. The first step in providing efficient solutions to the congestion challenge is to estimate traffic state to detect the congested traffic flow. After successful identification of the congestion, appropriate measures can be applied to reduce the congestion and its negative effects to overall traffic and quality of life.

This thesis presents the novel matrix-based traffic data representation, the STM, as a tool which can be used for building a variety of methods for traffic state estimation. Furthermore, this thesis's main objective and the contributions are related to the development of STM-based methods for traffic state estimation which can be used for traffic congestion estimation, traffic anomaly detection, and bottlenecks detection. The primary aim of introducing the STM is to provide traffic data representation that is more informative because it does not discard many datapoints in the pre-processing step than commonly used speed profiles and it is less complex than traffic tensors.

As the matrix-based representation, STM can be visualized as a heatmap representing the traffic pattern. During the research it was observed that the position of the traffic pattern is the most important indicator for estimating the traffic state. Furthermore, the CoM emerged as the main feature which can be extracted from the STM and with further analysis used for the development of traffic state estimation methods. This chapter discusses three main contributions of the thesis, which include traffic congestion estimation, traffic anomaly detection, and motorway bottleneck detection, and combines each paper's contributions into the single discussion.

This research discusses the development of the clustering congestion estimation method based on the STM. The main goal of the research was to present the novel traffic data representation method, compare it to commonly used traffic data representation methods, and present the first STM use case, the congestion estimation. The method included computation of the CoM represented with equations (2.4) and (2.5). The computation of CoM resulted in transformation from the STM (matrix) to the one point with its coordinates c_x and c_y within the STM representing observed traffic pattern's center of mass. The transition from the STM into the CoM representation is shown in Figure 2.3. Due to its characteristic of simply, yet effectively representing the observed traffic pattern, the CoM is used in further research as the main feature extracted from the STM which is used for the traffic state estimation. After the CoM extraction for every computed STM, an Elbow method is used to determine the cluster number which represents the congestion states. The result showed the best possible cluster number of three congestion states. Then the agglomerative clustering approach is used to divide input CoMs into the three congestion classes "Free flow", "Stable flow", and "Congestion". This approach is visualized in Figure 2.4 which presents the result of the Elbow method and the corresponding dendrogram representing clustering result. The method was applied on the real-life dataset collected at the medium sized European city described in Table 3.2. The result in Table 2.1Figure 2.5 Results of traffic state estimation visualized on the map of the City of Zagreb. and Figure 2.5 indicate the three estimated congestion states with the corresponding probabilities of the occurrence spanned through different daily intervals. The results indicate the greatest probability of congestion during the rush hours due to daily commuting. After the validation of method using cross validation (Table 2.2 and Table 2.3) and the domain knowledge (Table 2.4 and Table 2.5) with the corresponding accuracy of 97% and 91%, the results are further used for the analysis of most congested bridges at the City of Zagreb. The results showed that congestion is generally very high, and traffic behaves very differently at bridges when morning and evening rush hours are compared. This research resulted in successfully applying the STM-based method for clustering traffic data which resulted in identifying three congestion classes. In the other hand, it showed the advantages of using the more informative traffic data representation with using the STM to extract the spatiotemporal correlations within the observed road segments. Consequently, this research places STM-based clustering as a method with a clear applicable value for further research or implementation in industry applications.

This research continued from identifying clusters in traffic data into detection of unexpected events, the anomaly detection. The second goal of the research was to develop an STM-based method for anomaly detection in traffic datasets. The proposed methodology includes data pre-processing steps and the anomaly detection method by identifying anomalies from a constructed tensor. In the pre-processing step, the observed area (medium sized European city) was divided into N cells with the physical size of 500×500 m. One cell was then used to extract STMs which are then used to construct a traffic tensor. In this case, every cell was represented with a tensor $\tau \in \mathbb{R}^{m \times n \times t}$ further explained in Figure 3.4. With the constructed tensor, traffic can be modeled, and the traffic state can be estimated by analysis of the factor matrices which were the result of the tensor decomposition represented with the Equation (3.1). The main anomaly detection input matrix was the factor matrix $A \in \mathbb{R}^{400 \times 10}$ which consisted of the characteristic traffic patterns represented with the STMs. On the other hand, factor matrices $B \in \mathbb{R}^{n \times 10}$ and $C \in \mathbb{R}^{8 \times 10}$, where used for spatial and temporal anomaly placement, respectively. After computing factor matrices, a novel anomaly detection method based on the CoM was proposed. The method relies on the computation of the d_A, the relative distance from the CoM and the diagonal of the matrix which spans from the upper left to the lower right corner, represented in Figure 3.1 and Figure 3.4. In Figure 3.6 a) all CoMs relative distances from the diagonal was plotted in the histogram. It showed significant right skew because most of the CoMs were very close to the diagonal as it represents the normal traffic. Due to the intense skew Figure 3.6 b) and Table 3.1 showed that standard anomaly detection methods are not able to detect the STM anomalies efficiently. This is the reason adjusted box plot is used to detect the anomalies as it can detect the anomalies from the skewed distributions. The research resulted in proposing an STM-based anomaly for anomaly detection represented with Algorithm 3.1. With this research, a novel paradigm for detecting anomalies in traffic networks using the STM is proposed. The approach defines an anomalous traffic state as an instance of unexpected traffic flow behavior. By specifically targeting sudden braking and intense accelerations, the method effectively avoids misclassifying recurrent congestion as anomalies. Many existing methods for road traffic anomaly detection rely on identifying large deviations in traffic parameters within a defined time period. However, when analyzing data hour by hour, such methods may mistakenly flag recurrent traffic congestion as anomalies, as it is substantially different from the normal traffic which occurs in the non-rush hours. This is due to peak traffic loads during rush hours, which vary in intensity across different areas of the city. For instance, anomaly detection methods based solely on traffic volume computations may fail to detect anomalies within a specific time interval if the average daily traffic volume remains unchanged. Leveraging the STM, our proposed approach overcomes these challenges, ensuring accurate anomaly detection while remaining adaptable for near real-time and real-time applications. Compared to the other approaches, this method is more focused on detecting local anomalies which scope is limited to the transition between two road segments. Concretely, it is adopted for detecting the dangerous situations which can lead to traffic accidents which places this method as a tool for the prediction of the anomalous patterns. After the successful prediction of the potentially dangerous event, the necessary actions can be taken to prevent its happening.

The research continued from detecting the general anomalous patterns in the urban areas to detecting the specific anomaly events on the motorways, the bottlenecks. The bottleneck occurs on the motorway when the speed of the vehicle suddenly drops due to vehicles merging on ramps or traffic accidents and causes a shock wave in the upstream direction. As the authors discussed in (Wegerle et al. 2020), most commonly used traffic flow models based on measuring standard traffic parameters cannot be used to represent traffic flow breakdowns which causes bottlenecks without incorporating spatio-temporal correlations. In this context, novel motorway bottleneck detection method is proposed based on the STM at it shows spatio-temporal correlations between the consecutive segments. The proposed STM-based method is based on extracting the speed transitions between consecutive road segments with STM representation (Figure 4.1), extracting the CoM (Equations (2.2)-(2.5)), and measuring two distances: i) distance from the CoM to the origin of the STM, and already mentioned ii) distance from the CoM to the diagonal of the STM. Distances are then used together for the bottleneck detection because of the ability to represent the congestion state and detect the anomaly which forms the bottleneck, represented in Figure 4.4. With this method three distinct traffic events can be detected which are part of the bottleneck manifestation: (i) sudden breaks, as the start of a bottleneck where vehicles are approaching the congested area, (ii) heavy congested area, where vehicles are slowed down or not moving, and (iii) sudden acceleration area, where vehicles are leaving the congested area. Figure 4.3 shows all three traffic events with the corresponding STMs used for the bottleneck detection. This method differs from the proposed methods in other literature due to its ability to estimate the probability of bottleneck occurrence which can be used for the bottleneck prediction. This can be used as actionable information for the authorities to prevent its negative effects on the traffic flow on motorways. The second differentiation is the ability to use the method in both real-time and history dataset analysis. After the initial method calibration on the historical dataset by setting the fuzzy rules represented in Table 4.1, and setting the threshold for the bottleneck detection shown in Table 4.3 the method can be efficiently used in a real-time scenarios and simulations. The third distinctive feature of the bottleneck detection method is the ability to detect the bottleneck length which is shown in Figure 4.7 represented using heatmap. The bottleneck length can be estimated by counting the transitions represented with the STMs from following the three mentioned bottleneck events from the sudden breaks to the sudden acceleration area.

There is a wide range of possible STM usages related to scientific contributions and industry applications like visualizing, quantifying, and classifying traffic state, identifying anomalous behavior, and predicting traffic flow. One of the possible use cases is expanding the STM to traffic state prediction problems by extracting traffic flow-related features captured in STMs traffic patterns. Where multiple features from the traffic pattern can be extracted related to the traffic flow speed, density, or some other external features which are influencing the traffic state. Another possible use case can include extraction of STMs as traffic images for training deep learning models to predict future traffic flow state. Here, every traffic image can represent the traffic state on the consecutive road segments on the observed part of the traffic network. Since every STM represents the traffic pattern, it can be used as an input image for training the deep learning model like CNN. The model can be trained to extract the visual features and predict the future traffic behavior or to classify the traffic state into pre-defined traffic state classes. Furthermore, the STM-based traffic state prediction can be incorporated into the routing application to find the optimal route due to its primary feature of containing traffic state and anomaly information. In this use case, STM is used to represent spatiotemporal relations between congestion at the consecutive road segments which are a part of the route. Here, STM is used as a data source for estimating traffic congestion, which will determine the optimal route based on the chosen loss function like fastest route, the most energy effective route, or similar goal.

Alongside the advantages and presented application of the proposed method, there are several limitations that need to be addressed in further research. One limitation is that the traffic state value is only based on speed values, which can lead to inaccuracies on short road segments bounded by unsynchronized traffic lights. Another limitation is that the dataset used for the experiment only includes working days to capture the most extreme congestion conditions in the urban road network. A possible improvement would be to include weekend data to analyze the differences between traffic on working days and weekends. Additionally, the anomaly detection method is based on speed, which can lead to false anomaly detection on short road segments bounded by non-synchronized intersections. More narrow time intervals like 5, 15 or 30 minutes are used in most road traffic-related research. Narrower time intervals could provide more informative results, especially in environments with mixed traffic flows. The modeling process of the STM also needs to be addressed, with important parameters like size, discretization period, and cell size analyzed for optimization purposes. Furthermore, the method should be evaluated on a dataset that contains mixed traffic flow (containing CVs and not connected vehicles) to be applicable to current motorway scenarios. Speed is used as the only traffic parameter for bottleneck detection and incorporating more transition matrices containing parameters like density or traffic flow could increase the accuracy and reliability of the method in more versatile motorway traffic scenarios. Finally, the trade-off between the time interval of data collection and the traffic pattern accuracy represented by the STM needs to be considered. The time interval depends on the type of research being conducted and should be set to a wide range for macroscopic considerations and a narrow range for micro-scale considerations such as traffic control on highways or intersections.

Chapter 6

Conclusion

Within this thesis, three methods, all related to traffic flow state estimation are represented. The novelty of the methods is reflected in using newly proposed GNSS traffic data representation named STM. Methods represent the joint contribution to the transport and traffic research field by showing the STM usage possibilities as a novel traffic data representation. Three methods are presented to the scientific community by publishing journal papers in journals raked according to the Ordinance on doctoral studies at the University of Zagreb and the recommendations of the Commission for Postgraduate Studies and Doctorates of the Faculty of Transport science.

Based on the results of this thesis, the following conclusions and answers to the hypotheses can be made:

1. Speed transition matrix can be used for a development of the traffic congestion classification method on a road traffic network.

The results of this research present the methodology for traffic congestion estimation and classification on a citywide scale based on the medium-sized European city. The CoM for every matrix, as a most important feature representing the position of the traffic pattern, was extracted to classify the STMs regarding the traffic congestion with classes "free traffic flow", "unstable traffic flow" and "congestion", respectively. The CoM extraction resulted in simplification of the classification process due to dimensionality reduction and higher interpretability of the resulting classes. The results show that STMs can be used for traffic congestion estimation on a citywide scale and on micro-locations. The results are validated using the cross-validation method, and specific domain knowledge, which resulted in an accuracy of 97% and 91%, respectively.

__2. Speed transition matrix can be used for a development of the tensor-decomposition-based method for anomaly detection on a road network.__

This thesis presents a novel tensor-based method that is using STMs as an input for anomaly detection. The main confirmed hypothesis within this research was that CoM position within the STM can represent the anomaly efficiently. The research showed that the distance between the main diagonal and the CoM can be used as an anomaly detection feature when working with STMs. This method integrates the advantages of using tensor-based data representation with the novel STM-based approach to extract the traffic patterns, which can be described as anomalous or not based on the distance between the CoM and the STM's diagonal. The result presents valuable traffic insights useful for the routing application, responsible urban planners, and road infrastructure maintenance authorities. When compared to other anomaly detection methods that uses tensor-based representation, this method is more focused on detecting the local anomalies that strictly relates to traffic flow and can lead to the development of the potentially dangerous situations like traffic accidents.

__3. Speed transition matrix can be used for a development of the method for traffic bottleneck detection on motorways and probability of their occurrences.__

This thesis presents a novel method for detecting and estimating motorway bottleneck occurrences. The method is based on using STMs as an input, while traffic bottleneck features were extracted from CoM, then fed to the FIS to estimate the bottleneck's probability of occurrence. The main bottleneck detection features were related to measuring the distance between the CoM and the STM's source and the main diagonal, respectively. The method was validated using a traffic simulator with different scenarios including traffic incident, recurrent traffic, and the moving bottleneck, with a validation accuracy of over 92%. The proposed method can be used in an environment with high penetration rate of CVs and as an input to the motorway traffic control algorithms.

Further research directions for the academic community in the field of traffic state estimation and classification include expanding upon the use of STMs as a data model. One potential direction includes training a deep learning model, such as a Convolutional Neural Network (CNN), as a traffic state classifier, using STMs as input data. Another direction is to use tensor-based analysis, by creating traffic tensors as multiple STMs placed in a tensor-based on the time interval in which the STM is collected, which could provide more spatiotemporal insight into traffic conditions. Additionally, future work could involve expanding the proposed method for real-time anomaly detection, as well as implementing STM-based bottleneck detection as an input to classical motorway traffic control systems like Variable Speed Limits (VSL), where the bottleneck probability and length are used as a reward function for training a reinforcement learning algorithm to control speed limits at VSL signs. The agent will be rewarded if the bottleneck length decreases or the bottleneck probability drops, indicating a more efficient control strategy.

Bibliography

Afrin, Tanzina, and Nita Yodo. 2020. "A Survey of Road Traffic Congestion Measures towards a Sustainable and Resilient Transportation System." *Sustainability* 12(11):4660. doi: 10.3390/su12114660.

Anon. 2010. *HCM 2010 : Highway Capacity Manual*. 5th ed. Washington, D.C.: Transportation

Research Board, National Research Council.

Barbieri, Luca, Stefano Savazzi, Mattia Brambilla, and Monica Nicoli. 2022. "Decentralized Federated Learning for Extended Sensing in 6G Connected Vehicles." *Vehicular Communications* 33. doi: 10.1016/J.VEHCOM.2021.100396.

Bro, Rasmus, and Henk A. L. Kiers. 2003. "A New Efficient Method for Determining the Number of Components in PARAFAC Models." *Journal of Chemometrics* 17(5):274–86. doi: 10.1002/CEM.801.

Capparuccini, David Mario, Ardeshir Faghri, Abishai Polus, and Robert E. Suarez. 2008. "Fluctuation and Seasonality of Hourly Traffic and Accuracy of Design Hourly Volume Estimates." *Transportation Research Record* (2049):63–70. doi: 10.3141/2049-08.

Carić, T., T. Erdelić, J. Fosin, M. Matulin, M. Erdelić, L. Rožić, and A. Milošević. 2016. *Technical Report of Project SORDITO*. Zagreb.

Carić, Tonči, and Juraj Fosin. 2020. "Using Congestion Zones for Solving the Time Dependent Vehicle Routing Problem." *Promet - Traffic - Traffico* 32(1):25–38. doi: 10.7307/PTT.V32I1.3296.

Chandola, Varun. 2009. "Anomaly Detection : A Survey." *ACM Comput. Surv.* 41(3):1–58.

Chen, Longbiao, Jérémie Jakubowicz, Dingqi Yang, Daqing Zhang, and Gang Pan. 2017. "Fine-Grained Urban Event Detection and Characterization Based on Tensor Cofactorization." *IEEE Transactions on Human-Machine Systems* 47(3):380–91. doi: 10.1109/THMS.2016.2596103.

Chen, Shuyan, Wei Wang, and Henk van Zuylen. 2010. "A Comparison of Outlier Detection Algorithms for ITS Data." *Expert Systems with Applications* 37(2):1169–78. doi: 10.1016/J.ESWA.2009.06.008.

Chen, Xinyu, Zhaocheng He, Yixian Chen, Yuhuan Lu, and Jiawei Wang. 2019. "Missing Traffic Data Imputation and Pattern Discovery with a Bayesian Augmented Tensor Factorization Model." *Transportation Research Part C: Emerging Technologies* 104:66–77. doi: 10.1016/J.TRC.2019.03.003.

Chow, Andy, Alex Santacreu, Ioannis Tsapakis, Garavig Tanasaranond, and Tao Cheng. 2014. "Empirical Assessment of Urban Traffic Congestion." *Journal of Advanced Transportation* 48(8):1000–1016.

Chowdhury, Debashish, Ludger Santen, and Andreas Schadschneider. 2000. "Statistical Physics of Vehicular Traffic and Some Related Systems." *Physics Reports* 329(4–6):199–329. doi: 10.1016/S0370-1573(99)00117-9.

Coifman, Benjamin, and Seoungbum Kim. 2011. "Extended Bottlenecks, the Fundamental Relationship, and Capacity Drop on Freeways." *Procedia - Social and Behavioral Sciences* 17:44–57. doi: 10.1016/J.SBSPRO.2011.04.507.

Daganzo, Carlos F. 1994. "The Cell Transmission Model: A Dynamic Representation of Highway Traffic Consistent with the Hydrodynamic Theory." *Transportation Research Part B* 28(4):269–87. doi: 10.1016/0191-2615(94)90002-7.

Djenouri, Youcef, Asma Belhadi, Jerry Chun Wei Lin, Djamel Djenouri, and Alberto Cano. 2019. "A Survey on Urban Traffic Anomalies Detection Algorithms." *IEEE Access* 7:12192–205. doi: 10.1109/ACCESS.2019.2893124.

Dülgar, Yildirim, Sven Eric Molzahn, Hubert Rehborn, Micha Koller, Boris S. Kerner, Dominik Wegerle, Michael Schreckenberg, Michael Menth, and Sergey L. Klenov. 2020. "Empirical Random Phase Transitions between Free Flow and Synchronized Flow at Highway Bottlenecks." *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations* 24(6):539–55. doi: 10.1080/15472450.2019.1615488.

Elefteriadou, Lily A. 2016. "Highway Capacity Manual 6th Edition." *Highway Capacity Manual 6th Edition*. doi: 10.17226/24798.

Erdelić, Martina, Tonči Carić, Edouard Ivanjko, and Niko Jelušić. 2019. "Classification of Travel Modes Using Streaming GNSS Data." Pp. 209–16 in *Transportation Research Procedia*. Vol. 40. Elsevier B.V.

Erdelić, T., M. Ravlić, and T. Carić. 2016. "Travel Time Prediction Using Speed Profiles for Road Network of Croatia." Pp. 97–100 in *2016 International Symposium ELMAR*. Zadar, Croatia.

Erdelić, Tomislav, Tonči Carić, Martina Erdelić, Leo Tišljarić, Ana Turković, and Niko Jelušić. 2021. "Estimating Congestion Zones and Travel Time Indexes Based on the Floating Car Data." *Computers, Environment and Urban Systems* 87. doi: 10.1016/J.COMPENVURBSYS.2021.101604.

Erdelić, Tomislav, and Martina Ravlić. 2016. "Sordito: System for Route Optimization in Dynamic Transport Environment." *Promet - Traffic - Traffico* 28(2):193–94. doi: 10.7307/ptt.v28i2.2145.

European Commission. 2017. "Sustainable Urban Mobility: European Policy, Practice and Solutions."

Fanaee-T, Hadi, and João Gama. 2015. "EigenEvent: An Algorithm for Event Detection from Complex Data Streams in Syndromic Surveillance." *Intelligent Data Analysis* 19(3):597–616. doi: 10.3233/IDA-150734.

Fanaee-T, Hadi, and João Gama. 2016. "Event Detection from Traffic Tensors: A Hybrid Model." *Neurocomputing* 203:22–33. doi: 10.1016/J.NEUCOM.2016.04.006.

Feng, Zhenni, and Yanmin Zhu. 2016. "A Survey on Trajectory Data Mining: Techniques and Applications." *IEEE Access* 4:2056–67. doi: 10.1109/ACCESS.2016.2553681.

Fernandes, Sofia, Hadi Fanaee-T, João Gama, Leo Tišljarić, and Tomislav Šmuc. 2021. "WINTENDED: WINdowed TENsor Decomposition for Densification Event Detection in Time-Evolving Networks." *Machine Learning*. doi: 10.1007/S10994-021-05979-8.

Gauvin, Laetitia, André Panisson, and Ciro Cattuto. 2014. "Detecting the Community Structure and Activity Patterns of Temporal Networks: A Non-Negative Tensor Factorization Approach." *PLoS ONE* 9(1). doi: 10.1371/JOURNAL.PONE.0086028.

Gong, Jinli, and Wen Yang. 2009. "The Traffic Bottleneck Analysis on Urban Expressway under Information Condition." *PEITS 2009 - 2009 2nd Conference on Power Electronics and Intelligent Transportation System* 1:400–403. doi: 10.1109/PEITS.2009.5406984.

Gregurić, Martin, Sadko Mandžuka, and Krešimir Vidović. 2020. "Advanced Applications for Urban Motorway Traffic Control." Pp. 395–400 in *Lecture Notes in Networks and Systems*. Vol. 76. Springer.

Gregurić, Martin, Sadko Mandžuka, and Miroslav Vujić. 2020. "Improvement of Variable Speed Limit Control Effectiveness in Context of Connected Vehicles." *Lecture Notes in Networks and Systems* 128 LNNS:560–63. doi: 10.1007/978-3-030-46817-0_65.

Gregurić, Martin, Miroslav Vujić, Charalampos Alexopoulos, and Mladen Miletić. 2020. "Application of Deep Reinforcement Learning in Traffic Signal Control: An Overview and Impact of Open Traffic Data." *Applied Sciences (Switzerland)* 10(11). doi: 10.3390/APP10114011.

Guo, Jianhua, Wei Huang, and Billy M. Williams. 2015. "Real Time Traffic Flow Outlier Detection Using Short-Term Traffic Conditional Variance Prediction." *Transportation Research Part C: Emerging Technologies* 50:160–72. doi: 10.1016/J.TRC.2014.07.005.

Gupta, Manish, Jing Gao, Charu C. Aggarwal, and Jiawei Han. 2014. "Outlier Detection for Temporal Data: A Survey." *IEEE Transactions on Knowledge and Data Engineering* 26(9):2250–67. doi: 10.1109/TKDE.2013.184.

Herrera, Juan C., and Alexandre M. Bayen. 2010. "Incorporation of Lagrangian Measurements in Freeway Traffic State Estimation." *Transportation Research Part B: Methodological* 44(4):460–81. doi: 10.1016/J.TRB.2009.10.005.

Herrera, Juan C., Daniel B. Work, Ryan Herring, Xuegang (Jeff) Ban, Quinn Jacobson, and Alexandre M. Bayen. 2010. "Evaluation of Traffic Data Obtained via GPS-Enabled Mobile Phones: The Mobile Century Field Experiment." *Transportation Research Part C: Emerging Technologies* 18(4):568–83. doi: 10.1016/J.TRC.2009.10.006.

Hubert, M., and E. Vandervieren. 2008. "An Adjusted Boxplot for Skewed Distributions." *Computational Statistics and Data Analysis* 52(12):5186–5201. doi: 10.1016/J.CSDA.2007.11.008.

Inoue, Ryo, Akihisa Miyashita, and Masatoshi Sugita. 2016. "Mining Spatio-Temporal Patterns of Congested Traffic in Urban Areas from Traffic Sensor Data." *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC* 731–36. doi: 10.1109/ITSC.2016.7795635.

Iordanidou, Georgia Roumpini, Claudio Roncoli, Ioannis Papamichail, and Markos Papageorgiou. 2015. "Feedback-Based Mainstream Traffic Flow Control for Multiple Bottlenecks on Motorways." *IEEE Transactions on Intelligent Transportation Systems* 16(2):610–21. doi: 10.1109/TITS.2014.2331985.

Jin, Xin, Dipti Srinivasan, and Ruey Long Cheu. 2001. "Classification of Freeway Traffic Patterns for Incident Detection Using Constructive Probabilistic Neural Networks." *IEEE Transactions on Neural Networks* 12(5):1173–87. doi: 10.1109/72.950145.

Jordaan, Ian J. 2005. *Decisions under Uncertainty : Probabilistic Analysis for Engineering Decisions*. Cambridge University Press.

Juhász, J. 2017. "Influence of Different Route-Choice Decision Modes." *Transportation Research Procedia* 27:246–52. doi: 10.1016/J.TRPRO.2017.12.058.

Kan, Zihan, Luliang Tang, Mei Po Kwan, Chang Ren, Dong Liu, and Qingquan Li. 2019. "Traffic Congestion Analysis at the Turn Level Using Taxis' GPS Trajectory Data." *Computers, Environment and Urban Systems* 74:229–43. doi: 10.1016/J.COMPENVURBSYS.2018.11.007.

Keler, Andreas, Linfang Ding, and Jukka M. Krisp. 2016. "Visualization of Traffic Congestion Based on Floating Taxi Data." *KN - Journal of Cartography and Geographic Information* 66(1):7–13. doi: 10.1007/BF03545180.

Kerner, Boris S. 2002. "Empirical Macroscopic Features of Spatial-Temporal Traffic Patterns at Highway Bottlenecks." *Physical Review E* 65(4):046138. doi: 10.1103/PhysRevE.65.046138.

Kerner, Boris S. 2007. "Control of Spatiotemporal Congested Traffic Patterns at Highway Bottlenecks." *IEEE Transactions on Intelligent Transportation Systems* 8(2):308–20. doi: 10.1109/TITS.2007.894192.

Kerner, Boris S., Hubert Rehborn, Mario Aleksic, and Andreas Haug. 2004. "Recognition and Tracking of Spatial-Temporal Congested Traffic Patterns on Freeways." *Transportation Research Part C: Emerging Technologies* 12(5):369–400. doi: 10.1016/J.TRC.2004.07.015.

Kim, Jiwon, and Hani S. Mahmassani. 2015. “Spatial and Temporal Characterization of Travel Patterns in a Traffic Network Using Vehicle Trajectories.” *Transportation Research Part C: Emerging Technologies* 59:375–90. doi: 10.1016/J.TRC.2015.07.010.

Kolda, Tamara G., and Brett W. Bader. 2009. “Tensor Decompositions and Applications.” *SIAM Review* 51(3):455–500. doi: 10.1137/07070111X.

Kong, Qing Jie, Qiankun Zhao, Chao Wei, and Yuncai Liu. 2013. “Efficient Traffic State Estimation for Large-Scale Urban Road Networks.” *IEEE Transactions on Intelligent Transportation Systems* 14(1):398–407. doi: 10.1109/TITS.2012.2218237.

Li, Changle, Wenwei Yue, Guoqiang Mao, and Zhigang Xu. 2020. “Congestion Propagation Based Bottleneck Identification in Urban Road Networks.” *IEEE Transactions on Vehicular Technology* 69(5):4827–41. doi: 10.1109/TVT.2020.2973404.

Li, Dawei, Min Yang, Cheng Jie Jin, Gang Ren, Xianglong Liu, and Haode Liu. 2021. “Multi-Modal Combined Route Choice Modeling in the MaaS Age Considering Generalized Path Overlapping Problem.” *IEEE Transactions on Intelligent Transportation Systems* 22(4):2430–41. doi: 10.1109/TITS.2020.3030707.

Lin, Chaoguang, Qiuhan Zhu, Shunan Guo, Zhuochen Jin, Yu Ru Lin, and Nan Cao. 2018. “Anomaly Detection in Spatiotemporal Data via Regularized Non-Negative Tensor Analysis.” *Data Mining and Knowledge Discovery* 32(4):1056–73. doi: 10.1007/S10618-018-0560-3.

Lipan, Florin, and Adrian Groza. 2010. “Mining Traffic Patterns from Public Transportation GPS Data.” *Proceedings - 2010 IEEE 6th International Conference on Intelligent Computer Communication and Processing, ICCP10* 123–26. doi: 10.1109/ICCP.2010.5606450.

Liu, Xinran, Xingwu Liu, Yuanhong Wang, Juhua Pu, and Xiangliang Zhang. 2016. “Detecting Anomaly in Traffic Flow from Road Similarity Analysis.” Pp. 92–104 in *Web-Age Information Management*, edited by B. Cui, N. Zhang, J. Xu, X. Lian, and D. Liu. Cham: Springer International Publishing.

Lopez, Pablo Alvarez, Michael Behrisch, Laura Bieker-Walz, Jakob Erdmann, Yun Pang Flotterod, Robert Hilbrich, Leonhard Lucken, Johannes Rummel, Peter Wagner, and Evamarie Wiebner. 2018. “Microscopic Traffic Simulation Using SUMO.” *IEEE Conference* *on Intelligent Transportation Systems, Proceedings, ITSC* 2018-November:2575–82. doi: 10.1109/ITSC.2018.8569938.

Lu, Ning, Nan Cheng, Ning Zhang, Xuemin Shen, and Jon W. Mark. 2014. "Connected Vehicles: Solutions and Challenges." *IEEE Internet of Things Journal* 1(4):289–99. doi: 10.1109/JIOT.2014.2327587.

Lykov, Stanislav, and Yasuo Asakura. 2020. "Anomalous Traffic Pattern Detection in Large Urban Areas: Tensor-Based Approach with Continuum Modeling of Traffic Flow." *International Journal of Intelligent Transportation Systems Research* 18(1):13–21. doi: 10.1007/S13177-018-0167-5.

Ma, Xiaolei, Zhuang Dai, Zhengbing He, Jihui Ma, Yong Wang, and Yunpeng Wang. 2017a. "Learning Traffic as Images: A Deep Convolutional Neural Network for Large-Scale Transportation Network Speed Prediction." *Sensors (Switzerland)* 17(4):1–16. doi: 10.3390/s17040818.

Ma, Xiaolei, Zhuang Dai, Zhengbing He, Jihui Ma, Yong Wang, and Yunpeng Wang. 2017b. "Learning Traffic as Images: A Deep Convolutional Neural Network for Large-Scale Transportation Network Speed Prediction." *Sensors (Switzerland)* 17(4):1–16. doi: 10.3390/s17040818.

Mandžuka, Sadko, Edouard Ivanjko, Miroslav Vujić, Pero Škorput, and Martin Gregurić. 2016. "The Use of Cooperative ITS in Urban Traffic Management." Pp. 272–88 in *Intelligent Transport Systems: Technologies and Applications*, edited by A. Perallos, U. Hernandez-Jayo, E. Onieva, and I. J. García-Zuazola. New Jersey, SAD: John Wiley & Sons.

Müller, Eduardo R., Rodrigo C. Carlson, and Werner Kraus. 2016. "Cooperative Mainstream Traffic Flow Control on Freeways." *IFAC-PapersOnLine* 49(32):89–94. doi: 10.1016/J.IFACOL.2016.12.195.

Nguyen, Hoang, Wei Liu, and Fang Chen. 2017. "Discovering Congestion Propagation Patterns in Spatio-Temporal Traffic Data." *IEEE Transactions on Big Data* 3(2):169–80. doi: 10.1109/TBDATA.2016.2587669.

Nguyen, Hong Nam, Panchamy Krishnakumari, Hai L. Vu, and Hans van Lint. 2016. "Traffic Congestion Pattern Classification Using Multi-Class SVM." Pp. 1–4 in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. Rio de Janeiro.

Nguyen, Tin T., Panchamy Krishnakumari, Simeon C. Calvert, Hai L. Vu, and Hans van Lint. 2019. "Feature Extraction and Clustering Analysis of Highway Congestion." *Transportation Research Part C: Emerging Technologies* 100(January):238–58. doi: 10.1016/j.trc.2019.01.017.

Nkoro, A. B., and Y. A. Vershinin. 2014. "Current and Future Trends in Applications of Intelligent Transport Systems on Cars and Infrastructure." *2014 17th IEEE International Conference on Intelligent Transportation Systems, ITSC 2014* 514–19. doi: 10.1109/ITSC.2014.6957741.

Ouyang, Qi, Yongbo Lv, Jihui Ma, and Jing Li. 2020. "An LSTM-Based Method Considering History and Real-Time Data for Passenger Flow Prediction." *Applied Sciences 2020, Vol. 10, Page 3788* 10(11):3788. doi: 10.3390/APP10113788.

Pan, Bei, Yu Zheng, David Wilkie, and Cyrus Shahabi. 2013. "Crowd Sensing of Traffic Anomalies Based on Human Mobility and Social Media." *GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems* 334–43. doi: 10.1145/2525314.2525343.

Pan, Pingjun, Haiyang Wang, Longyuan Li, Yongkun Wang, and Yaohui Jin. 2018. "Peak-Hour Subway Passenger Flow Forecasting: A Tensor Based Approach." *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC* 2018-November:3730–35. doi: 10.1109/ITSC.2018.8569577.

Pan, Tianlu, Renzhong Guo, William H. K. Lam, Renxin Zhong, Weixi Wang, and Biao He. 2021. "Integrated Optimal Control Strategies for Freeway Traffic Mixed with Connected Automated Vehicles: A Model-Based Reinforcement Learning Approach." *Transportation Research Part C: Emerging Technologies* 123. doi: 10.1016/J.TRC.2021.102987.

Papageorgiou, M., H. Hadj-Salem, and F. Middelham. 1997. "ALINEA Local Ramp Metering: Summary of Field Results." *Transportation Research Record* (1603):90–98. doi: 10.3141/1603-12.

Papalexakis, Evangelos E. 2016. "Automatic Unsupervised Tensor Mining with Quality Assessment." Pp. 711–19 in *Proceedings of the 2016 SIAM International Conference on Data Mining*.

Pascale, Francesco, Ennio Andrea Adinolfi, Simone Coppola, and Emanuele Santonicola. 2021. "Cybersecurity in Automotive: An Intrusion Detection System in Connected Vehicles." *Electronics (Switzerland)* 10(15). doi: 10.3390/ELECTRONICS10151765.

Prada, Miguel A., Manuel DomÍnguez, Pablo Barrientos, and Sergio GarcÍa. 2012. "Dimensionality Reduction for Damage Detection in Engineering Structures." *International Journal of Modern Physics B* 26(25). doi: 10.1142/S0217979212460046.

Pun-Cheng, Lilian S. C. 2012. "An Interactive Web-Based Public Transport Enquiry System with Real-Time Optimal Route Computation." *IEEE Transactions on Intelligent Transportation Systems* 13(2):983–88. doi: 10.1109/TITS.2011.2181501.

Qi, Geqi, Ailing Huang, Wei Guan, and Lingling Fan. 2019. "Analysis and Prediction of Regional Mobility Patterns of Bus Travellers Using Smart Card Data and Points of Interest Data." *IEEE Transactions on Intelligent Transportation Systems* 20(4):1197–1214. doi: 10.1109/TITS.2018.2840122.

Rakha, Hesham, and Aly Tawfik. 2009. "Traffic Networks: Dynamic Traffic Routing, Assignment, and Assessment." *Encyclopedia of Complexity and Systems Science* 9429–70. doi: 10.1007/978-0-387-30440-3_562.

Rao, Mohan, and K. Ramachandra Rao. 2012. "Measuring Urban Traffic Congestion – A Review." *International Journal for Traffic and Transport Engineering* 2(12):286–305. doi: 10.7708/ijtte.2012.2(4).01.

Rao, Wenming, Yao Jan Wu, Jingxin Xia, Jishun Ou, and Robert Kluger. 2018. "Origin-Destination Pattern Estimation Based on Trajectory Reconstruction Using Automatic License Plate Recognition Data." *Transportation Research Part C: Emerging Technologies* 95:29–46. doi: 10.1016/J.TRC.2018.07.002.

Rendle, Steffen. 2012. "Factorization Machines with LibFM." *ACM Transactions on Intelligent Systems and Technology* 3(3). doi: 10.1145/2168752.2168771.

Schubert, Erich, Arthur Zimek, and Hans Peter Kriegel. 2014. "Local Outlier Detection Reconsidered: A Generalized View on Locality with Applications to Spatial, Video, and Network Outlier Detection." *Data Mining and Knowledge Discovery* 28(1):190–237. doi: 10.1007/S10618-012-0300-Z.

Servos, Nikolaos, Xiaodi Liu, Michael Teucke, and Michael Freitag. 2019. "Travel Time Prediction in a Multimodal Freight Transport Relation Using Machine Learning Algorithms." *Logistics 2020, Vol. 4, Page 1* 4(1):1. doi: 10.3390/LOGISTICS4010001.

Škrinjar, Jasmina Pašagić, Borna Abramović, Lucija Bukvić, and Željko Marušić. 2020. "Managing Fuel Consumption and Emissions in the Renewed Fleet of a Transport Company." *Sustainability 2020, Vol. 12, Page 5047* 12(12):5047. doi: 10.3390/SU12125047.

Sun, Huijun, Jianjun Wu, Dan Ma, and Jiancheng Long. 2014. "Spatial Distribution Complexities of Traffic Congestion and Bottlenecks in Different Network Topologies." *Applied Mathematical Modelling* 38(2):496–505. doi: 10.1016/J.APM.2013.06.027.

Tan, Huachun, Yuankai Wu, Bin Shen, Peter J. Jin, and Bin Ran. 2016. "Short-Term Traffic Prediction Based on Dynamic Tensor Completion." *IEEE Transactions on Intelligent Transportation Systems* 17(8):2123–33. doi: 10.1109/TITS.2015.2513411.

Tan, Huachun, Zhongxing Yang, Guangdong Feng, Wuhong Wang, and Bin Ran. 2013. "Correlation Analysis for Tensor-Based Traffic Data Imputation Method." *Procedia - Social and Behavioral Sciences* 96:2611–20. doi: 10.1016/J.SBSPRO.2013.08.292.

Tan, Pang-ning, Michael Steinbach, and Vipin Kumar. 2006. "Anomaly Detection." Pp. 651–84 in *Introduction to Data Mining*. Pearson Addison Wesley.

Tang, Kun, Shuyan Chen, and Zhiyuan Liu. 2018. "Citywide Spatial-Temporal Travel Time Estimation Using Big and Sparse Trajectories." *IEEE Transactions on Intelligent Transportation Systems* 19(12):4023–34. doi: 10.1109/TITS.2018.2803085.

Thianniwet, T., and S. Phosaard. 2009. "Classification of Road Traffic Congestion Levels from GPS Data Using a Decision Tree Algorithm and Sliding Windows." in *Proceedings of the World Congress on Engineering*. London, U.K.

Tian, Xuemin, Xiaoling Zhang, Xiaogang Deng, and Sheng Chen. 2009. "Multiway Kernel Independent Component Analysis Based on Feature Samples for Batch Process Monitoring." *Neurocomputing* 72(7–9):1584–96. doi: 10.1016/J.NEUCOM.2008.09.003.

Tišljarić, L., S. Fernandes, T. Carić, and J. Gama. 2020. "Spatiotemporal Traffic Anomaly Detection on Urban Road Network Using Tensor Decomposition Method." Pp. 674–88 in *Discovery Science. Lecture Notes in Computer Science*. Vol. 12323, edited by A. Appice, G. Tsoumakas, Y. Manolopoulos, and S. Matwin. Cham: Springer.

Tišljarić, Leo, Tonči Carić, Borna Abramović, and Tomislav Fratrović. 2020. “Traffic State Estimation and Classification on Citywide Scale Using Speed Transition Matrices.” *Sustainability (Switzerland)* 12(18). doi: 10.3390/SU12187278.

Tišljarić, Leo, Sofia Fernandes, Tonči Carić, and João Gama. 2021. “Spatiotemporal Road Traffic Anomaly Detection: A Tensor-Based Approach.” *Applied Sciences 2021, Vol. 11, Page 12017* 11(24):12017. doi: 10.3390/APP112412017.

Tišljaric, Leo, Edouard Ivanjko, Zvonko Kavran, and Tonci Caric. 2021. “Fuzzy Inference System for Congestion Index Estimation Based on Speed Probability Distributions.” *Transportation Research Procedia* 55:1389–97. doi: 10.1016/J.TRPRO.2021.07.124.

United States. Federal Highway Administration. 2005. *Traffic Congestion and Reliability: Trends and Advanced Strategies for Congestion Mitigation*. doi: 10.21949/1503647.

Vinitsky, Eugene, Kanaad Parvate, Aboudy Kreidieh, Cathy Wu, and Alexandre Bayen. 2018. “Lagrangian Control through Deep-RL: Applications to Bottleneck Decongestion.” *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC* 2018-November:759–65. doi: 10.1109/ITSC.2018.8569615.

Vrbanić, Filip, Edouard Ivanjko, Krešimir Kušić, and Dino Čakija. 2021. “Variable Speed Limit and Ramp Metering for Mixed Traffic Flows: A Review and Open Questions.” *Applied Sciences (Switzerland)* 11(6). doi: 10.3390/APP11062574.

Vrbanic, Filip, Edouard Ivanjko, Sadko Mandzuka, and Mladen Miletic. 2021. “Reinforcement Learning Based Variable Speed Limit Control for Mixed Traffic Flows.” *2021 29th Mediterranean Conference on Control and Automation, MED 2021* 560–65. doi: 10.1109/MED51440.2021.9480215.

Vrbanic, Filip, Mladen Miletic, Edouard Ivanjko, and Zeljko Majstorovic. 2021. “Creating Representative Urban Motorway Traffic Scenarios: Initial Observations.” *Proceedings Elmar - International Symposium Electronics in Marine* 2021-September:183–88. doi: 10.1109/ELMAR52657.2021.9550867.

Wahle, J., O. Annen, Ch Schuster, L. Neubert, and M. Schreckenberg. 2001. “A Dynamic Route Guidance System Based on Real Traffic Data.” *European Journal of Operational Research* 131(2):302–8. doi: 10.1016/S0377-2217(00)00130-2.

Wang, Jingyuan, Fei Gao, Peng Cui, Chao Li, and Zhang Xiong. 2014. "Discovering Urban Spatio-Temporal Structure from Time-Evolving Traffic Networks." Pp. 93–104 in *Web Technologies and Applications*, edited by L. Chen, Y. Jia, T. Sellis, and G. Liu. Cham: Springer International Publishing.

Wang, X., A. Fagette, P. Sartelet, and L. Sun. 2019. "A Probabilistic Tensor Factorization Approach to Detect Anomalies in Spatiotemporal Traffic Activities." Pp. 1658–63 in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*.

Wang, Youcheng, Jian Xu, Ming Xu, Ning Zheng, Jinsheng Jiang, and Kaiwei Kong. 2016. "A Feature-Based Method for Traffic Anomaly Detection." *Proceedings of the 2nd ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics, UrbanGIS 2016*. doi: 10.1145/3007540.3007545.

Wang, Zhe, Kai Hu, Ke Xu, Baolin Yin, and Xiaowen Dong. 2012. "Structural Analysis of Network Traffic Matrix via Relaxed Principal Component Pursuit." *Computer Networks* 56(7):2049–67. doi: 10.1016/J.COMNET.2012.02.017.

Wegener, Axel, Michat Piórkowski, Maxim Raya, Horst Hellbrück, Stefan Fischer, and Jean Pierre Hubaux. 2008. "TraCI: An Interface for Coupling Road Traffic and Network Simulators." *Proceedings of the 11th Communications and Networking Simulation Symposium, CNS'08* 155–63. doi: 10.1145/1400713.1400740.

Wegerle, Dominik, Boris S. Kerner, Michael Schreckenberg, and Sergej L. Klenov. 2020. "Prediction of Moving Bottleneck through the Use of Probe Vehicles: A Simulation Approach in the Framework of Three-Phase Traffic Theory." *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations* 24(6):598–616. doi: 10.1080/15472450.2019.1652825.

Xu, Guoxia, Sheheryar Khan, Hu Zhu, Lixin Han, Michael K. Ng, and Hong Yan. 2018. "Discriminative Tracking via Supervised Tensor Learning." *Neurocomputing* 315:33–47. doi: 10.1016/J.NEUCOM.2018.05.108.

Zhang, Bin, Shuyan Chen, Yongfeng Ma, Tiezhu Li, and Kun Tang. 2020. "Analysis on Spatiotemporal Urban Mobility Based on Online Car-Hailing Data." *Journal of Transport Geography* 82. doi: 10.1016/J.JTRANGEO.2019.102568.

Zhang, Hailong, Yuankai Wu, Huachun Tan, Hanxuan Dong, Fan Ding, and Bin Ran. 2022. "Understanding and Modeling Urban Mobility Dynamics via Disentangled Representation Learning." *IEEE Transactions on Intelligent Transportation Systems* 23(3):2010–20. doi: 10.1109/TITS.2020.3030259.

Zhang, Hang, Zhibin Li, Pan Liu, Chengcheng Xu, and Hao Yu. 2013. "Control Strategy of Variable Speed Limits for Improving Traffic Efficiency at Merge Bottleneck on Freeway." *Procedia - Social and Behavioral Sciences* 96:2011–23. doi: 10.1016/J.SBSPRO.2013.08.227.

Zhang, Yicai, Min Zhao, Dihua Sun, Shi hui Wang, Shuai Huang, and Dong Chen. 2021. "Analysis of Mixed Traffic with Connected and Non-Connected Vehicles Based on Lattice Hydrodynamic Model." *Communications in Nonlinear Science and Numerical Simulation* 94. doi: 10.1016/J.CNSNS.2020.105541.

Zheng, Yu. 2015. "Trajectory Data Mining: An Overview." *ACM Transactions on Intelligent Systems and Technology* 6(3). doi: 10.1145/2743025.

Żochowska, Renata, and Grzegorz Karoń. 2016. "ITS Services Packages as a Tool for Managing Traffic Congestion in Cities." *Studies in Systems, Decision and Control* 32:81–103. doi: 10.1007/978-3-319-19150-8_3.

List of Figures

List of Tables

Nomenclature

CoM	Center of Mass
CAV	Connected Autonomous Vehicle
CV	Connected Vehicle
CORCONDIA	Core Consistency Diagnostic
CP	CANDECOMP/PARAFAC
FIS	Fuzzy Inference System
GNSS	Global Navigation Satellite System
HCM	Highway Capacity Manual
HDV	Heavy Duty Vehicle
ITS	Intelligent Transport Systems
NTD	Non-negative Tensor Decomposition
O-D	Origin-Destination
PCA	Principal Components Analysis
RM	Ramp Metering
STM	Speed Transition Matrix
SUMO	Simulation of Urban MObillity
VSL	Variable Speed Limit

Biography

Leo Tišljarić was born on 29th of September 1994. at Zagreb, Croatia. He received his B.Sc. and M.Sc. degree (title: "Analysis of Queues and Level of Service on Urban Roads Using Machine Learning Algorithms and NoSql Database") in Intelligent Transport Systems and Logistics from Faculty of Transport and Traffic Sciences, University of Zagreb in 2016. and 2018. respectively. After obtaining his B.Sc. he worked as undergraduate assistant at Department for Applied Computing. While he was on master study he participated on internships at Span Ltd. as Windows server administrator, Ericsson Nikola Tesla Ltd. on project „Smart Internet of Things Analytics" and went to international internship to Audi AG (Germany) as C# and Python developer. While he was working in Audi AG, he submitted a patent with title „Verfahren zum Verarbeiten eines technischen Notrufs sowie System zum Durchführen des Verfahrens" (no. 102018200570.2) related to methods for data analysis in E-call service. In 2016, he receives a Rector reward for his paper „Measurement of Energy Consumption of a Small Electric Vehicle with Respect to Configuration of the Terrain to Optimize Vehicle Routes" and Dean award in 2017. for best student GPA.

In September 2018, he starts to work as a Junior Researcher on the project „Datacross – Advanced methods and technologies in Data Science and Cooperative Systems ". Same year, he enrolled in the Ph.D. study „Traffic" at the Faculty of

Transport and Traffic Sciences at University of Zagreb (study adviser: prof.dr.sc. Tonči Carić). He participates in teaching on the course of Algorithms and programming, Data Mining, and Computer Science.

In February 2022, he resumes his professional career at Intis d.o.o. as an Artificial intelligence Engineer. He works on problems related to sales forecasting, geo clustering, optimization of the whole supply chain from warehouse, routing process to field operations in goods delivery procedures.

Research/professional interests: artificial intelligence, data mining, machine learning, intelligent transportation systems

List of published papers:

Patents:

Audi AG, Ingolstadt, Germany, DE Verfahren zum Verarbeiten eines telefonischen Notrufs sowie System zum Durchführen des Verfahrens, DE102018200570. (https://www.bib.irb.hr/1085156)

Book chapters:

Tišljarić, L., Ribić, F., Majstorović, Ž. & Carić, T. (2022) Speed Transition Matrix Feature Extraction for Traffic State Estimation Using Machine Learning Algorithms. In: Petrović, M., Novačko, L., Božić, D. & Rožić, T. (ed.) The Science and Development of Transport—ZIRP 2021. Cham, Springer, str. 61-74 doi:10.1007/978-3-030-97528-9_5.

Scientific journals:

Vrbanić, F., Tišljarić, L., Majstorović, Ž. & Ivanjko, E. (2023) Reinforcement Learning-Based Dynamic Zone Placement Variable Speed Limit Control for Mixed Traffic Flows Using Speed Transition Matrices for State Estimation. Machines, 11 (4), 11040479, 15 doi:10.3390/machines11040479.

Fernandes, S., Fanaee-T, H., Gama, J., Tišljarić, L. & Šmuc, T. (2023) WINTENDED: WINdowed TENsor decomposition for Densification Event Detection in time-evolving networks. Machine learning, 112 (2), 459-481 doi:10.1007/s10994-021-05979-8.

Majstorović, Ž., Tišljarić, L., Ivanjko, E. & Carić, T. (2023) Urban Traffic Signal Control under Mixed Traffic Flows: Literature Review. Applied Sciences, 13 (7), 4484, 19 doi:10.3390/app13074484.

Vrbanić, F., Miletić, M., Tišljarić, L. & Ivanjko, E. (2022) Influence of Variable Speed Limit Control on Fuel and Electric Energy Consumption, and Exhaust Gas Emissions in Mixed Traffic Flows. Sustainability, 14 (2), 932, 20 doi:10.3390/su14020932.

Tišljarić, L., Vrbanić, F., Ivanjko, E. & Carić, T. (2022) Motorway Bottleneck Probability Estimation in Connected Vehicles Environment Using Speed Transition Matrices. Sensors, 22 (7), 2807, 20 doi:10.3390/s22072807.

Erdelić, M., Carić, T., Erdelić, T. & Tišljarić, L. (2022) Transition State Matrices Approach for Trajectory Segmentation Based on Transport Mode Change Criteria. Sustainability, 14 (5), 2756, 20 doi:10.3390/su14052756.

Cvetek, D., Muštra, M., Jelušić, N. & Tišljarić, L. (2021) A Survey of Methods and Technologies for Congestion Estimation Based on Multisource Data Fusion. Applied Sciences-Basel, 11 (5), 2306, 19 doi:10.3390/app11052306.

Tišljarić, L., Fernandes, S., Gama, J. & Carić, T. (2021) Spatiotemporal Road Traffic Anomaly Detection: A Tensor-Based Approach. Applied Sciences-Basel, 11 (24), 12017, 17 doi:10.3390/app112412017.

Erdelić, T., Carić, T., Erdelić, M., Tišljarić, L., Turković, A. & Jelušić, N. (2021) Estimating congestion zones and travel time indexes based on the floating car data. Computers environment and urban systems, 87, 101604, 17 doi:10.1016/j.compenvurbsys.2021.101604.

Tišljarić, L., Carić, T., Abramović, B. & Fratrović, T. (2020) Traffic State Estimation and Classification on Citywide Scale Using Speed Transition Matrices. Sustainability, 12 (18), 7278, 16 doi:10.3390/su12187278.

Conference papers:

Majstorović, Ž., Tišljarić, L., Ivanjko, E. & Carić, T. (2022) Intersection Traffic State Estimation using Speed Transition Matrix and Fuzzy-based Systems. U: Gini, G., Nijmeijer, H., Burgard, W. & Filev, D. (ur.) Proceedings of the 19th International Conference on Informatics in Control, Automation and Robotics - ICINCO. Lisabon, str. 193-200 doi:10.5220/0011275500003271.

Vaiti, T., Tišljarić, L., Erdelić, T. & Carić, T. (2022) Traffic Emissions Clustering Using OBD-II Dataset Based on Machine Learning Algorithms. U: Petrovic, M., Dovbischuk, I. & Cunha, A. (ur.) International Scientific Conference "The Science and Development of Transport - Znanost i razvitak prometa". Šibenik, Hrvatska, Elsevier BV, str. 364-371 doi:10.1016/j.trpro.2022.09.040.

Vrbanić, F., Tišljarić, L., Majstorović, Ž. & Ivanjko, E. (2022) Reinforcement Learning Based Variable Speed Limit Control for Mixed Traffic Flows Using Speed Transition Matrices for State Estimation. U: 2022 30th Mediterranean Conference on Control and Automation (MED). Atena, Grčka, IEEE, str. 1093-1098 doi:10.1109/med54222.2022.9837279.

Mardešić, N., Erdelić, T., Tišljarić, L. & Carić, T. (2022) Trajectory Estimation from Sparse Cellular Network Data Based on the Historical Vehicular Data. U: Petrovic, M., Dovbischuk, I. & Cunha, A. (ur.) International Scientific Conference "The Science and Development of Transport - Znanost i razvitak prometa". Šibenik, Hrvatska, Elsevier BV, str. 166-173 doi:10.1016/j.trpro.2022.09.020.

Tišljarić, L., Ivanjko, E., Kavran, Z. & Carić, T. (2021) Fuzzy Inference System for Congestion Index Estimation Based on Speed Probability Distributions. U: Bujňák, J. & Guagliano, M. (ur.) 14th International scientific conference on sustainable, modern and safe transport. Visoke Tatre, Slovačka, Elsevier BV, str. 1389-1397 doi:10.1016/j.trpro.2021.07.124.

Tišljarić, L., Cvetek, D., Vareškić, V. & Gregurić, M. (2021) Classification of Travel Modes from Cellular Network Data Using Machine Learning Algorithms. U: Muštra, M., Vuković, J. & Zovko-Cihlar, B. (ur.) Proceedings of ELMAR-2021. Zadar, University of Zagreb, str. 173-177 doi:10.1109/ELMAR52657.2021.9550817.

Majstorović, Ž., Garašić, A., Tišljarić, L. & Carić, T. (2021) Simulation of the Urban Mobility Using Cellular Network Data: Case Study for the City of Rijeka Croatia. U: Ivanjko, E., Stanković, R. & Nikšić, M. (ur.) Proceedings of the International Scientific Conference "The Science and Development of Transport" (ZIRP 2021). Zagreb, Faculty of Transport and Traffic Sciences, University of Zagreb, str. 69-78. (https://www.bib.irb.hr/1148639).

Tišljarić, L., Fernandes, S., Carić, T. & Gama, J. (2020) Spatiotemporal Traffic Anomaly Detection on Urban Road Network Using Tensor Decomposition Method. U: Appice, A., Tsoumakas, G., Manolopoulos, Y. & Matwin, S. (ur.) Discovery Science. Cham, Springer, str. 674-688 doi:10.1007/978-3-030-61527-7_44.

Tišljarić, L., Majstorović, Ž., Erdelić, T. & Carić, T. (2020) Measure for Traffic Anomaly Detection on the Urban Roads Using Speed Transition Matrices. U: Skala, K. (ur.) MIPRO 2020 43rd International Convention Proceedings. Rijeka, Croatian Society for Information, Communication and Electronic Technology - MIPRO, str. 268-275 doi:10.23919/MIPRO48935.2020.9245327.

Tišljarić, L., Cvetek, D., Muštra, M. & Jelušić, N. (2020) Mixed Impact of the Covid-19 Pandemic and the Earthquake on Traffic Flow in the Narrow City Center: A Case Study for Zagreb-Croatia. U: Ivanjko, E. & Stanković, R. (ur.) Proceedings of the International Scientific Conference "The Science and Development of Transport" (ZIRP 2020). Zagreb, Faculty of Transport and Traffic Sciences, str. 293-300. (https://www.bib.irb.hr/1082336).

Cvetek, D., Tišljarić, L., Mandžuka, B. & Jurčević, M. (2018) APPLICATION OF NLP ALGORITHMS IN ITS MARKET ANALYSIS. U: Proceedings of the International Scientific Conference "Science and Traffic Development" (ZIRP 2018)- Transport and Logistics Industry in Digital Age. Zagreb, Faculty of Transport and Traffic Sciences, University of Zagreb, str. 33-42.

Tišljarić, L., Erdelić, T. & Carić, T. (2018) Analysis of Intersection Queue Lengths and Level of Service using GPS data. U: Muštra, M., Grgić, M., Zovko-Cihlar, B. & Vitas, D. (ur.) Proceedings of 60th International Symposium ELMAR-2018. Zagreb, Faculty of Electrical Engineering and Computing, University of Zagreb, str. 43-46 doi:10.23919/ELMAR.2018.8534668.

Tišljarić, L., Carić, T., Erdelić, T. & Erdelić, M. (2020) Traffic State Estimation Using Speed Profiles and Convolutional Neural Networks. U: Skala, K. (ur.) MIPRO 2020 43rd International Convention Proceedings. Rijeka, Croatian Society for Information, Communication and Electronic Technology - MIPRO, str. 2147-2152 doi:10.23919/MIPRO48935.2020.9245177.

Erdelić, T., Carić, T., Erdelić, M. & Tišljarić, L. (2019) Electric vehicle routing problem with single or multiple recharges. U: Bujňák, J. & Guagliano, M. (ur.) TRANSCOM 2019 13th

International Scientific Conference on Sustainable, Modern and Safe Transport. Novy Smokovec, Slovačka, Elsevier, str. 217-224 doi:10.1016/j.trpro.2019.07.033.

Extended abstracts:

Tišljarić, L., Vrbanić, F., Ivanjko, E. & Carić, T. (2021) Motorway Bottleneck Detection Using Speed Transition Matrices. U: 6th Int'l Workshop on Data Science (IWDS 2021). Zagreb, Hrvatska, 24.11.2021. (https://www.bib.irb.hr/1159648).

Tišljarić, L., Carić, T., Fernandes, S. & Gama, J. (2020) Traffic State Estimation and Anomaly Detection Using Tensor-Based Method. U: 5th Int'l Workshop on Data Science (IWDS 2020). Zagreb, Hrvatska, 24.11.2020. (https://www.bib.irb.hr/1093157).

Tišljarić, L. & Carić, T. (2020) Clustering of the Anomalous Spatiotemporal Traffic Patterns Using Tensor Decomposition Method. U: The 3rd Symposium on Management of Future Motorway and Urban Traffic Systems. Luksemburg, Luksemburg, 06.06.2020. - 08.06.2020. (https://www.bib.irb.hr/1123575).