Data-driven fundamental models for pedestrian movements

THÈSE N^{O} 7613 (2017)

PRÉSENTÉE LE 5 MAI 2017

À LA FACULTÉ DE L'ENVIRONNEMENT NATUREL, ARCHITECTURAL ET CONSTRUIT
LABORATOIRE TRANSPORT ET MOBILITÉ
PROGRAMME DOCTORAL EN GÉNIE ÉLECTRIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Marija NIKOLIĆ

acceptée sur proposition du jury:

Prof. P. Frossard, président du jury
Prof. M. Bierlaire, directeur de thèse
Prof. H. Mahmassani, rapporteur
Prof. S. Hoogendoorn, rapporteur
Prof. N. Geroliminis, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE
Suisse
2017

"The scientific man does not aim at an immediate result.
He does not expect that his advanced ideas will be readily taken up.
His work is like that of the planter - for the future.
His duty is to lay the foundation for those who are to come, and point the way.
He lives and labors and hopes."

— Nikola Tesla, 1934

To my beloved family...
Mojoj voljenoj porodici...

Acknowledgments

In September 2012 I embarked on a long and interesting adventure called PhD studies. If one word could describe all my experiences and memories of the past four and a half years, it would definitely be *rafting*. Similar to rafting, PhD life refers to a new and challenging environment, where dealing with unknown and the need for teamwork are the essential parts of the adventure. Many colleagues, friends and family members have helped me with their paddles to manage some rough water, waves, rocks, drops and sharp maneuvers, and enjoy calm parts during this amazing journey. I am immensely grateful to all of them.

First and foremost, I would like to express my sincere gratitude to my supervisor Prof. Michel Birelaire, both professionally and personally. It has been an honor to be his PhD student. Without his guidance, patience, motivation and impressive knowledge this thesis would not have been possible. His vision and enthusiasm have motivated me to grow as a research scientist and enriched my way of thinking. I could not have imagined having a better advisor and mentor for my PhD study. And not only that. He is also a great friend, with whom I could talk about all possible aspects of life and count on his support. I am thankful to him for making the lab a fantastic place for sharing exciting research ideas, but Belgian beers as well. In the future I will always try to follow his motto: *Being excellent is not good enough!*

Besides my supervisor, I would like to thank the rest of my thesis committee: Prof. Pascal Frossard, Prof. Nikolaos Geroliminis, Prof. Serge Hoogendoorn and Prof. Hani Mahmassani, for their insightful comments and encouragement, and for the challenging questions which incited me to improve the quality of my thesis from various perspectives. I am also thankful to them for making my private PhD defense an enjoyable moment.

This thesis has been supported by the Swiss National Science Foundation. I am also thankful to the Swiss Federal Railways (SBB) and Alexandre Alahi for providing us with the unique data sets, which represent the core of the thesis.

My sincere thanks goes to my current and former colleagues and friends from the Transport and Mobility Laboratory. They made me enjoy every single day of my PhD life. I love the stimulating discussions we had in the lab and during our breaks, and I am really grateful for all the fun and crazy things we have had during our trips. Thank you: Anne Curchod for your help with IT related issues; Anna Fernández Antolin for your close friendship, for being always there to laugh and cry with me, for supporting Serbian water polo team and for all the nice gifts you left on my white board; Evanthia Kazagli and Iliya Markov for all the talks about Balkan food, music, culture and history;

Marianne Ruegg for making my life easier since I came to Lausanne; Matthieu de Lapparent for many interesting discussions, great French food and wine, and for being my accomplice; Meritxell Pacheco Paneque for teaching me how to dance like a real Catalan; Mila Manorma Bender for smiling all the time, for helping me with many administrative tasks, for organizing the best parties in your "cabane", for cooking the best Surinamese food; Nicholas Alan Molyneaux for improving my French and having the best British pronunciation; Riccardo Scarinci for relaxed and positive atmosphere in the office, and the best pizzas you make; Shadi Sharif Azadeh and Yousef Maknoon for the tones of unforgettable moments, and for becoming my siblings; Tomáš Robenek for being a great friend, an awesome office mate, for being next to me from the beginning, and for pointing out my weakness, and emphasizing my strengths; Stefan Binder for being my private pilot, and for making me realize that beside tennis, basketball and water polo I can enjoy hockey matches as well; Virginie Lurkin for coloring my days with bright colors, and choosing to be my friend. I am also thankful to Bilal Farooq for many beautiful moments, Dimitrios Efthymiou for being the best, Bilge Atasoy and Aurélie Glerum for helping me smoothly sail into PhD life, Flurin Hänseler and Antonin Danalet for many professional and less professional discussions, and Amanda Stathopoulos, Daniel Emery, Jianghang Chen, Jingmin Chen, Michaël Thémans, Nitish Umang, Ricardo Hurtubia, Stefano Moret and our friends from the Urban Transport Systems Laboratory (LUTS) for providing a great working environment.

Special thanks to my friends Milena Gojković, Aleksandar Djordjević, Marko Dedić, Vladimir Višnjić, Nikola Obrenović and Marina Oros for believing in me, for motivating and supporting me throughout this period. I am also grateful to my Serbian, Greek, Swiss and Catalan gangs: Andrej, Milena and Miloš Nikolić, Ana and Predrag Milosavljević, Klara, Edit and Milenko Tokić, Tamara Tasić, Sana, Mima and Predrag Spasojević, Dražen Nadoveza, Nevena Šaponjić, Jovanče Trajković, Ana Miličić, Anastasia Founta, Sofia Samoili, Stefanos Giannakis, Jean-Daniel Bender (JD), Aitor Pérez, and many others that made my life in Switzerland just amazing.

Last but not the least, I would like to thank my beloved family for being the greatest source of my happiness. I am thankful to my mum and dad, Zorica and Dragan Nikolić, my sister, Jelena Savić, my little nephews, Nikola and Mihailo Savić, and my grand mother, Nada Cvetković, for their boundless love and the support throughout my life. Without you, all of this would not have been possible. I am also thankful to my second family, my parents-in-law Ljiljana and Željko Bardić, my brothers-in-law and sister-in-law Igor Bardić, Aleksandar Savić and Nataša Bardić for their love, support and understanding, and Lana Bardić for being my little angel. And finally, the biggest thanks to my beloved husband and my best friend Darko Bardić. Thank you for encouraging me to embark on this adventure. Thank you for being the wind at my back and my calm harbor. Thank you for making all my dreams come true. *Hvala vam! Volim vas najviše na svetu!*

Lausanne, 20 March 2017 Marija Nikolić

Abstract

The focus of the thesis is the utilization of the data collected using state-of-the-art tracking technologies for the characterization and modeling of pedestrian movements. In this context, the main objectives are the development of (i) data-driven definitions of fundamental variables and (ii) data-inspired mathematical formulations of fundamental relationships characterizing pedestrian traffic.

The motivation of this research comes from the analysis of a real dataset collected in the train station in Lausanne, Switzerland. To collect the raw data, a large-scale network of smart sensors has been deployed in the station. We consider this case study to illustrate and validate our methodology.

The definitions of fundamental traffic variables (speed, density and flow), existing in the literature are extended through a data-driven discretization framework. The framework is based on spatio-temporal Voronoi diagrams, designed using pedestrian trajectory data. The new definitions are (i) independent from an arbitrarily chosen discretization, (ii) appropriate for the multi-directional composition of pedestrian traffic, (iii) able to reflect the heterogeneity of pedestrian population and (iv) applicable to pedestrian trajectories described either analytically or as a sample of points. The performance of the approach and its advantages are illustrated empirically. Our approach outperforms the existing methodologies from the literature, in terms of the smoothness of the results, and in terms of the robustness with respect to the simulation noise and sampling frequency.

To represent fundamental relationships of pedestrian traffic, we introduce probabilistic speed-density models. The approach is motivated by the high scatter in the data that we have analyzed. To characterize the observed pattern we relax the homogeneity assumption of the equilibrium relationships, and propose two models. The first model is based on distributional assumptions. The second model is more advanced, and it includes structures that are designed to capture specific aspects of the walking behavior. Various empirical tests validate the specification of both models. Contrasted with existing approaches, they yield a more realistic representation of the empirically observed phenomena.

This thesis contributes with respect to the utilization of data potential in modeling of fundamental aspects related to pedestrian traffic. This becomes essential in the context of the growing data revolution and interconnected technologies that can help improve the safety and convenience of pedestrians. The methodological framework is fairly general, and it can be adapted to various pedestrian facilities.

Keywords: individual pedestrian trajectories, time and space discretization, three-dimensional Voronoi tessellations, fundamental traffic variables, fundamental traffic relationships, speed-density relationship, probabilistic model, heterogeneity, latent classes

Résumé

Le sujet principal de cette thèse concerne l'utilisation de trajectoires de piétons collectées par des techniques à la pointe de la recherche afin de caractériser et modéliser les mouvements de piétons. Dans ce contexte, les principaux objectifs sont (i) le développement de variables fondamentales à partir d'observations empiriques et (ii) la formulation mathématique des relations fondamentales caractérisant les flux de piétons.
Cette recherche est inspirée d'analyses effectuées sur des données réelles collectées dans la gare de Lausanne, Suisse. Pour récolter ces trajectoires de piétons, un large réseau de capteurs intelligents a été installé dans la gare. Nous utilisons ce cas d'étude pour valider notre méthodologie et nos modèles.
Les variables fondamentales (vitesse, densité et débit) définies dans la littérature sont étendues avec une discrétisation inspirée des données enregistrées. Plus précisément, des diagrammes de Voronoi spatio-temporels sont créés en se basant sur les trajectoires des piétons. Les nouvelles définitions qui en découlent sont (i) indépendantes d'une discrétisation arbitrairement choisie, (ii) adéquates pour adresser l'aspect multidirectionnel des flux piétonniers, (iii) capables de représenter l'hétérogénéité de la population et (iv) peuvent être appliquées pour les trajectoires de piétons représentées de façon analytique ou comme suite de points. Comme mentionné précédemment, nous avons validé les performances de notre approche de façon empirique. Celle-ci produit de meilleurs résultats que les méthodologies existantes à la fois en terme de régularité des résultats, de robustesse face à l'échantillonnage et de robustesse face au bruit induit par les simulations.
Une nouvelle méthode inspirée des données récoltées permet la définition des relations fondamentales de trafic. Nous utilisons de façon novatrice les distributions statistiques combinées avec les diagrammes vitesse-densité afin d'ajouter une dimension probabiliste à ces diagrammes. Ce choix est motivé par la grande variabilité présente dans les données. De plus, afin de caractériser les motifs observés dans les données, nous proposons deux modèles dans lesquels nous assouplissons la contrainte d'homogénéité des équations d'équilibre. Le premier est basé sur des distributions statistiques tandis que le second est plus sophistiqué en introduisant des structures qui prennent en compte des aspects particuliers du comportement des piétons. Afin de valider les résultats plusieurs tests empiriques ont été effectués. Contrairement aux modèles classiques, les nôtres produisent des résultats représentant de façon plus fidèle les phénomènes observés en réalité.
Cette thèse contribue à l'exploitation du potentiel offert par la quantité croissante de

données dans le domaine de la modélisation des principes fondamentaux des flux de piétons. Ceci est essentiel aujourd'hui, au vu de l'accroissement de la quantité et du type de données résultant des technologies connectées, afin d'améliorer le confort et la sécurité des piétons. La méthodologie développée dans cette thèse est suffisamment générale pour être appliquée à différentes situations, notamment des infrastructures différentes.

Mots-clés: trajectoires individuelles de piétons, discrétisation spatio-temporel, pavage tridimensionnel de Voronoi, variables fondamentales de trafic, relations fondamentales de trafic, diagrammes vitesse-densité, modèle probabiliste, hétérogénéité, classes latentes

Contents

List of Figures

List of Tables

1 Introduction

This chapter is organized into four parts. The first part (Section 1.1) provides essential context for the thesis, and outlines the main research motivation. The second (Section 1.2) and the third part (Section 1.3) present objectives, respectively contributions of the research that is proposed in the thesis. The last part (Section 1.4) describes the structure of the thesis.

1.1 Context and motivation

The data created in various fields has dramatically increased over the last years. Thanks to the advances in information technology, the data volume in the world expanded by nearly five times during only five years (Gantz and Reinsel, 2011). The main sources of data nowadays are (i) the operation and trading information in enterprises, (ii) human interaction and position information in the world of Internet and mobile devices (e.g. chatting records, forum posts, searching entries, video, audio and image files, etc.) and (iii) scientific research. By the year of 2030, Internet of Things (IoT) is predicted to become dominant in data generation (Chen et al., 2014). It refers to a vast amount of networking sensors deployed all over the world, that are collecting and transmitting data from different areas (e.g. industry, agriculture, transportation, medical care, geography, etc.).

The growth of data allows for answering questions previously considered to be beyond

our reach. It brings about new opportunities for discovering new values, and helps to gain an in-depth understanding of unknown, but potentially useful information. The applications include enterprise management, the analysis of social networks-oriented data, medical and healthcare services, collective intelligence, smart grid, etc. The smart city is an example of the application of IoT data (Bowerman et al., 2000; Petrolo et al., 2014). It has arised as a new paradigm to tackle key challenges related to the rapid urban growth over recent years (e.g. increased traffic and congestion, climate change, exhaustion of local resources, changes in the physical environment, and in the form and spatial structure of cities). The cities of the future are expected to improve the quality of their residents' lives by relaying on the growing data revolution, and interconnected and networked technologies. Transportation and mobility of people are essential fields for building innovative and effective solutions within this context. Traffic lights that adjust based on vehicular flows (Lämmer and Helbing, 2008), real-time estimation of travel times (Skabardonis and Geroliminis, 2008), autonomous and connected vehicles (Talebpour and Mahmassani, 2016), and monitoring and management of pedestrian traffic (Daamen and Hoogendoorn, 2003; Alahi et al., 2014) are just a few of the ideas that can help improve the safety and convenience of people in the future.

As a consequence of an intense urbanization, congestion has become an issue of the modern society. As such, it is typical for pedestrians facilities as well. The examples range from transportation hubs, such as airports (Kalakou and Moura, 2014) and train stations (Daamen, 2004; Van den Heuvel and Hoogenraad, 2014; Hänseler et al., 2014, 2017), to museums (Yoshimura et al., 2014; Kanda et al., 2007), music festivals (Naini et al., 2011; Duives et al., 2014a), commercial centers (Lam and Cheung, 2000; Yaeli et al., 2014), university campuses (Danalet et al., 2014), crosswalks (Lam et al., 2002; Rastogi et al., 2013) or even religious infrastructures (Algadhi and Mahmassani, 1990; Helbing et al., 2007). Congestion in pedestrian facilities represents a phenomenon with a negative impact on pedestrian dynamics. It prevents pedestrians from achieving efficient movements and may lead to an increase in travel time, delays and potential collisions among pedestrians. A simple application of a particular policy, without the previous study of the concrete problem, might lead to some very costly trial and error solutions. Thus, a sophisticated understanding and modeling of the data behind complex pedestrian movement patterns (Bierlaire and Robin, 2009) is necessary for (i) an efficient planning and management of future pedestrian facilities and (ii) the optimization of current infrastructure and operations.

Data collection for pedestrian flow and behavior analysis used to be particularly cumbersome. Typically, manual counting methods (on-site or on videos) and surveys distributed to randomly selected individuals were the main sources of data. Nowadays, automatic pedestrian detection and tracking methods have evolved tremendously, allowing for more comprehensive analyses (Bauer et al., 2009). In order to obtain an empirical understanding of pedestrian movements, different empirical studies were conducted and reported in the literature. For instance, it was observed that directness, habits, pleasantness, safety,

pollution and noise levels are some of the important attributes for pedestrian route choice (Bovy and Stern, 2012; Seneviratne and Morrall, 1985). Other empirical studies have revealed the existence of relationships between traffic indicators (Weidmann, 1993). Self-organized structures in pedestrian flows, such as lane formation (Daamen and Hoogendoorn, 2003; Hoogendoorn and Daamen, 2004), and phenomena like stop and go waves (Helbing et al., 2007), herding (Helbing et al., 2005), the faster is slower (Helbing and Johansson, 2010), the zipper effect (Hoogendoorn and Daamen, 2005), were also empirically discovered.

Data is also crucial to the process of model formulation, calibration and validation. The empirical observations have inspired a number of theories and models of pedestrian movements. They are utilized to describe and predict pedestrian movement at strategic, tactical, and operational level (Hoogendoorn and Bovy, 2004). The models concerned with strategic decisions (departure time choice, activity pattern choice) are important for the assessment of pedestrian demand (Danalet, 2015). The models at tactical level are focused on activity scheduling and route-choice (Borgers and Timmermans, 1986; Cheung and Lam, 1998; Stubenschrott et al., 2014). Together with the models at operational level (walking behavior), they are used to evaluate quality levels of pedestrian traffic, and have been applied to support the design and planning processes in many areas related to pedestrians (Daamen, 2004). The modeling of pedestrian walking behavior attracts a significant attention. For instance, there are approaches that are based on social force-fields (Helbing and Molnar, 1995), cellular automata (Blue and Adler, 2001), continuum flow (Hughes, 2002; Hoogendoorn et al., 2014), utility maximization (Robin et al., 2009) and queuing theory (Cheah and Smith, 1994; Løvås, 1994). A comprehensive review of the existing approaches and their evaluation can be found in Duives et al. (2013).

The fundamental quantities used at both levels, to observe and to model the traffic of pedestrians, are density (k), speed (v) and flow (q). *Density* is expressed as the number of pedestrians per unit of space at a given moment in time; *flow* is interpreted as the number of pedestrians per unit of time and per unit of length; *velocity* is expressed in unit of length per unit of time (Daamen, 2004). Several definitions of these variables are proposed in the literature (Duives et al., 2015; Zhang, 2012). However, little concern is dedicated to the nature of spatial and temporal discretization underlying the definitions. The basic issue is that there are many possible ways to discretize continuous space and time for the purpose of defining traffic variables. Yet, studies normally report the analysis for one particular discretization scheme whose choice is often arbitrary. Furthermore, it is the highly heterogeneous and complex nature of pedestrian movement behavior that accounts for simplifications regarding pedestrian flow characterization. The definitions of pedestrian traffic variables are usually developed on the grounds of drawing the parallels between pedestrian and vehicular traffic. This is due to the fact that the field of vehicular flow modeling is quite well established (Daganzo and Newell, 1995; Hoogendoorn and Bovy, 2001). However, this analogy is considered useful to some extent, because of large

differences that exist between the two types of traffic flow. In comparison to roadways, where vehicular flow is separated by directions and hence is regulated, pedestrian flow is such that pedestrian facilities permit them to move in a multi-directional fashion. The lack of strict rules for pedestrians to follow allows them to occupy any part of walkway so they are able to change their speed and direction faster than vehicles (Blue and Adler, 2001).

The relationships between density and flow, density and speed, and flow and speed, are referred to as the *fundamental relationships*. They are specified under the assumption that the traffic system is at equilibrium, that is stationary and homogenous. The fundamental relationships play an important role in the field (Weidmann, 1993; Daamen, 2004). They represent simple models capable of predicting pedestrian flow under specific circumstances, and are thus useful for planning and designing of pedestrian facilities (Zhang, 2012). They are also a required input or calibration criterion for models of pedestrian dynamics (Hänseler et al., 2014; Blue and Adler, 2001). The existing models of these macroscopic relationships are rather descriptive than explanatory, and they lack generality (Hoogendoorn, 2001). More importantly, empirical analyses reported in the literature reveal significant scatter in the data, which eliminates the use of a unique equilibrium relationship (Cheung and Lam, 1998; Daamen et al., 2005; Steffen and Seyfried, 2010). A possible approach to capture this complex phenomenon consists in modeling explicitly the exact underlying walking process and the explicit interactions at the disaggregate level. It would allow to test various hypotheses, at the expense of losing computational efficiency (Duives et al., 2013). An alternative approach, based on an aggregate representation of the pedestrian traffic, is still missing.

Although advanced theories and pedestrian models at strategic, tactical, and operational level exit, the fundamental aspects related to pedestrian movement are still not adequately treated. This is where this thesis aims at making a contribution, by exploiting the great potential of the available data. We propose a data-driven approach to pedestrian traffic characterization, and an econometric perspective in modeling the relationship between traffic characteristics. Our analyses are predominantly based on a real dataset of pedestrian trajectories, which is collected in the train station in Lausanne, Switzerland (Alahi et al., 2011, 2014). The infrastructure data and train timetables are also available for this case study. Additionally, we make use of the data from a controlled experiment performed by the Technical University of Delft (Daamen and Hoogendoorn, 2003), and the data obtained using pedestrian simulation tools (Campanella et al., 2014).

1.2 Objectives

The goal of this thesis is the exploitation of the data collected using state-of-the-art technologies in dealing with the issues related to the characterization and modeling of

pedestrian movements. In that context, the objectives are threefold:

1. Pedestrian flow characterization
   (a) To develop a data-driven mathematical framework for spatio-temporal discretization suitable for pedestrian flow studies.
   (b) To define fundamental traffic variables in a pedestrian-oriented manner, and within developed discretization scheme.
2. Pedestrian flow modeling
   (a) To process and analyze a rich set of pedestrian data, in order to identify phenomena characterizing the relationships between fundamental traffic variables.
   (b) To propose a theory representing discovered phenomena.
   (c) To operationalize the proposed theory through the development of suitable mathematical framework.
3. Application
   (a) To apply developed models on empirical data, so as to evaluate the performance and applicability of the framework.
   (b) To provide directions for the use of the models in practice.

1.3 Contributions

This thesis contributes with respect to the utilization of data potential in modeling of fundamental aspects related to pedestrian traffic. The main scientific contributions of this thesis include:

1. Pedestrian flow characterization
   (a) *Data-driven discretization* (Chapter 3) - The derivation of a discretization framework that is based on spatio-temporal Voronoi diagrams. It is designed through the utilization of pedestrian trajectories.
   (b) *Data-driven characterization* (Chapter 3) - The definition of fundamental traffic variables that are (i) independent of an arbitrarily chosen discretization, (ii) appropriate for the multi-directional composition of pedestrian traffic, (iii) able to reflect the heterogeneity of pedestrian population and (iv) applicable to pedestrian trajectories described either analytically or as a sample of points.

2. Pedestrian flow modeling - The development of methodologies for representing speed-density relationships for pedestrian traffic, in particular taking the observed nature of the data into account. The scatter is represented by relaxing the equilibrium assumption in two ways

   (a) *Probabilistic speed-density relationship* (Chapter 4) - Implicitly accounting for the heterogeneity of speeds by bringing together first principles and an econometric approach. The approach is probabilistic and utilizes only the information contained in pedestrian trajectory data.

   (b) *Multi-class speed-density relationship* (Chapter 5) - Explicitly explaining the heterogeneity of speeds by using latent class methodology. The approach is probabilistic and combines different data sources (e.g. traffic condition data, infrastructure data, characteristics of individuals, etc.).

The developed fundamental models are applied to different case studies in order to evaluate their performance and to illustrate their advantages, with respect to existing approaches. Our data-driven characterization is shown to outperform the existing methodologies from the literature, in terms of the smoothness of the results, and in terms of the robustness with respect to the simulation noise and sampling frequency. With respect to speed-density relationships, various empirical tests validate the proposed data-inspired model specifications. Contrasted with existing approaches, our approach yields a more realistic representation of the empirically observed phenomena.

1.4 Thesis structure

The thesis is structured as described in the following.

Chapter 2 presents a review of the state-of-the-art with respect to the data that is available in the context of pedestrian movements, and characterization and modeling of fundamental aspects of pedestrian movements. The chapter also provides a review of the literature associated with vehicular traffic theory, that is relevant for pedestrians.

Chapter 3 describes the proposed methodology for the derivation of the spatio-temporal discretization framework. Based on this framework, we derive the definitions of the pedestrian traffic variables, that is density, flow and velocity.

The preliminary ideas of the methodology presented in this chapter are published as:

> Nikolić, M., and Bierlaire, M. (2014). Pedestrian-oriented flow characterization, *Transportation Research Procedia* 2:359-366.
>
> Nikolić, M., and Bierlaire, M. (2015). Pedestrian flow characterization

based on spatio-temporal Voronoi tessellations. *Proceedings of the 15th Swiss Transportation Research Conference (STRC) 15-17 April, 2015.*

This chapter is based on the article published as:

Nikolić, M. and Bierlaire, M. (2016). Data-driven characterization of pedestrian flows. *Technical report TRANSP-OR 160815.*

The article has been accepted for publication in the proceedings and a special issue of an academic journal for the 22nd International Symposium on Transportation and Traffic Theory, Northwestern University, July 24-26, 2017.

Chapter 4 introduces a probabilistic speed-density relationship for pedestrians. The model is able to implicitly account for the heterogeneity of pedestrian flows. As such, it is useful in the context where pedestrian trajectory data is the only available information.

The preliminary ideas of the methodology presented in this chapter are published as:

Nikolić, M., Farooq, B., and Bierlaire, M. (2013). Exploratory analysis of pedestrian flow characteristics in mobility hubs using trajectory data. *Proceedings of the Swiss Transportation Research Conference (STRC) 24-26 April, 2013.*

Nikolić, M., Bierlaire, M., and Farooq, B. (2014). Probabilistic speed-density relationship for pedestrians based on data driven space and time representation. *Proceedings of the Swiss Transportation Research Conference (STRC) 14-16 May, 2014.*

This chapter has been published as:

Nikolić, M., Bierlaire, M., and Farooq, B. (2015). Probabilistic speed-density relationship for pedestrian traffic: a data-driven approach. *Technical report TRANSP-OR 150411.*

Nikolić, M.,Bierlaire, M., Farooq, B. and de Lapparent, M. (2016). Probabilistic speed-density relationship for pedestrian traffic, *Transportation Research Part B: Methodological* 89: 58 - 81.

Chapter 5 develops the model of speed-density relationship using the latent class methodology. The heterogeneity of speeds is, in this case, explicitly explained using variables related to train timetables, pedestrian type and infrastructure data.

This chapter is based on the article:

Nikolić, M., Bierlaire, M., Lapparent, M. and Scarinci, R. (2017). Multi-class speed-density relationship for pedestrian traffic. *Technical report TRANSP-OR 170115*.

Chapter 6 summarizes the contributions of the thesis and determines future research directions.

2

State of the Art

This chapter is organized into four parts. The first part (Section 2.1) focuses on data that is available in the context of pedestrian movements. The second part (Section 2.2) deals with the characterization of pedestrian traffic. The third part (Section 2.3) discusses the models of speed-density relationships for pedestrians. The chapter is finalized by drawing some conclusions in Section 2.4

In addition to the approaches proposed for pedestrian movements characterization and modeling, we also consider those from the vehicular traffic theory. The approaches derived in the field of vehicular traffic are relevant for pedestrians as well, but for most applications in pedestrian flow theory they can not be directly used.

2.1 Pedestrian data

Data is essential in understanding and modeling of different phenomena, as well as in the process of calibration and validation of the models. A number of different technologies have been developed and used in recent years to collect data related to pedestrians. The data collections have been performed for different purposes (e.g. health, marketing, transportation, etc.) and at different scales (e.g. city, facilities), as elaborated in Danalet (2015). We focus on pedestrian movement data collected in normal situations. Within this context, the data is predominantly used to gain better understanding about pedestrian movement behavior, and to access the demand and the level of service of

pedestrian facilities. This is important in order to provide services related to pedestrian comfort, convenience, efficiency and safety (Bauer et al., 2009).

Different types of pedestrian movement data are reported and utilized in the literature. They can be roughly classified into three main categories: survey data, counts and trajectories. Each data type is associated with specific data collection technologies, and the corresponding opportunities and limitations. In the following, we provide a description of these data types, and discuss their importance.

2.1.1 Survey data

Survey data on pedestrian movement behavior are typically obtained using revealed preference questionnaires. The questionnaires ask the respondent to describe their past choices, such as the path they followed for a specific trip, and to report about it. They are usually distributed to randomly selected individuals in pedestrian facilities (Kalakou and Moura, 2014), or conducted via mail, telephone or computer assisted tools (Kazagli et al., 2014). This technique allows to gather additional information, such as socio-demographic, trip context data (e.g. purpose), and potentially description of unchosen alternatives (Kazagli et al., 2014). Complex logistics, selection bias, low temporal and spatial accuracy and recall bias, since it relies on respondents memory, are the main drawbacks related to this data type.

The data collected by means of surveys is commonly used at strategic and tactical level, solely or together with other data sources (Hoogendoorn and Bovy, 2004). For instance, Verlander and Heydecker (1997) have presented an application of a sample survey of daily walks in an urban area in studying route choice for pedestrians. In Kalakou and Moura (2014), a survey data is used to estimate location choice models for a given activity type. Lin and Chen (2013) report the use of passenger survey data, collected at Taiwan's Taoyuan International Airport, to examine the relationship between passengers' shopping motivations and their commercial activities at airports.

2.1.2 Count data

Pedestrian count data are typically represented as a pair of the number of people crossing a particular (virtual) section and a time stamp. The traditional method to obtain pedestrian counts is the manual one. This method is simple, low-cost and reliable. The human factor and the complexity of the scene appear to be crucial for counting accuracy in this case (Diogenes et al., 2007; Greene-Roesel et al., 2008).

Turnstiles, infra-red beams and switching mats are exemplary representatives of the automatic counting methods. They allow for the collection of count data at low cost.

On the other hand, the accuracy of such methods is low and it depends on the separation of pedestrians. Also, the installation of the equipment might be an issue, as well as the sparsity of the observations (Bauer et al., 2009).

More sophisticated counting systems rely on video-based techniques. A number of computer vision algorithms have been developed for extracting pedestrian counts from video footage (Sexton et al., 1995; Zhang and Sexton, 1997; Tan et al., 2011; Chen et al., 2006). These systems deliver reliable observations, but are limited by higher cost and installation issues (e.g. require top view locations). Counts can also be obtained from sales (Hänseler, 2016) or automatic fare collection data, i.e., check-in and check-out of passengers (Van den Heuvel and Hoogenraad, 2014).

Pedestrian count data, obtained manually or automatically, can be used to obtain the information on flow rates or calibrate pedestrian demand and flow models (Cascetta and Russo, 1997; Hänseler, 2016).

2.1.3 Trajectory data

Pedestrian trajectory data are provided in the form of individual-specific pairs of consecutive time and location observations. Different tracking methods have been designed to collect this type of data.

Pedestrian trajectories are usually collected exploiting video footage. In Daamen and Hoogendoorn (2003) the individual trajectories were collected using digital video cameras during controlled experiments at the Technical University of Delft in the Netherlands. Wong et al. (2010) carried out the experiments in a sports hall in Hong Kong, and collected video footage on 89 controlled experiments. Plaue et al. (2014) conducted a multi-directional pedestrian flow experiment in the entrance hall of a university building at TU Berlin, Germany, and the setting is video-recorded using three networked cameras. Hoogendoorn et al. (2003) propose a computer vision-based approach to extract individual pedestrian data from video sequences. A methodology for generating individualized trajectories from video footage is also discussed in Mehner et al. (2015). Antonini et al. (2006) propose a framework for detection and tracking of pedestrians in video sequences, which is based on a combination of image processing methods and behavioral models of pedestrian dynamics. In addition to high cost of video-based systems, Bauer et al. (2009) indicate that abrupt object motion, changing appearance patterns and occlusions lead to major difficulties in video-based tracking.

To overcome the high cost and limitations of traditional data collection methods, simpler approaches have emerged. They are based on wireless technologies, among which WiFi and Bluetooth are particularly attractive (Lesani and Miranda-Moreno, 2016). With Bluetooth and WiFi, a unique media access control (MAC) address for each device is

obtained and thus each device can be monitored as it moves through a network. WiFi and Bluetooth traces offer the opportunity to collect panel data in the long term. The issue with these technologies is the low accuracy and temporal resolution, and strong sample bias (Liu et al., 2007). Pedestrian data collected using wirelesses technologies are topically employed to study route and activity choice. Yaeli et al. (2014) analyze pedestrian behavior in stores based on WiFi traces. Danalet (2015) proposes methodologies to detect and model activity episodes from WiFi data. The data collected via wireless technologies is also used to obtain the information on OD patterns (Versichele et al., 2012; Kim et al., 2015) and to analyze and improve the attractiveness of pedestrian facilities (Kanda et al., 2007). Also, in Hänseler (2016) it is discussed that the combination of WiFi or Bluetooth traces with count data may provide better understanding about prevailing traffic conditions.

Technological progress have encouraged the use of smart sensors (e.g. optical, thermal and depth sensors) in pedestrian facilities for surveillance purposes (Alahi et al., 2013). Due to their falling cost, a large number of optical sensors is already installed in public spaces. Their usage in tracking pedestrians is however limited by their high sensitivity to reflections. Thermal sensors are less affected by changes of lighting conditions, but should be installed high enough to avoid self-occlusion. To deal with the occlusion problems, depth sensors are preferred. The main drawbacks of depth sensors is their limited monitoring range (up to 7 meters) (Alahi et al., 2013). Clearly, every sensing modality features some advantages and limitations. Requirements for protection from vandalism and bad weather conditions play an additional role in all the cases (Hänseler, 2016). Alahi et al. (2014, 2011) presented the mathematical framework that handles the inputs from different sensors to allow for high accuracy in pedestrian detection and tracking. Their methodology is applied to a real case study involving a Lausanne railway station, Switzerland. Pedestrian trajectories from this case study are utilized to calibrate a dynamic pedestrian propagation model proposed by Hänseler et al. (2014), which is applicable to congested and multi-directional flows.

Pedestrian trajectory data contain useful information regarding pedestrian motion. Detailed observations obtained using cameras and smart sensors are extremely important for the estimation and validation of pedestrian models at operational level (Johansson et al., 2007; Hoogendoorn and Daamen, 2007). However, socio-economic information is missing in this case.

2.1.4 Other data

Depending on the purpose of a particular study, the availability of other data sources might be also desirable. Pedometers can be useful for the analysis of walking patterns (Whitt et al., 2004). Several studies also employ information available from online social media websites to study pedestrian movements. Wood et al. (2013) use the locations of

photographs in *flickr* to estimate visitation rates, and information from the profiles of the photographers to derive travelers' origins. Dashti et al. (2014) show how *tweets* can be used to support a digital survey of a region and to navigate optimal paths. In some circumstances, the attributes of a particular facility, such as detailed dimensions and locations of relevant parts (e.g. access doors, shops, etc.) is of particular importance (Hui et al., 2009; Hillier and Tzortzi, 2006; Haq and Luo, 2012; Kalakou and Moura, 2014). In transportation hubs, for instance, timetables and schedules have a significant impact on the usage of facilities (Daamen, 2004). Domain knowledge by practitioners about attractive areas or peak periods may also provide a significant insight regarding the model specification.

2.2 Traffic characterization

Traffic characterization is usually based on the quantities such as velocity, density and flow. They represent the main Level of Service (LoS) indicators. Also, they are the fundamental variables used to model traffic related phenomena.

The definitions of traffic variables rely on spatial and temporal units obtained using particular discretization scheme. The definition of the discretization scheme is not trivial. It is related to different issues that are well recognized in the field of geography (Openshaw, 1983; Çöltekin et al., 2011) and dynamic systems (Beck and Roepstorff, 1987). The research from geography field have demonstrated that the results of any spatio-temporal analysis depend severely on the underlying discretization. The problem appears in two dimensions, space and time (known as Modifiable Areal Unit Problem - MAUP and Modifiable Temporal Unit Problem - MTUP). For instance, analysis of data using grid-based spatial discretization differs from analysis performed using hexagon cells. Similarly, temporal discretization may distort or exaggerate the actual temporal pattern existing in data if it is based on an arbitrary choice. It is therefore essential that the discretization rely on a meaningful basis relevant for the purpose of the study. The definition of discretization scheme has to precede any attempt to define characteristics based on it.

This section first focuses on vehicular traffic characterization. We then present the approaches specific to pedestrian traffic characterization and their comparison.

2.2.1 Vehicular traffic

The most general and widely used definitions of vehicular traffic variables are proposed by Edie (1963). The definitions are derived based on the trajectories of vehicles $i = 1, ..., N$ in the time-space region A. The shape of the region A is usually rectangular

with duration dt and length dx. The definitions are given as

$$k(A) = \frac{\sum_{i=1}^{N} t_i}{dxdt}, \tag{2.1}$$

$$q(A) = \frac{\sum_{i=1}^{N} x_i}{dxdt}, \tag{2.2}$$

$$v(A) = \frac{\sum_{i=1}^{N} x_i}{\sum_{i=1}^{N} t_i}, \tag{2.3}$$

where t_i and x_i are the time spent by vehicle i, respectively the distance traversed by vehicle i in the region A. This approach is applicable to any time-space domain of interest and provides consistent results in observations and modeling. The determination of the shape, the size and the placement of the time-space region A is however left to the modeler.

Some authors propose a "vehicle-based" discretization (Jabari et al., 2014; Treiber and Kesting, 2013). The definitions of the indicators with this discretization are consistent with the classical definitions of Edie (1963), but with the space-time intervals chosen to fit exactly one vehicle each. Let $x_{i-1}(t)$ and $x_i(t)$ denote the positions of the leader, $i-1$, and the follower, i, at time t. The spacing is defined as $s_i(t) = x_{i-1}(t) - x_i(t)$. The density at time t is defined as the inverse of the spacing $s_i(t)$ measured at that time

$$k(x,t) = \frac{1}{s_i(t)}, \text{ for } x \in [x_i(t), x_{i-1}(t)). \tag{2.4}$$

Let $t_i(x)$ denote the time when vehicle i crosses position x. The time headway is defined as $h_i(x) = t_i(x) - t_{i-1}(x)$. The flow at position x is defined as the inverse of the time headway $h_i(x)$ measured at that location

$$q(x,t) = \frac{1}{h_i(x)}, \text{ for } t \in (t_{i-1}(x), t_i(x)]. \tag{2.5}$$

Speed is defined as the ratio between flow and density

$$v(x,t) = \frac{s_i(t)}{h_i(x)}, \text{ for } x \in [x_i(t), x_{i-1}(t)), t \in (t_{i-1}(x), t_i(x)], \tag{2.6}$$

and it represents a mean speed for vehicle i. This microscopic approach allows to preserve the heterogeneity of driver population.

2.2.2 Pedestrian traffic

One of the first approaches to pedestrian flow characterization was proposed by Fruin (1971). In this method a grid-based spatial discretization is considered and density is defined as

$$k(x, y, t) = \frac{N_A(t)}{|A|}, \text{ for } (x, y) \in A, \tag{2.7}$$

where A is a grid cell, $|A|$ is the area of A, and $N_A(t)$ represents the number of pedestrians present in the cell A at a specific time instant t. The instantaneous velocity of pedestrian i at t is specified using the following formulation

$$\vec{v}_i(t) = \frac{\begin{pmatrix} x_i(t_2) \\ y_i(t_2) \end{pmatrix} - \begin{pmatrix} x_i(t_1) \\ y_i(t_1) \end{pmatrix}}{t_2 - t_1}, \tag{2.8}$$

where $(x_i(t), y_i(t))^T$ refers to the position of pedestrian i, and t_1 and t_2 define the time instants before, respectively after time t. In general, no guidance is provided for the selection of these time instants. The velocity within the cell A is then given as the average of individual instantaneous velocities

$$\vec{v}(x, y, t) = \frac{\sum_{i=1}^{N_A} \vec{v}_i(t)}{N_A}, \text{ for } (x, y) \in A. \tag{2.9}$$

The flow is determined using the fundamental flow equation

$$\vec{q}(x, y, t) = k(x, y, t)\vec{v}(x, y, t), \tag{2.10}$$

which holds only if (2.9) represents space-mean speed, that is when t_2 approaches t_1 in (2.8). In the rest of the thesis, we refer to this method as the grid-based method (GB).

The range-based method (RB) is similar to the grid-based method (Duives et al., 2015). The difference is that a circle defined by radius r at any discrete location in space is used instead of rectangular cells.

In van Wageningen-Kessels et al. (2014) (similar to Saberi and Mahmassani (2014)) the definitions of Edie (1963) are extended by studying pedestrian traffic in a three-dimensional time-space diagram A (of length dx, width dy and duration dt) with pedestrians $i = 1, ..., N$. The density is defined as the average number of pedestrians in the

region $[dx \times dy]$ during time period dt

$$k(A) = \frac{\sum_{i=1}^{N} t_i}{dxdydt}, \tag{2.11}$$

where t_i is the time during which pedestrian i is present in the region A. The flow is defined in x and y directions as

$$\vec{q}(A) = \begin{pmatrix} q_x(A) \\ q_y(A) \end{pmatrix} = \begin{pmatrix} \frac{\sum_{i=1}^{N} x_i}{dxdydt} \\ \frac{\sum_{i=1}^{N} y_i}{dxdydt} \end{pmatrix}, \tag{2.12}$$

where x_i and y_i are the distances traveled in A in direction x, respectively y by pedestrian i. The velocity in direction x (or y) is defined as the average distance traveled in x direction (or in y direction) divided by the total time spent

$$\vec{v}(A) = \begin{pmatrix} v_x(A) \\ v_y(A) \end{pmatrix} = \begin{pmatrix} \frac{\sum_{i=1}^{N} x_i}{\sum_{i=1}^{N} t_i} \\ \frac{\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} t_i} \end{pmatrix}. \tag{2.13}$$

At the limit $dt \to 0$, the density converges to the number of pedestrians present in $[dx \times dy]$ at a specific moment in time. At the limit $dx \to 0$ ($dy \to 0$) flow converges to the number of pedestrians per unit of time and per unit of length. In the rest of the thesis, we refer to this method as the XY-T method.

Another approach is provided by Helbing et al. (2007). This method defines the characteristics at any point (x, y) by weighting the relative influence of the surrounding pedestrians using Gaussian distance-dependent weight function

$$f\left(\begin{pmatrix} x_i(t) \\ y_i(t) \end{pmatrix} - \begin{pmatrix} x \\ y \end{pmatrix}\right) = \frac{1}{\pi R^2} \exp\left(-\frac{\left\|\begin{pmatrix} x_i(t) \\ y_i(t) \end{pmatrix} - \begin{pmatrix} x \\ y \end{pmatrix}\right\|^2}{R^2}\right), \tag{2.14}$$

where R represents the distance up-to-which the influence of pedestrians is taken into account, and $(x_i(t), y_i(t)^T$ the location of pedestrian i. The density is defined as

$$k(x, y, t) = \sum_i f\left(\begin{pmatrix} x_i(t) \\ y_i(t) \end{pmatrix} - \begin{pmatrix} x \\ y \end{pmatrix}\right). \tag{2.15}$$

The velocity is given by

$$\vec{v}(x,y,t) = \frac{\sum_i \vec{v}_i(t) f\left(\begin{pmatrix} x_i(t) \\ y_i(t) \end{pmatrix} - \begin{pmatrix} x \\ y \end{pmatrix}\right)}{\sum_i f\left(\begin{pmatrix} x_i(t) \\ y_i(t) \end{pmatrix} - \begin{pmatrix} x \\ y \end{pmatrix}\right)}, \tag{2.16}$$

where $\vec{v}_i(t)$ is the velocity of pedestrian i at time t, which is given by (2.8). The flow is determined using the fundamental flow equation (2.10). In the rest of the thesis, we refer to this method as the exponentially weighted distance method (EW).

Steffen and Seyfried (2010) propose the method in which the spatial discretization is adjusted to the data through the use of Voronoi diagrams (Okabe et al., 2000). The Voronoi space decomposition assigns a personal region A_i to each pedestrian i, in such a way that each point in the personal region is closer to i than to any other pedestrian, with respect of the Euclidean distance. The density at position (x, y) at time t is defined as

$$k(x,y,t) = \frac{1}{|A_i|}, \text{ for } (x,y) \in A_i, \tag{2.17}$$

where $|A_i|$ is the area of A_i. The velocity is defined based on position differences of pedestrian i between time instances t_1 and t_2

$$\vec{v}(x,y,t) = \frac{\begin{pmatrix} x_i(t_2) \\ y_i(t_2) \end{pmatrix} - \begin{pmatrix} x_i(t_1) \\ y_i(t_1) \end{pmatrix}}{t_2 - t_1}, \text{ for } (x,y) \in A_i, \tag{2.18}$$

where $(x_i(t), y_i(t)^T$ is the location of pedestrian i at time t. Time instances t_1 and t_2 are determined such that the effect of the swaying movement of pedestrians is reduced, which requires an extensive pre-processing of each pedestrian trajectory. The flow within an interval is defined using fractional counts obtained from Voronoi cells: half a person has passed a segment if half of the Voronoi cell has passed it. With the procedure defined above different definitions of density in an area A are possible. For instance, Steffen and Seyfried (2010) proposed

$$k_A = \frac{N}{\sum_{i=1}^{N} |A_i|}, \tag{2.19}$$

where N refers to the number of Voronoi cells overlapping with the area A. The space-mean speed in the observation area A is then

$$v_A = \frac{\sum_{i=1}^{N} v(x,y,t)}{N}, \tag{2.20}$$

where the sum is taken over the instantaneous speeds of all persons that are in A at time t. In the rest of the thesis, we refer to this method as the Voronoi-based method (VB).

There also exist headway-based approaches for the definition of density variable, such as Harmonically Weighted Mean Distance and Minimum Distance, both with or without a vision field taken into account. According to Duives et al. (2015), these approaches are not capable of providing correct and consistent estimation and are therefore excluded from the further analysis in our study.

2.2.3 Comparison of methods

A summary of general characteristics of the approaches to pedestrian traffic characterization is provided in Table 2.1. We first compare the approaches in terms of the scale that is considered, that is in terms of whether the characterization is defined by using the information about a single pedestrian (microscopic) or multiple pedestrians (macroscopic). Then, the analysis is made with respect to the exact way the spatial and temporal aggregation is performed at a given scale. Finally, the approaches are contrasted in terms of the type of data required to perform the characterization.

Method	Scale	Spatial aggregation		Temporal aggregation		Data type
		Unit	Assumptions	Unit	Assumptions	
XY-T	Macroscopic	Area	Shape Size Location	Interval	Duration	Trajectories
Grid-based (GB)	Macroscopic	Cell	Size Location	Interval	Duration	Trajectories Sync. sample
Range-based (RB)	Macroscopic	Circle	Radius Location	Interval	Duration	Trajectories Sync. sample
Exponentially-weighted (EW)	Macroscopic	Range	Influence function Range of influence	Interval	Duration	Trajectories Sync. sample
Voronoi-based (VB)	Microscopic	Voronoi cell	Boundary conditions	Interval	Duration	Trajectories Sync. sample

Table 2.1: Characteristics of the approaches to pedestrian traffic characterization

Most of the methods (XY-T, GB, RB, EW) rely on macroscopic approach. This approach does not always comply with the nature of the underlying system. Pedestrians differ in many ways (Weidmann, 1993; Bierlaire and Robin, 2009) and studying pedestrian movement at the macroscopic level may lead to the loss of heterogeneity. Also, by using macroscopic definitions, velocity and flow vectors may nullify if the pedestrians do not all move in the same direction. As for the XY-T method, van Wageningen-Kessels et al. (2014) state that "if about half of the pedestrians walks from left to right, and the rest walks in the other direction, this causes the flows and velocities in x direction to (almost) cancel out". The same example can be used to conclude that the GB, the RB and the EW method suffer from the same issue. On the other hand, microscopic characterization

(employed in the VB method) is able to reflect these particularities of pedestrian traffic. It is further supported by detailed movement data (at the individual level) that is more and more available due to the advances in tracking technologies (Bauer et al., 2009). The microscopic approach is characterized by higher computational burden, which becomes less problematic in the era of high-performance computers.

All the approaches have in common the arbitrary chosen temporal intervals for the specification of velocity and flow indicators. Most of them (XY-T, GB, RB, EW) additionally depend on an arbitrary spatial aggregation. These may generate noise in the data and the results may be highly sensitive to minor changes. The choice of the shape, size and locations of the spatial units in the methods XY-T, GB and RB influences the results significantly (Steffen and Seyfried, 2010). Also, the use of fixed aggregation over time might cause large fluctuations in the indicator values when pedestrians cross the boundaries of the aggregation units. An additional level of arbitrariness is introduced when a pedestrian is exactly at the border between two units, and an arbitrary decision must be made about what unit she belongs to (Duives et al., 2015). Indicators obtained using the EW approach strongly depend on the radius R and, in general, on the choice of the influence function f, given by (2.14). The VB method of Steffen and Seyfried (2010) is the only one that addresses the issue of arbitrary aggregation in space through a data-driven approach. The spatial units in this approach are not fixed over time. Aggregation follows the trend of the data by computing Voronoi diagrams for every time step. The issue is that Voronoi diagram is potentially not enclosed. There is no clear understanding about where to put the Voronoi boundaries in directions where no other pedestrians are present. Steffen and Seyfried (2010) use a restriction of the individual cells in size ($2m^2$) to deal with this issue, which is active only for a few cells.

An analytical description of the trajectories is required for the XY-T method. Consequently, interpolation has to be used when sampled data is available, which is another source of errors. Note that in Table 2.1 we consider the general formulation of the XY-T method. If the limit conditions are considered (e.g. $dt \to 0$), the method can be also applied on discrete trajectory data. The methods GB, RB, EW and VB can be applied on trajectories described either analytically or as a sample of points. If a sample of points is used, all the approaches require a tracking technology that produces synchronized samples. Often, the cameras used to track pedestrians in a distributed network of cameras operate with different sampling frequency or produce observations at irregular intervals. In this case not all trajectories have observations at the same time instants, which might lead to the underestimation of the indicators. It is therefore necessary to perform the interpolation of trajectories when using non-synchronized samples, before applying the methods.

2.3 Speed-density relationships

The relationships between density and flow, density and speed, and flow and speed are referred to as the fundamental relationships. Speed-density relationships are predominantly used in the literature. They are useful for planning and designing of pedestrian facilities. They are also a required input or calibration criterion for models of pedestrian dynamics. The relationships are specified under the assumption that the traffic system is at equilibrium, that is stationary and homogenous.

2.3.1 Equilibrium relationships

The fundamental diagram that corresponds to stationary and homogenous traffic was first introduced empirically in the field of vehicular traffic by Greenshields et al. (1935). This study established the relationship between spacing (the inverse of density) and speed in a form of a simple linear equation. Since then there have been many empirical studies that were aimed at improving this relationship. A comprehensive review of the models proposed in this field is given in Wang et al. (2013). Some of the established deterministic empirical relationships are listed in Table 2.2, where v_f is the free-flow speed, v_0 is the average travel speed in stop-and-go conditions, k_j is the jam density, k_c is the critical density, and λ, θ, θ_1 and θ_2 are parameters.

Source	Specification	Parameters
Greenshields et al. (1935)	$v(k) = v_f \left(1 - \frac{k}{k_j}\right)$	v_f, k_j
Underwood (1961)	$v(k) = v_f \exp\left(-\frac{k}{k_c}\right)$	v_f, k_c
Newell (1961)	$v(k) = v_f(1 - \exp(-\frac{\lambda}{v_f}(\frac{1}{k} - \frac{1}{k_j}))$	v_f, k_j, λ
Drake et al. (1967)	$v(k) = v_f \exp\left(-\theta k^2\right)$	v_f, θ
Wang et al. (2013)	$v(k) = v_0 + \frac{v_f - v_0}{(1+\exp(\frac{k-k_c}{\theta_1})^{\theta_2}}$	$v_0, v_f, k_c, \theta_1, \theta_2$
Units: $k[veh/km], v[km/h]$		

Table 2.2: Deterministic fundamental relationships - vehicular traffic

Another stream of the literature establishes the fundamental relations by analyzing the behavior of microscopic car-following models under the equilibrium conditions at the aggregate level. Different car-following models result in different specifications of fundamental relations (Jabari et al., 2014). Also, the study of Geroliminis and Daganzo (2008) revels that a macroscopic fundamental diagram (MFD) linking space-mean flow, density and speed exists on a large urban area.

In the context of pedestrian traffic, fundamental relations are usually established by

fitting deterministic curves to empirical data. Both linear and nonlinear speed-density models have been proposed, as reported in Table 2.3, where v_f is the free flow speed, k_j the jam density, and θ and γ are parameters. The linearity of the speed-density relationship has long been questioned for both vehicular and pedestrian flows (Daamen, 2004). An alternative specification has been proposed by Tregenza (1976) where speed decreases exponentially with the increase in density, whereas Weidmann (1993) proposed the so-called Kladek-formula, with a double S-form. The exponential specifications of the relationship appeared to be better for describing the behavior of pedestrian walking speed (Cheah and Smith, 1994). In comparison to fundamental relationships from vehicular traffic, the relationship proposed by Weidmann (1993) corresponds to the model proposed by Newell (1961), while the relationship proposed by Tregenza (1976) can be regarded as the generalization of the model proposed by Underwood (1961). Rastogi et al. (2013) have shown that the speed-density relationship of pedestrian flow on sidewalks also follows the model presented in Underwood (1961).

The proposed relationships clearly differ in terms of functional form, but also in terms of the values of their parameters and supports. For instance, jam density (the maximum density achieved under congestion) goes from 3.8 ped/m^2 to 10 ped/m^2, the reported critical density (the maximum density achievable under free flow) ranges from 1.7 ped/m^2 to 7 ped/m^2 (Seyfried et al., 2010) and the mean of the free-flow speed estimated in different studies is 1.34 m/s while its standard deviation is 0.37 m/s (Daamen, 2004). The researchers have suggested several explanations for these deviations: the cultural differences, the differences between pedestrian facilities and the effects of the environment, flow composition, measurement methods, etc. (Seyfried et al., 2010).

Source	Specification	Parameters
Older (1968) Navin and Wheeler (1969) Fruin (1971) Tanaboriboon et al. (1986) Lam et al. (1995)	$v(k) = v_f - \theta k$	v_f, θ
DiNenno (2002)	$v(k) = v_f - v_f \theta k$	v_f, θ
Tregenza (1976)	$v(k) = v_f \exp\left(-(\frac{k}{\theta})^{\gamma}\right)$	v_f, γ, θ
Weidmann (1993)	$v(k) = v_f \left\{1 - \exp\left(-\gamma\left(\frac{1}{k} - \frac{1}{k_j}\right)\right)\right\}$	v_f, k_j, γ
Rastogi et al. (2013)	$v(k) = v_f \exp(-\frac{k}{\theta})$	v_f, θ
Units: $k[ped/m^2], v[m/s]$		

Table 2.3: Deterministic fundamental relationships - pedestrian traffic

Simulation-based fundamental relationships are predominantly obtained via cellular automaton models. For instance, Blue and Adler (1998) specified an unidirectional cellular automaton model that produces a speed-density relationship similar to the one proposed

by Weidmann (1993). Few studies propose the specifications of fundamental relations derived from first principles (Flötteröd and Lämmel, 2015; Hoogendoorn et al., 2014).

2.3.2 Scatter in empirical observations

The findings from several studies (Cheung and Lam, 1998; Daamen et al., 2005; Steffen and Seyfried, 2010), question the deterministic approach of the listed studies. They indeed report a significant scatter in the empirical speed-density relationship. The observed scatter is not possible to predict by the proposed deterministic models. Cheung and Lam (1998) have reported different distributions of the speed data observed for various ranges of density. In this study, speeds are less evenly distributed for lighter traffic conditions, which is explained by higher freedom that pedestrians have in controlling their movements. This indicates that in addition to density, other factors are likely to influence the speed of pedestrians. In the field of vehicular traffic, the observed scattering is explained by non-stationary dynamical aspects of the traffic system (Treiber and Helbing, 2003), or by the impact of driver heterogeneity (Kim and Zhang, 2008; Jabari et al., 2014).

Weidmann (1993) has empirically shown that the trip purpose of pedestrians represents one of the relevant factors. According to this study free-flow speed of shopping pedestrians is 1.04 m/s, it is 1.45 m/s for commuters and 0.99 m/s for tourists. The speed of pedestrians appears to be affected by the age and the gender as well. According to Bowman and Vecellio (1994), the walking speed of pedestrians who are 60 years old and older is significantly lower than for the rest of the adult population. Weidmann (1993) has reported that children (under 12 years) are not capable of attaining the same speed as adults. According to the same study, walking speed of men is found to be 1.41 m/s, whereas for women it is lower (1.27 m/s). Microscopic approaches capture this complex phenomena by modeling the exact underlying walking process and interactions at the level of individuals (Johansson et al., 2007; Hoogendoorn and Bovy, 2004). Although being highly precise, these approaches suffer from high computational time and require a great deal of dissagregate data.

There are several ways to account for the observed heterogeneity in speed at more aggregate level. They can be roughly divided into two categories: the approaches based on physical laws and those based on econometric principles. The examples from the first category are macroscopic models of vehicular traffic that have adopted individual-level speed-density relations (Colombo, 2003; Khoshyaran and Lebacque, 2009; Lebacque and Khoshyaran, 2013). Mesoscopic models, on the other hand, describe pedestrian behavior in terms of probabilities (e.g. velocity distributions). For instance, the gas-kinetic pedestrian flow model proposed by Hoogendoorn and Bovy (2000) describes the dynamics of the so-called pedestrian phase-space density, which represents a combination of the density and the speed distribution.

The models from the second category can be derived from a deterministic model by allowing for uncertainty, or from a model which itself is probabilistic. In the field of vehicular traffic, Wang et al. (2013) proposed a model derived by adding Gaussian noise to the existing deterministic relationships. This can potentially lead to unrealistic outcomes (e.g. negative speed values). Jabari et al. (2014) proposed a probabilistic speed-density relationship based on a microscopic car-following model. Probabilistic features are incorporated by introducing random parameters capturing population heterogeneity. However, limited behavioral basis exists to help in the specification of the distribution of these parameters. Also, such approach has the disadvantage of potentially allowing for behaviorally and physically implausible parameter values, depending on the choice of distributions. An alternative for dealing with heterogeneity is a two-stage approach, where the data is first segmented based on some observed characteristics (e.g. socio-economic or demographic variables). The segmentation can be performed using automated clustering schemes (Ge et al., 2012; Lee et al., 2007) or manually. The assignment of the individual observations to different segments is deterministic in this approach. In the second stage, a separate model is estimated for each predefined segment in the population (Weidmann, 1993). The issue of imprecise parameter estimates may arise due to potentially small sample sizes in some segments. Also, segmentation is usually performed based on a single characteristic and assumed to be error-free. In reality, the heterogeneity may come from multiple factors, which may introduce errors in the second stage.

2.4 Summary

Various forms of data have been collected for the purpose of studying pedestrian flow and behavior. The main representatives include survey, count and trajectory data. Each of them is related to certain advantages and disadvantages. In most of the studies, typically only one type of data is used. The studies in which multiple data types have been combined are extremely rare. The most detailed information on pedestrian motion is contained in trajectory data. However, in many cases, cost and privacy issues do not allow the use of high precision sensors covering an entire pedestrian infrastructure.

From the above literature review, it can be concluded that the characterization of pedestrian movements that is (i) independent of arbitrarily chosen values, in both space and time, (ii) capable to reflect the heterogeneity of the population, (iii) in accordance with multi-directional pedestrian flows, and (iv) applicable to continuous and discrete pedestrian trajectory data, is still missing. Also, the existing models of speed-density relationships for pedestrian traffic are not designed to capture the complex aspects related to high scattering. This is where our study makes a contribution.

We propose methodologies for pedestrian flow characterization and modeling that utilize

the potential of the data itself. The approach for the definition of fundamental traffic variables proposed in Chapter 3 is data-driven. It relies on the microscopic definitions of Edie (1963), adopted for pedestrian traffic, and extended through a data-driven discretization framework. We propose an alternative approach to account for the scatter in pedestrian fundamental relationships, as observed in the data. First, a probabilistic model of speed-density relationship is developed in Chapter 4. The model implicitly accounts for the heterogeneity of pedestrian flows by bringing together first principles and a data-inspired approach. The model specification ensures the physical correctness of the results, but lacks behavior-oriented explanatory power. The second model, proposed in Chapter 5, deals with this issue by relying on the latent class modeling (LCM) methodology (Frühwirth-Schnatter, 2006). The LCM approach has been proven to be valuable in capturing unobserved heterogeneity and characterization of the latent classes (Walker and Li, 2007).

3

Data-driven spatio-temporal discretization for pedestrian flow characterization

This chapter is based on the article:

Nikolić, M. and Bierlaire, M. (2016). Data-driven characterization of pedestrian flows. Technical report TRANSP-OR 160815.

It has been accepted for publication in the proceedings and a special issue of an academic journal for the 22nd International Symposium on Transportation and Traffic Theory.

The work has been performed by the candidate under the supervision of Prof. Michel Bierlaire.

In this chapter, we propose a novel approach to pedestrian flow characterization. The definitions of density, flow and velocity existing in the literature are extended through a spatio-temporal discretization framework that is independent from arbitrarily chosen values. We propose to adjust the discretization to the data itself using a data-driven approach. The approach is based on spatio-temporal Voronoi diagrams designed through the utilization of pedestrian trajectories. The methodology proposed here relies on the microscopic definitions of Edie (1963) adopted for pedestrian traffic, and extended through a data-driven discretization.

The structure of the chapter is as follows. Section 3.1 provides a formal introduction of the basic elements involved in our analysis. Section 3.2 describes the proposed methodology for the derivation of the spatio-temporal discretization framework. Based on this framework, we derive the definitions of the pedestrian traffic variables, that is density, flow and velocity. Section 3.3 empirically illustrates the performance of the approach by using synthetic data. Finally, Section 3.4 summarizes the outcomes of the proposed methodology.

3.1 Preliminaries

We consider a space-time representation and denote the area of interest by $\Omega \subset \mathbb{R}^3$. An orthonormal basis of this space is considered. The distance along each of the two spatial axes is expressed in meters, and the unit for time is seconds. The triplet $p = (p_x, p_y, p_t) = (x, y, t) \in \Omega$ represents a physical position (x, y) in space at a specific time t. It is assumed that Ω is convex, that is obstacle-free, and bounded.

The trajectory of pedestrian i is a curve in space and time. It is a set of points

$$\Gamma_i : \{p_i(t) | p_i(t) = (x_i(t), y_i(t), t)\}, \tag{3.1}$$

indexed by time t that spans the horizon of the analysis, and $x_i(t)$ and $y_i(t)$ are the coordinates of the position of pedestrian i at time t.

In practice, the analytical description of a trajectory is seldom available. Instead, the pedestrian trajectory data is collected through an appropriate tracking technology (Alahi et al., 2014; Daamen and Hoogendoorn, 2003). In this case time is discretized and the trajectory is described as a finite collection of triplets (a sample of points)

$$\Gamma_i : \{p_{is} | p_{is} = (x_{is}, y_{is}, t_s)\}, \tag{3.2}$$

where $s = [1, 2, ..., T_i]$ and $t_s = [t_1, t_2, ..., t_{T_i}]$ correspond to the available sample.

The speed along the continuous trajectory of pedestrian i is given by

$$v_i(t) = (x_i'(t), y_i'(t), 1). \tag{3.3}$$

Interpolation methods or finite differences (forward, backward or central) approximation can be used with sampled data.

3.2 Methodology

This section is organized in four parts. The first part presents the derivation of the spatio-temporal discretization framework using a data-driven approach. In the second part, we define pedestrian traffic variables, that is density, flow and velocity. The variables are defined by revising the existing microscopic definitions according to the proposed discretization (as motivated in Chapter 2). In the third part, we present concrete suggestions for the operationalization of the general and abstract concepts related to the discretization framework. The last part focuses on the implementation details.

3.2.1 Data-driven discretization

We propose the discretization in space and time that is defined based on three-dimensional (3D) Voronoi diagrams associated with pedestrian trajectories. We call the set of trajectories the generator set $\Gamma = \{\Gamma_1, ..., \Gamma_n\}$, consistently with the literature. We assume that elements in Γ do not intersect each other. This assumption is reasonable, as two pedestrians cannot be at the exact same place at the exact same time. The main idea for defining the partition of Ω is that (i) every point $p \in \Omega$ belongs to a unique discretization unit, (ii) each discretization unit is assigned to one generator $\Gamma_i \in \Gamma$ according to a certain assignment rule and (iii) the resulting discretization units associated with the trajectories are collectively exhaustive and mutually exclusive. Therefore, the partitioning is characterized by the assignment of each point $p \in \Omega$ to one generator from Γ. The discretization units are then defined as the set of points p assigned to the same generator.

Given a non-empty space Ω and a generator set Γ, the assignment rule δ_Γ of a point $p \in \Omega$ to an element of Γ is in the literature (Okabe et al., 2000) often specified in terms of distance relations D (not necessary distance metric). The point p is assigned to the "closest" generator in term of a given distance:

$$\delta_\Gamma(p, \Gamma_i) = \begin{cases} 1, & D(p, \Gamma_i) \leq D(p, \Gamma_j), \forall j \neq i \\ 0, & \text{otherwise.} \end{cases} \tag{3.4}$$

Note that this rule is ambiguous for points p that are equidistant to two trajectories. In this case, an additional arbitrary rule must be used. For instance, if $D(p, \Gamma_i) = D(p, \Gamma_j)$, then it can be decided that p is assigned to Γ_i if $i \leq j$.

If the generators are continuous trajectories (3.1), the distance may be defined as

$$D(p, \Gamma_i) = \min_t \{d(p, p_i(t)) | p_i(t) \in \Gamma_i, \Gamma_i \in \Gamma, p \in \Omega\}, \tag{3.5}$$

where $d(p, q)$ is the distance between two points p and q in Ω. Concrete examples of this

distance function are discussed in Section 3.2.3. Similarly, if the generators are sampled (3.2), the distance may be defined as

$$D(p, \Gamma_i) = \min_s \{d(p, p_{is}) | p_{is} \in \Gamma_i, \Gamma_i \in \Gamma, p \in \Omega\}. \tag{3.6}$$

Under the assignment rule (3.4), we consider the set of points V_i assigned to Γ_i

$$V_i = \{p | \delta_\Gamma(p, \Gamma_i) = 1, p \in \Omega, \Gamma_i \in \Gamma\}, \tag{3.7}$$

which represents a personal spatio-temporal region associated with pedestrian i. For each i, V_i is a convex subset of Ω called a Voronoi cell. Collectively, they represent a Voronoi diagram. The assumption that Ω is obstacle-free and bounded (Section 3.1) allows for the creation of non-degenerate Voronoi diagrams

$$\mathcal{V} = \{V_1, ..., V_n\}, \tag{3.8}$$

generated by Γ.

In a three-dimensional space Ω the plane through the point $p_0 = (x_0, y_0, t_0)$ and with non-zero normal vector $\vec{n} = (a, b, c)$ has equation

$$\mathcal{P}_{\vec{n},p_0} : ax + by + ct + d = 0, \tag{3.9}$$

where $d = -ax_0 - by_0 - ct_0$. We define the set of points $A(V_i, \mathcal{P}_{\vec{n},p_0})$ corresponding to the intersection of the cell V_i and the plane $\mathcal{P}_{\vec{n},p_0}$

$$A(V_i, \mathcal{P}_{\vec{n},p_0}) = \{p | p \in \{V_i \cap \mathcal{P}_{\vec{n},p_0}\}\}. \tag{3.10}$$

For $\vec{n} = (0, 0, 1)$ we have a plane parallel to the x-y plane and its intersection with V_i is given as

$$A(V_i, \mathcal{P}_{(0,0,1),p_0}) = \{p | p \in V_i \text{ and } p_t = t_0\}. \tag{3.11}$$

It represents a set of dimension 2 or a physical area on the floor (illustrated in Figure 3.1 (left)), at time t_0. The area of this cell is denoted by $|A(V_i, \mathcal{P}_{(0,0,1),p_0})|$, with the unit in m^2. Similarly, for $\vec{n} = (a, b, 0)$ we have

$$A(V_i, \mathcal{P}_{(a,b,0),p_0}) = \{p | p \in V_i \text{ and } ap_x + bp_y = ax_0 + by_0\}. \tag{3.12}$$

It is the set of dimension 2 or a segment on the floor occupied by pedestrian i in the direction perpendicular to $\vec{n} = (a, b, 0)$ during the time interval spanning V_i. The area of the cell is denoted by $|A(V_i, \mathcal{P}_{(a,b,0),p_0})|$, with the unit in ms. Note that if $\vec{n} = (1, 0, 0)$ and $\vec{n} = (0, 1, 0)$, the corresponding planes are parallel to the x-t, respectively the y-t plane (illustrated in Figure 3.1 (right)).

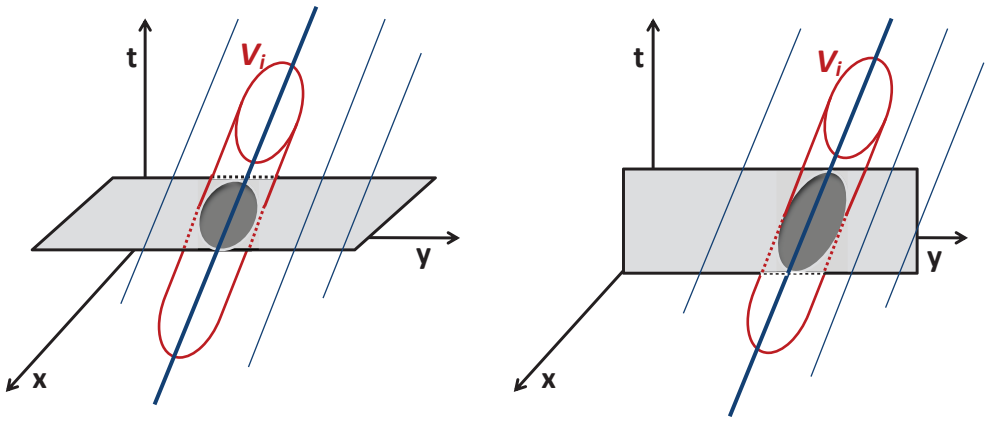


Figure 3.1: 3D Voronoi-based discretization: Set $A(V_i, \mathcal{P}_{(0,0,1),p_0})$ (left); Set $A(V_i, \mathcal{P}_{(a,b,0),p_0})$ (right)

3.2.2 Definitions of pedestrian traffic variables

Assume that the Voronoi cell V_i is associated with position $(x, y, t) \in \Omega$. The density at (x, y, t) is defined as the inverse of the area of the set $A(V_i, \mathcal{P}_{(0,0,1),(x,y,t)})$ assigned to pedestrian i at time t

$$k(x, y, t) = \frac{1}{|A(V_i, \mathcal{P}_{(0,0,1),(x,y,t)})|}, \tag{3.13}$$

where $A(V_i, \mathcal{P}_{(0,0,1),(x,y,t)})$ is given by (3.11) and $|A(V_i, \mathcal{P}_{(0,0,1),(x,y,t)})|$ is the area of this set. The location (x, y) determines pedestrian i and the corresponding set $A(V_i, \mathcal{P}_{(0,0,1),(x,y,t)})$. The unit of $k(x, y, t)$ is a number of pedestrians per square meter. This definition is consistent with (2.11), adapted to this 3D Voronoi context.

The flow in the direction determined by the vector $\vec{e} = (a, b)$ is defined as the inverse of the area of the set $A(V_i, \mathcal{P}_{(a,b,0),(x,y,t)})$ assigned to pedestrian i

$$\vec{q}_e(x, y, t) = \frac{1}{|A(V_i, \mathcal{P}_{(a,b,0),(x,y,t)})|}, \tag{3.14}$$

where $A(V_i, \mathcal{P}_{(a,b,0),(x,y,t)})$ is given by (3.12) and $|A(V_i, \mathcal{P}_{(a,b,0),(x,y,t)})|$ is the area of this set. The point (x, y, t) determines pedestrian i and the corresponding set $A(V_i, \mathcal{P}_{(a,b,0),(x,y,t)})$. The set $(V_i, \mathcal{P}_{(a,b,0),(x,y,t)})$ belongs to the spatio-temporal domain. Therefore, the unit of $\vec{q}_e(x, y, t)$ is a number of pedestrians per meter per second. The flow in x and y directions is obtained considering the inverse of the areas $|A(V_i, \mathcal{P}_{(1,0,0),(x,y,t)})|$ and $|A(V_i, \mathcal{P}_{(0,1,0),(x,y,t)})|$, consistently with (2.12).

Adopting the usual definition of the velocity (the ratio between the flow and density), from (3.13) and (3.14) we have

$$\vec{v}_e(x, y, t) = \frac{\vec{q}_e(x, y, t)}{k(x, y, t)} = \frac{|A(V_i, \mathcal{P}_{(0,0,1),(x,y,t)})|}{|A(V_i, \mathcal{P}_{(a,b,0),(x,y,t)})|}. \tag{3.15}$$

It represents the mean speed of pedestrian i (Jabari et al., 2014) in the direction determined by the vector $\vec{e} = (a, b)$, expressed in meters per second. Note that the framework allows for the specification and measurement of the indicators $\vec{q}_e(x, y, t)$ and $\vec{v}_e(x, y, t)$ in any direction $\vec{e}$ of interest.

The proposed definitions (3.13)-(3.15) are independent from arbitrarily chosen intervals in space and time, due to the fact that they rely on a data-driven discretization. Our approach is microscopic and therefore suitable for multi-directionl nature of pedestrian flows and able to preserve the heterogeneity of pedestrians (Jabari et al., 2014). The issues characteristic for the macroscopic approach, discussed in Chapter 2, do not appear at the microscopic level. Also, the proposed definitions can be applied on continuous or sampled trajectories. The samples, in general, do not need to be synchronized.

We refer to the proposed approach as the *3D Voronoi characterization* (3DVoro).

3.2.3 Spatio-temporal distances

The proposed framework is fairly general, and can accommodate various methods to generate the Voronoi diagrams, based on different definitions of the distance between two points. To construct 3D Voronoi diagrams (Section 3.2.1), we need to define the exact form of the distance relation d used in (3.5) and (3.6). Applying the Euclidean distance in $\mathbb{R}^3$ looks like a natural choice. However, it is important to keep in mind that it would mix units in square meters with units in seconds. We propose here several ways to deal with it. First, we propose to restrict the Euclidean distance in the spatial dimension, and consider each point in time as independent. Second, we propose distances in $\mathbb{R}^3$ that convert seconds into meters using speed. Third, we account for the pedestrian dynamics to define the distance, anticipating his future position. Finally, we define a distance through the identification of points that are equidistant. Their respective performance is empirically evaluated in Section 3.3.

We denote by $p = (x, y, t)$ a point from Ω. An observation from the trajectory of pedestrian i is denoted by $p_i = (x_i, y_i, t_i)$. It refers to either $p_i(t)$ or p_{is}, depending on the context.

3.2.3.1 Spatial Euclidean distance

The first distance that we propose is defined with respect to the standard Euclidean distance in the spatial dimension, that is

$$d_E(p, p_i) = \begin{cases} \sqrt{(x - x_i)^2 + (y - y_i)^2}, & t = t_i \\ \infty, & \text{otherwise.} \end{cases} \tag{3.16}$$

Intuitively, each point in time is independent. This is motivated by the availability of snapshots of the floor area at given points in time. This implies that all pedestrians in the area must be observed at the exact same time.

We refer to the characterization obtained using this distance as the *Euclidean 3D Voronoi characterization* (E-3DVoro).

3.2.3.2 Time-Transform distances

We define the set of three distances that apply a conversion parameter, expressed in meters per second, to transform the temporal difference between the points into the spatial one. They are denoted as the Time-Transform distances ($d_{TT_1}, d_{TT_2}, d_{TT_3}$). The distances differ in terms of the choice of the conversion parameter and in the way of coupling the spatial and temporal component. They are defined as

$$d_{TT_1}(p, p_i) = \sqrt{(x - x_i)^2 + (y - y_i)^2 + v^2(t - t_i)^2}, \tag{3.17}$$

where v is a parameter representing the typical speed of pedestrians (a value of v = 1.34 m/s is used in our experiments, Weidmann (1993)),

$$d_{TT_2}(p, p_i) = \sqrt{(x - x_i)^2 + (y - y_i)^2 + \hat{v}_i(t_i)^2(t - t_i)^2}, \text{ and} \tag{3.18}$$

$$d_{TT_3}(p, p_i) = \sqrt{(x - x_i)^2 + (y - y_i)^2} + \hat{v}_i(t_i)|t - t_i|, \tag{3.19}$$

where $\hat{v}_i(t_i)$ is the speed at time t on trajectory Γ_i. The choice of the conversion parameter $\hat{v}_i(t_i)$ in (3.18) and (3.19) allows to treat moving pedestrians in a different way than standing pedestrians.

The distances (3.17) and (3.18) combine the spatial and temporal components based on the Euclidean norm, using two different values for the speed. In (3.19), the components are considered as independent and kept separately. The distance d_{TT_3} is defined as a weighted sum of two norms. When $t = t_i$, all distances are equivalent to (3.16).

We refer to the characterization obtained using these three distances as the *Time-Transform 3D Voronoi characterization* (TT_1-3DVoro, TT_1-3DVoro, TT_1-3DVoro).

3.2.3.3 Predictive distance

The Predictive distance anticipates the forward movement of pedestrians. The anticipated positions x_i^a and y_i^a are extrapolated from the current velocities of pedestrians for a time determined by the anticipation time $t - t_i$

$$x_i^a = x_i^a(t) = x_i + (t - t_i)v_i^x(t_i), \tag{3.20}$$

$$y_i^a = y_i^a(t) = y_i + (t - t_i)v_i^y(t_i), \tag{3.21}$$

where $v_i^x(t_i)$ and $v_i^y(t_i)$ are the speed of pedestrian i at t_i in x, respectively y, direction.

The distance is specified as

$$d_P(p, p_i) = \begin{cases} \sqrt{(x_i^a - x)^2 + (y_i^a - y)^2}, & t - t_i \geq 0 \\ \infty, & \text{otherwise.} \end{cases} \tag{3.22}$$

Note that it is not a metric distance, as it is not symmetric. The anticipation time extends from zero to a positive value $(t - t_i)$. Points p that are backward in time with respect to the current positions of pedestrian are considered infinitely distant. When $t = t_i$, the distance reduces to the standard $\mathbb{R}^2$ Euclidean distance. The consideration of individual speeds allows for the distinction between pedestrians that perform movement from those that stand.

We refer to the characterization obtained using this distance as the *Predictive 3D Voronoi characterization* (P-3DVoro).

3.2.3.4 Mahalanobis distance

The Mahalanobis distance is specified as

$$d_M(p, p_i) = \sqrt{(p - p_i)^T M_i (p - p_i)}, \tag{3.23}$$

where M_i is a change of variable matrix. It is a symmetric, positive-definite matrix, which defines how the distances are measured in different spatio-temporal directions from the perspective of pedestrian i. To implement this distance we need to determine the matrix M_i. We do so by identifying 6 points in Ω such that they are equidistant to p_i for the Mahalanobis distance. We take into account the information about the speed and direction of pedestrians, in the sense that the points that are in the movement direction of a pedestrian are "closer" than the points from other directions.

Formally, we consider three directions of interest. First, we define the normalized direction of movement in the space-time dimensions

$$d^1(t_i) = \frac{v_i(t_i)}{||v_i(t_i)||}, ||d^1(t_i)|| = 1, \tag{3.24}$$

where $v_i(t_i)$ is the speed along the trajectory of pedestrian i given by (3.3). We next define a normalized spatial direction orthogonal to $d^1(t_i)$, that is

$$d^2(t_i) = \begin{pmatrix} d_x^1(t_i) \\ d_y^2(t_i) \\ 0 \end{pmatrix}, \tag{3.25}$$

such that $d^1(t_i)^T d^2(t_i) = 0$ and $||d^2(t_i)|| = 1$. The third direction is for time

$$d^3(t_i) = \begin{pmatrix} 0 \\ 0 \\ \Delta t \end{pmatrix}, \tag{3.26}$$

where Δt is typically determined by the sampling frequency and $||d^3(t_i)|| = \Delta t$.

We determine the matrix M_i, and the distance d_M, such that the following points in the defined directions are all at distance α from the point p_i. The key feature is that, in the direction of movement, the distances do not refer to the position at time t, but the positions at time $t + \Delta t$ and $t - \Delta t$. The points S_1 and S_2 in the d^1 direction are at α and $-\alpha$ from the positions at time $t + \Delta t$, respectively $t - \Delta t$

$$S_1(t_i, \alpha) = p_i + \Delta t v_i(t_i) + \alpha d^1(t_i), \tag{3.27}$$

$$S_2(t_i, \alpha) = p_i - \Delta t v_i(t_i) - \alpha d^1(t_i). \tag{3.28}$$

In the direction d^2 we consider the point S_3 that is at α from the point p_i

$$S_3(t_i, \alpha) = p_i + \alpha d^2(t_i), \tag{3.29}$$

and the point S_4 that is at $-\alpha$ from the point p_i

$$S_4(t_i, \alpha) = p_i - \alpha d^2(t_i). \tag{3.30}$$

Similarly, in time direction d^3 we consider the point S_5 that is at α from the point p_i

$$S_5(t_i, \alpha) = p_i + \alpha d^3(t_i), \tag{3.31}$$

and the point S_6 that is at $-\alpha$ from the point p_i

$$S_6(t_i, \alpha) = p_i - \alpha d^3(t_i). \tag{3.32}$$

This is illustrated in Figure 3.2.

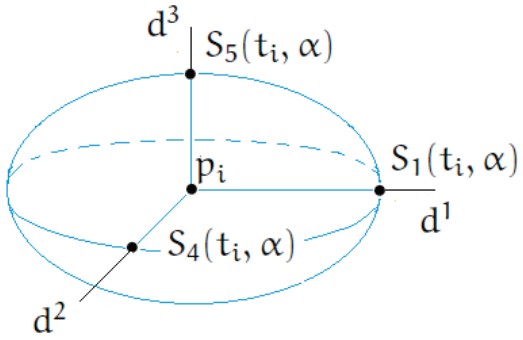


Figure 3.2: Mahalanobis distance - illustration

In standard Euclidean space we have that

$$||S_1(t_i, \alpha) - p_i|| = ||S_2(t_i, \alpha) - p_i|| = \Delta t||v_i(t_i)|| + \alpha. \tag{3.33}$$

It shows that, in the direction d^1, the forward and backward distances are stretched by the quantity $\Delta t||v_i(t_i)||$. This is designed to anticipate the movement of pedestrians. The additional term vanishes when $\Delta t \to 0$. The approach also allows to deal with moving pedestrians and standing pedestrians in a different way. The distance in the d^2 direction is consistent with Euclidean distance

$$||S_3(t_i, \alpha) - p_i|| = ||S_4(t_i, \alpha) - p_i|| = ||\alpha d^2(t_i)|| = \alpha. \tag{3.34}$$

The distance in time direction d^3 is proportional to the time discretization

$$||S_5(t_i, \alpha) - p_i|| = ||S_6(t_i, \alpha) - p_i|| = ||\alpha d^3(t_i)|| = \alpha \Delta t. \tag{3.35}$$

In particular, it shrinks to zero when $\Delta t \to 0$.

In our experiments, a value of $\alpha = 1$ is used.

We refer to the characterization obtained using this distance as the *Mahalanobis 3D Voronoi characterization* (M-3DVoro).

3.2.4 Implementation details

The Voronoi diagrams for higher order generators (such as curves) in higher dimensions are difficult to compute in an exact way (Hoff III et al., 1999). For a number of applications an approximation of the exact diagram is considered (Schueller, 2007; Rong and Tan, 2007; Park et al., 2006; Fuchida et al., 2005).

We also use an approximate algorithm for the implementation of the Voronoi diagram detailed in Section 3.2.1. The area of interest, Ω, is discretized into regular cells. The cells are used in the assignment rule (3.4), instead of points p. In the literature, this approximation is known as the Naïve algorithm (Van der Putte, 2009).

The presented algorithm is merely one possibility. It is intuitive, but does not feature high computational efficiency. To improve the efficiency of the algorithm, specialized data structures can be considered (e.g. kd-tree), as discussed in Rigaux et al. (2001). However, the aim of our study is to illustrate the performance of the proposed methodology, and not to contribute to the field of computational geometry. The analysis of other algorithms to construct Voronoi diagrams is therefore out of the scope of this thesis.

3.3 Empirical analysis

The performance of our approach is evaluated based on synthetic data that is generated using the NOMAD simulation tool (Campanella et al., 2014; Hoogendoorn and HL Bovy, 2003). The flow is simulated in a 4 meters by 4 meters area, during 10 seconds for unidirectional (Uni) flow composition. The data is generated for two different scenarios. In the first scenario, we consider lower demand (1.2 pedestrians per second) and homogeneous pedestrian population ($LD - HomoPop$). The homogeneity of the population is reflected through (approximately) homogenous walking speed of pedestrians. We use the average speed value of 1.34 m/s, according to the study of Weidmann (1993). In the second scenario, we consider higher demand (3.6 pedestrians per second) and heterogeneous pedestrian population ($HD - HeteroPop$). To represent the heterogeneity in the population, we consider three sub-populations (slow, average and fast) with respective speeds of 0.5 m/s, 1.3 m/s and 2.1 m/s. Each sub-population has roughly the same size.

Our objective is to analyze the nature of the results obtained using our approach, and the robustness of the approach with respect to the simulation noise and with respect to the sampling frequency. The performance of the 3DVoro method (for all the distances)

is compared with the performance of the XY-T and VB methods. The consideration of these two methods is motivated by the empirical comparisons of the existing approaches presented in Duives et al. (2015), where it has been concluded that the XY-T and VB methods perform the best. In the application of the XY-T method, the parameters of the cells A reported in Duives et al. (2015) are used: a time interval of 1 second, and a grid cell size of 1×1 meter. Note that, the VB method corresponds to the E-3DVoro when used for the discrete time instants to discretize the spatial dimension only. Therefore, the VB method will not be considered separately.

3.3.1 Nature of the results

The aim of this section is to assess the effects of the discretization on the nature of the corresponding values of traffic variables. For this purpose we randomly select points in space where density, velocity and flow are measured across time. The observed pattern in the measures is similar across the selected points. It is illustrated for one specific point in Figure 3.3.

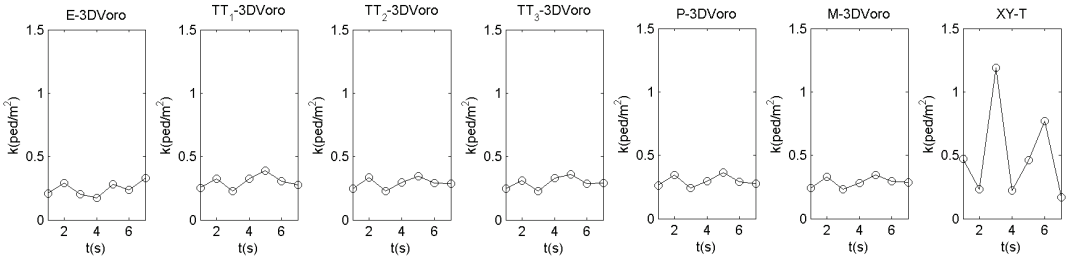
A discrete nature of the measures can be observed in case of the XY-T method. This method leads to large fluctuations, in particular for density and flow variables. The jumps occur due to the fact that these variables are directly linked to the number of pedestrians, and entering or exiting of the XY-T discretization units by pedestrians affects the indicators considerably.

The 3DVoro method for all the distances lead to similar results. Density and flow measures do not change abruptly. They are characterized by smoother transitions, compared to the XY-T method. Also, the microscopic nature of the approach allows to correlate the momentary speed of an individual pedestrian with the availability of space (as illustrated Figure 3.3).

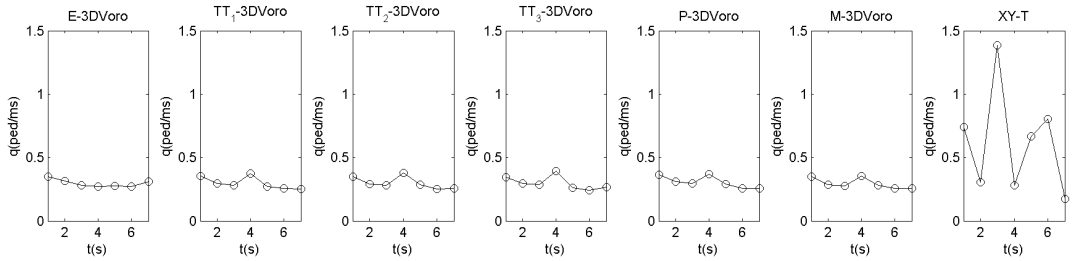
3.3.2 Robustness with respect to the simulation noise

In this section we analyze the performance of the approach when frequent pedestrian observations are available. For this purpose we consider synthetic trajectories obtained for minimal trajectory step (0.1 second). In the case of the XY-T method, the distances traveled and the times spent by pedestrians in regions A (Chapter 2) are obtained using interpolation.

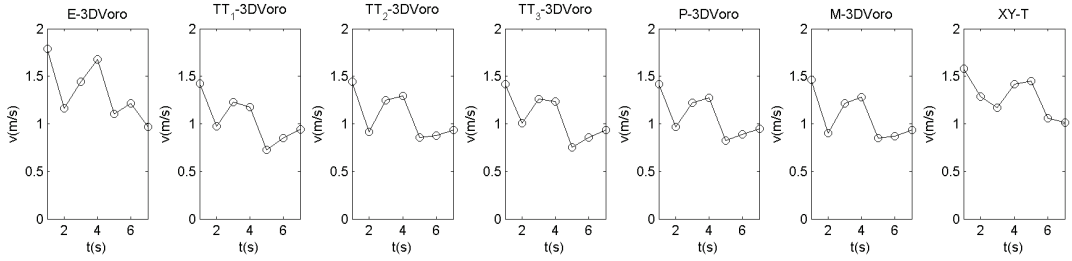
We synthesize 100 sets of pedestrian trajectories for each scenario and evaluate the variance of the indicators across these replications. The described settings of the simulator remain unchanged for a given scenario. The indicators (k, v, q) are calculated for each set of the trajectories via 3DVoro (for all the distances) and the XY-T method. The



(a) Density indicator



(b) Flow indicator



(c) Velocity indicator

Figure 3.3: Nature of the results - $Uni_{LD-HomoPop}$

methods are compared based on the standard deviation of the indicators at specific points due to simulation noise.

Let M represent the method (3DVoro or XY-T), r a realization of NOMAD simulation ($r = 1, ..., 100$) and p a point from Ω. We denote by $\theta_r^M(p) = (k_r^M(p), v_r^M(p), q_r^M(p))$ a vector of indicators at point p obtained by applying the method M to the r^{th} set of trajectories. For each method we calculate the standard deviation of the indicators at p as

$$\sigma_R^M(p) = \sqrt{\frac{1}{R}\sum_{r=1}^{R}(\theta_r^M(p) - \mu_R^M(p))^2}, \tag{3.36}$$

where $\mu_R^M(p) = \frac{1}{R}\sum_{r=1}^{R}\theta_r^M(p)$ and $R = 100$. This procedure is repeated for 1000 randomly selected points p (these points are the same across simulation). The results are reported using boxplot representation in Figure 3.4 for $Uni_{LD-HomoPop}$, and in Figure 3.5 for $Uni_{HD-HeteroPop}$.

The standard deviations of the indicators are larger for $Uni_{HD-HeteroPop}$ for all the methods, compared to $Uni_{LD-HomoPop}$. This can be explained by the larger changes in the data due to higher complexity of the system in $Uni_{HD-HeteroPop}$. However, a similar trend is noticeable in the results for both scenarios.

The results for the density indicator (Figure 3.4a and Figure 3.5a), and for the flow indicator (Figure 3.4b and Figure 3.5b), are similar across different 3DVoro distances for both scenarios. In the case of speed, E-3DVoro exhibits larger standard deviations as compared to other 3DVoro methods for $Uni_{LD-HomoPop}$ (Figure 3.4c). This suggests that 3DVoro with distances that account for the speed and/or movement direction of pedestrians lead to lower variance in the results when demand is low. When demand is higher (Figure 3.5c), higher number of pedestrian trajectories leads to similar behavior of different distances, and consequently to similar results.

In comparison to 3DVoro approach, in the XY-T method the simulation noise is significantly amplified. The results suggest that the changes of trajectories, even though for the same simulated settings, dramatically affect the measured indicators of the XY-T method.

3.3.3 Robustness with respect to the sampling frequency

In order to evaluate the effectiveness of the approach when sampled data is available, we consider the samples of points from the synthetic trajectories. The samples are obtained using different sampling frequencies (3 s^{-1}, 2 s^{-1}, 1 s^{-1} and 0.5 s^{-1}), as illustrated in Figure 3.6. The feature of interest is the robustness of the approach with respect to the sampling frequency. That is, the ability of the approach to produce stable results even in the lack of frequent observations.

We can deal with sampled data in two ways. First, we generate trajectories using linear interpolation (Figure 3.6). The speed along trajectories is approximated, using finite differences. The indicators are then obtained via 3DVoro applied to the interpolated trajectories. In the second case we apply 3DVoro directly to the sampled data. The indicators calculated on the true synthetic trajectories, generated using minimal step, are used as a benchmark. The indicators obtained using samples and interpolated trajectories are compared at 1000 randomly selected points to the corresponding benchmark values. These points are the same across all the methods.

We list in Table 3.1 and Table 3.2 the statistics (mean, mode, median and 90%-quantile) corresponding to resulting differences in the case of density indicator, for $Uni_{LD-HomoPop}$, respectively $Uni_{HD-HeteroPop}$. The statistics show similar trends for velocity and flow indicators, as illustrated in Appendix A (Table A.1 - Table A.4). To demonstrate the performance, we show the results corresponding to the extreme values of the considered sampling frequencies (3 s^{-1} and 0.5 s^{-1}). In the tables, IT refers to the value of a given statistic obtained based on interpolated trajectories; SoP refers to the value of a given statistic obtained based on sample of points. Gray color of the cells in the tables is used to indicate the overall best value of the considered statistic. Note that in this case the E-3DVoro and the XY-T method can be applied only when the points from samples are interpolated.

In general, the 3DVoro method outperforms the XY-T method. The lowest differences between the indicators calculated based on sample of points or interpolated trajectories and the benchmark vales are achieved using the 3DVoro method.

The interpolation appears to be a better choice for 3DVoro when the sampling frequency is high, in both scenarios. This is expected, given that more data points used for interpolation yield lower interpolation error. In this case, the Time-Transform distances lead to the best performance of the 3DVoro approach, and in particular TT_1-3DVoro.

When the sampling frequency is low, 3DVoro applied directly to the sample is associated with the best effectiveness. In $Uni_{LD-HomoPop}$, the distances that take into account the speed and/or direction of pedestrians (in particular TT_2-3DVoro, P-3DVoro and M-3DVoro) are the most satisfactory. In $Uni_{HD-HeteroPop}$, the preferred characterization is based on the Time-Transform distances (particularly TT_1-3DVoro).

We have also analyzed the robustness to the sampling frequency for bi-directional flow composition and for both scenarios, $Bi_{LD-HomoPop}$ and $Bi_{HD-HeteroPop}$. Table 3.3 and Table 3.4 report the statistics corresponding to resulting differences in the case of density indicator. The statistics show similar trends as for the uni-directional flow composition.

In summary, the *Time-Transform 3D Voronoi characterization* is the most robust with respect to the sampling frequency when more data is available (the sampling frequency equal to 3 s^{-1} or the demand equal to 3.6 pedestrians per second). When less data is available (the sampling frequency equal to 0.5 s^{-1} and the demand equal to 1.2 pedestrians per second), the distances accounting for the dynamics of pedestrians lead to the best robustness with respect to the sampling.

Method	Mean		Mode		Median		90% quantile	
	IT	SoP	IT	SoP	IT	SoP	IT	SoP
XY-T	$1.47e^{-02}$	/	$1.25e^{-02}$	/	$1.25e^{-02}$	/	$6.25e^{-02}$	/
E-3DVoro	$1.17e^{-02}$	/	0	/	$4.48e^{-04}$	/	$3.96e^{-02}$	/
TT_1-3DVoro	$2.70e^{-03}$	$6.70e^{-03}$	0	0	$3.00e^{-04}$	$2.30e^{-03}$	$7.30e^{-03}$	$1.02e^{-02}$
TT_2-3DVoro	$5.80e^{-03}$	$3.50e^{-02}$	0	$2.80e^{-03}$	$6.00e^{-04}$	$2.08e^{-02}$	$1.50e^{-02}$	$6.69e^{-02}$
TT_3-3DVoro	$5.40e^{-03}$	$4.34e^{-02}$	0	$8.00e^{-03}$	$6.00e^{-04}$	$2.83e^{-02}$	$1.32e^{-02}$	$9.22e^{-02}$
P-3DVoro	$8.20e^{-03}$	$5.36e^{-02}$	0	$6.10e^{-03}$	$2.40e^{-03}$	$3.03e^{-02}$	$1.30e^{-02}$	$1.14e^{-01}$
M-3DVoro	$4.50e^{-03}$	$5.65e^{-02}$	0	$6.80e^{-03}$	$1.10e^{-03}$	$4.55e^{-02}$	$1.28e^{-02}$	$1.04e^{-01}$

(a) Sampling frequency: 3 s^{-1}

Method	Mean		Mode		Median		90% quantile	
	IT	SoP	IT	SoP	IT	SoP	IT	SoP
XY-T	$1.90e^{-01}$	/	$1.00e^{-01}$	/	$1.50e^{-01}$	/	$3.38e^{-01}$	/
E-3DVoro	$1.64e^{-01}$	/	$1.12e^{-02}$	/	$1.46e^{-01}$	/	$3.02e^{-01}$	/
TT_1-3DVoro	$2.54e^{-01}$	$1.27e^{-01}$	$1.35e^{-02}$	$9.00e^{-03}$	$1.16e^{-01}$	$8.97e^{-02}$	$3.41e^{-01}$	$2.25e^{-01}$
TT_2-3DVoro	$1.64e^{-01}$	$1.22e^{-01}$	$1.44e^{-02}$	$1.06e^{-02}$	$1.21e^{-01}$	$7.30e^{-02}$	$3.52e^{-01}$	$2.33e^{-01}$
TT_3-3DVoro	$1.89e^{-01}$	$1.24e^{-01}$	$1.84e^{-02}$	$1.09e^{-02}$	$1.24e^{-01}$	$7.88e^{-02}$	$3.40e^{-01}$	$2.31e^{-01}$
P-3DVoro	$3.19e^{-01}$	$1.21e^{-01}$	$3.26e^{-02}$	$6.20e^{-03}$	$1.43e^{-01}$	$7.43e^{-02}$	$3.36e^{-01}$	$2.10e^{-01}$
M-3DVoro	$1.97e^{-01}$	$1.24e^{-01}$	$3.48e^{-02}$	$9.90e^{-03}$	$1.41e^{-01}$	$7.72e^{-02}$	$3.21e^{-01}$	$2.31e^{-01}$

(b) Sampling frequency: 0.5 s^{-1}

Table 3.1: Robustness to the sampling frequency of density indicator - $Uni_{LD-HomoPop}$

3.4 Summary

In this chapter a novel methodology for pedestrian traffic characterization is proposed. The definitions of pedestrian traffic variables that we have put forward are based on data-driven partitioning in space and time. As such, they resolve the issue of arbitrary selection of spatial discretization units (typical for XY-T, GB, RB) or spatial influence functions (typical for EW), and for the first time provide the way for data-driven temporal discretization. The discretization framework is designed via three-dimensional Voronoi diagrams directly generated from pedestrian trajectory data. It can be designed based on trajectories available either in the form of an analytical description or as a finite collection of points. The samples, in general, do not need to be synchronized. On the other hand, the existing methods (XY-T, GB, RB, EW, VB) can be applied on trajectories described either analytically or as a synchronized sample of points. The methodological framework is fairly general, and the exact characterization of the Voronoi diagrams can be adapted to specific situations. We have proposed different definitions of distances for the construction of the diagrams, and assessed them in quantitative terms. Also, the proposed definitions of the indicators are microscopic. They are therefore able to reflect the heterogeneity of pedestrians, and suitable for the multi-directional composition of pedestrian flows. Note that the methods XY-T, GB, RB and EW rely on macroscopic approach, which does not always comply with the nature of the underlying system (see Section 2.2.3).

The performance of the proposed approach is evaluated using synthetic data. It has been shown, for these datasets, that our approach outperforms the considered approaches from

Method	Mean		Mode		Median		90% quantile	
	IT	SoP	IT	SoP	IT	SoP	IT	SoP
XY-T	$2.05e^{-02}$	/	0	/	$1.25e^{-02}$	/	$5.00e^{-02}$	/
E-3DVoro	$1.43e^{-02}$	/	0	/	$2.67e^{-02}$	/	$2.64e^{-02}$	/
TT_1-3DVoro	$8.00e^{-03}$	$4.55e^{-02}$	0	0	$8.00e^{-04}$	$1.75e^{-02}$	$2.36e^{-02}$	$8.52e^{-02}$
TT_2-3DVoro	$1.49e^{-02}$	$1.07e^{-01}$	0	0	$3.20e^{-03}$	$5.72e^{-02}$	$3.33e^{-02}$	$2.21e^{-01}$
TT_3-3DVoro	$1.24e^{-02}$	$1.60e^{-01}$	0	0	$3.50e^{-03}$	$9.62e^{-02}$	$2.98e^{-02}$	$3.41e^{-01}$
P-3DVoro	$2.10e^{-02}$	$1.66e^{-01}$	0	0	$4.20e^{-03}$	$1.16e^{-01}$	$5.27e^{-02}$	$3.64e^{-01}$
M-3DVoro	$1.31e^{-02}$	$2.40e^{-01}$	0	0	$2.50e^{-03}$	$1.75e^{-01}$	$2.91e^{-02}$	$5.58e^{-01}$

(a) Sampling frequency: 3 s^{-1}

Method	Mean		Mode		Median		90% quantile	
	IT	SoP	IT	SoP	IT	SoP	IT	SoP
XY-T	$5.29e^{-01}$	/	$1.63e^{-01}$	/	$4.75e^{-01}$	/	$1.01e^{00}$	/
E-3DVoro	$4.02e^{-01}$	/	0	/	$2.49e^{-01}$	/	1.03E+00	/
TT_1-3DVoro	$4.06e^{-01}$	$2.90e^{-01}$	$3.10e^{-01}$	$2.48e^{-02}$	$2.64e^{-01}$	$1.65e^{-01}$	$9.21e^{-01}$	$7.12e^{-01}$
TT_2-3DVoro	$3.92e^{-01}$	$4.58e^{-01}$	$2.85e^{-01}$	$2.34e^{-01}$	$2.48e^{-01}$	$2.34e^{-01}$	$9.30e^{-01}$	1.11E+00
TT_3-3DVoro	$4.41e^{-01}$	$5.07e^{-01}$	$2.89e^{-01}$	$5.89e^{-02}$	$2.37e^{-01}$	$3.06e^{-01}$	$9.81e^{-01}$	1.17E+00
P-3DVoro	$4.31e^{-01}$	$3.71e^{-01}$	$1.40e^{-03}$	0	$2.58e^{-01}$	$1.80e^{-01}$	$9.43e^{-01}$	$7.29e^{-01}$
M-3DVoro	$4.34e^{-01}$	$5.01e^{-01}$	$3.16e^{-01}$	$1.36e^{-01}$	$2.75e^{-01}$	$3.52e^{-01}$	$9.96e^{-01}$	$9.80e^{-01}$

(b) Sampling frequency: 0.5 s^{-1}

Table 3.2: Robustness to the sampling frequency of density indicator - $Uni_{HD-HeteroPop}$

the literature (XY-T and VB), in terms of the smoothness of the results, the robustness to the simulation noise and the robustness with respect to the sampling frequency. As for the robustness to the simulation noise, 3DVoro with distances that account for the speed and/or movement direction of pedestrians lead to lower variance in the results when demand is low. When demand is higher, higher number of data leads to similar behavior of different distances. As for the robustness to the sampling frequency, when the sampling frequency is high, 3DVoro based on interpolated trajectories shows better results. When the sampling frequency is low, 3DVoro based on sample of points exhibit better performance. The analysis in the case of sampled data suggests that (i) when more data is available, either because of higher sampling frequency (3 s^{-1}) or higher demand (3.6 pedestrians per second), TT_1-3DVoro is the most robust with respect to the sampling frequency; (ii) when less data is available, due to lower sampling frequency (0.5 s^{-1}) and lighter traffic conditions (the demand of 1.2 pedestrians per second), the distances that account for the speed and the movement direction of pedestrians (TT_2-3DVoro, P-3DVoro and M-3DVoro) exhibit the best robustness.

A sensitivity analysis for the parameters affecting the characterization (parameters v and α in Section 3.2.3) is one direction of further investigation. Also, more research is needed to determine the performance of this approach in other scenarios and to understand its potential limitations.

Method	Mean		Mode		Median		90% quantile	
	IT	SoP	IT	SoP	IT	SoP	IT	SoP
XY-T	$6.50e^{-02}$	/	0	/	0	/	$8.65e^{-03}$	/
E-3DVoro	$1.20e^{-02}$	/	0	/	0	/	$4.66e^{-03}$	/
TT_1-3DVoro	$3.58e^{-03}$	$1.08e^{-02}$	0	0	0	$1.02e^{-03}$	$4.16e^{-03}$	$6.15e^{-03}$
$TT_2$3DVoro	$8.13e^{-03}$	$1.18e^{-02}$	0	0	0	$2.35e^{-03}$	$8.09e^{-03}$	$1.29e^{-02}$
TT_3-3DVoro	$1.49e^{-02}$	$2.06e^{-02}$	0	$3.91e^{-03}$	0	$8.43e^{-03}$	$7.46e^{-03}$	$3.10e^{-02}$
P-3DVoro	$2.29e^{-02}$	$5.42e^{-02}$	0	$1.94e^{-03}$	0	$2.72e^{-02}$	$9.25e^{-03}$	$1.06e^{-01}$
M-3DVoro	$2.15e^{-02}$	$4.82e^{-02}$	0	$4.31e^{-02}$	0	$2.42e^{-02}$	$7.69e^{-03}$	$1.29e^{-01}$

(a) Sampling frequency: 3 s^{-1}

Method	Mean		Mode		Median		90% quantile	
	IT	SoP	IT	SoP	IT	SoP	IT	SoP
XY-T	$1.66e^{-01}$	/	0	/	$6.84e^{-02}$	/	$7.00e^{-01}$	/
E-3DVoro	$1.65e^{-01}$	/	0	/	$1.19e^{-01}$	/	$3.40e^{-01}$	/
TT_1-3DVoro	$1.68e^{-01}$	$1.29e^{-01}$	$3.50e^{-02}$	$5.02e^{-02}$	$8.50e^{-02}$	$5.70e^{-02}$	$3.85e^{-01}$	$2.62e^{-01}$
TT_2-3DVoro	$1.70e^{-01}$	$1.02e^{-01}$	$4.52e^{-02}$	$5.63e^{-02}$	$8.49e^{-02}$	$6.15e^{-02}$	$3.82e^{-01}$	$5.57e^{-01}$
TT_3-3DVoro	$1.80e^{-01}$	$1.18e^{-01}$	$4.82e^{-02}$	$6.06e^{-02}$	$8.80e^{-02}$	$6.55e^{-02}$	$3.83e^{-01}$	$2.65e^{-01}$
P-3DVoro	$2.02e^{-01}$	$1.60e^{-01}$	$3.69e^{-02}$	$4.84e^{-02}$	$9.36e^{-02}$	$6.73e^{-02}$	$4.14e^{-01}$	$3.01e^{-01}$
M-3DVoro	$1.80e^{-01}$	$1.55e^{-01}$	$4.80e^{-02}$	$3.36e^{-02}$	$1.01e^{-01}$	$9.27e^{-02}$	$4.38e^{-01}$	$3.08e^{-01}$

(b) Sampling frequency: 0.5 s^{-1}

Table 3.3: Robustness to the sampling frequency of density indicator - $Bi_{LD-HomoPop}$

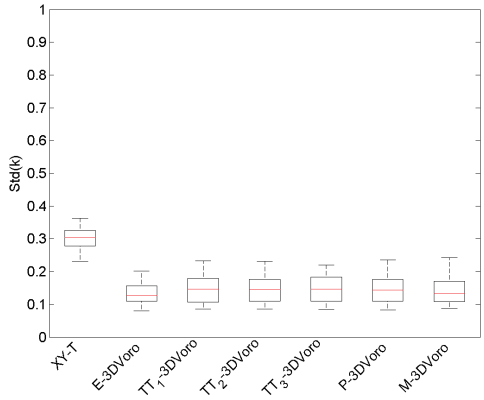
Method	Mean		Mode		Median		90% quantile	
	IT	SoP	IT	SoP	IT	SoP	IT	SoP
XY-T	$2.85e^{-02}$	/	0	/	$3.28e^{-03}$	/	$1.00e^{-01}$	/
E-3DVoro	$3.00e^{-02}$	/	0	/	$9.64e^{-03}$	/	$6.50e^{-02}$	/
TT_1-3DVoro	$1.15e^{-01}$	$2.78e^{-02}$	0	0	$7.90e^{-04}$	$8.78e^{-03}$	$2.32e^{-02}$	$4.94e^{-02}$
TT_2-3DVoro	$9.72e^{-02}$	$9.34e^{-02}$	0	0	$3.21e^{-03}$	$5.16e^{-02}$	$3.50e^{-02}$	$2.15e^{-01}$
TT_3-3DVoro	$4.89e^{-02}$	$1.05e^{-01}$	0	0	$2.83e^{-03}$	$5.91e^{-02}$	$3.56e^{-02}$	$2.62e^{-01}$
P-3DVoro	$1.15e^{-01}$	$1.70e^{-01}$	0	$3.33e^{-02}$	$4.79e^{-03}$	$6.28e^{-02}$	$4.65e^{-02}$	$2.61e^{-01}$
M-3DVoro	$1.15e^{-01}$	$1.52e^{-01}$	0	$8.33e^{-02}$	$4.55e^{-03}$	$7.20e^{-02}$	$5.35e^{-02}$	$3.51e^{-01}$

(a) Sampling frequency: 3 s^{-1}

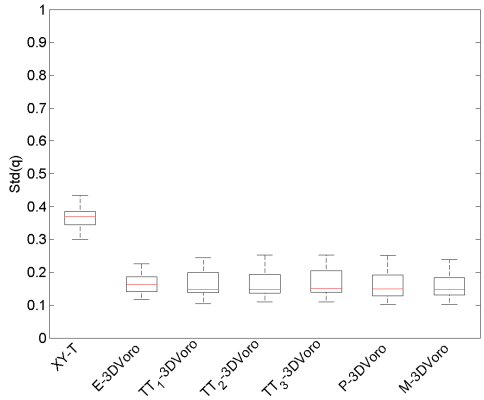
Method	Mean		Mode		Median		90% quantile	
	IT	SoP	IT	SoP	IT	SoP	IT	SoP
XY-T								
E-3DVoro	$2.79e^{-01}$	/	0	/	$1.29e^{-01}$	/	$7.14e^{-01}$	/
TT_1-3DVoro	$4.49e^{-01}$	$2.58e^{-01}$	$5.70e^{-03}$	$1.99e^{-03}$	$1.54e^{-01}$	$1.34e^{-01}$	$8.43e^{-01}$	$6.64e^{-01}$
TT_2-3DVoro	$3.71e^{-01}$	$2.98e^{-01}$	$4.28e^{-02}$	$9.34e^{-02}$	$1.61e^{-01}$	$1.40e^{-01}$	$8.07e^{-01}$	$7.90e^{-01}$
TT_3-3DVoro	$9.82e^{-01}$	$3.56e^{-01}$	$4.34e^{-02}$	$6.70e^{-03}$	$1.64e^{-01}$	$1.38e^{-01}$	$7.76e^{-01}$	$7.74e^{-01}$
P-3DVoro	$3.82e^{-01}$	$3.15e^{-01}$	$2.32e^{-03}$	$6.74e^{-03}$	$1.53e^{-01}$	$1.61e^{-01}$	$9.09e^{-01}$	$7.22e^{-01}$
M-3DVoro	$4.08e^{-01}$	$3.77e^{-01}$	$1.89e^{-02}$	$1.47e^{-02}$	$1.90e^{-01}$	$1.74e^{-01}$	$7.91e^{-01}$	$8.18e^{-01}$

(b) Sampling frequency: 0.5 s^{-1}

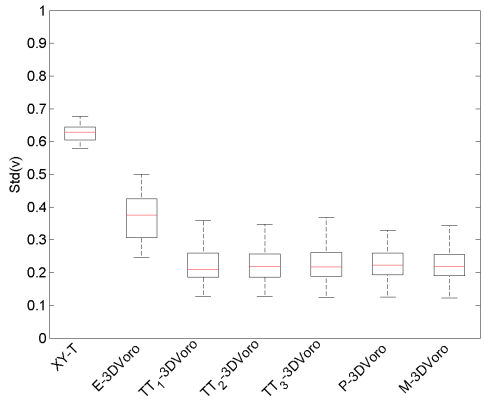
Table 3.4: Robustness to the sampling frequency of density indicator - $Bi_{HD-HeteroPop}$



(a) Standard deviation of density indicator;

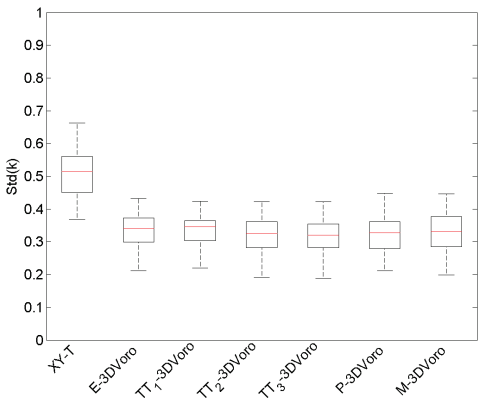


(b) Standard deviation of flow indicator;

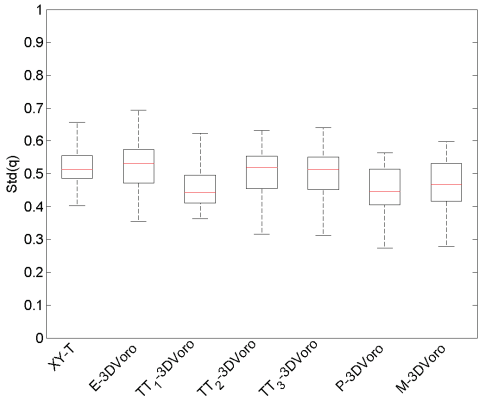


(c) Standard deviation of velocity indicator;

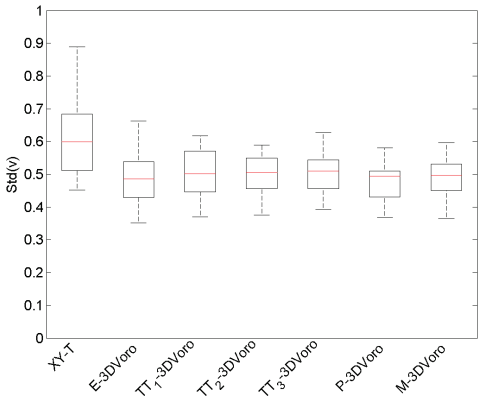
Figure 3.4: Robustness to the simulation noise - $Uni_{LD-HomoPop}$.



(a) Standard deviation of density indicator;



(b) Standard deviation of flow indicator;



(c) Standard deviation of velocity indicator;

Figure 3.5: Robustness to the simulation noise - $Uni_{HD-HeteroPop}$.

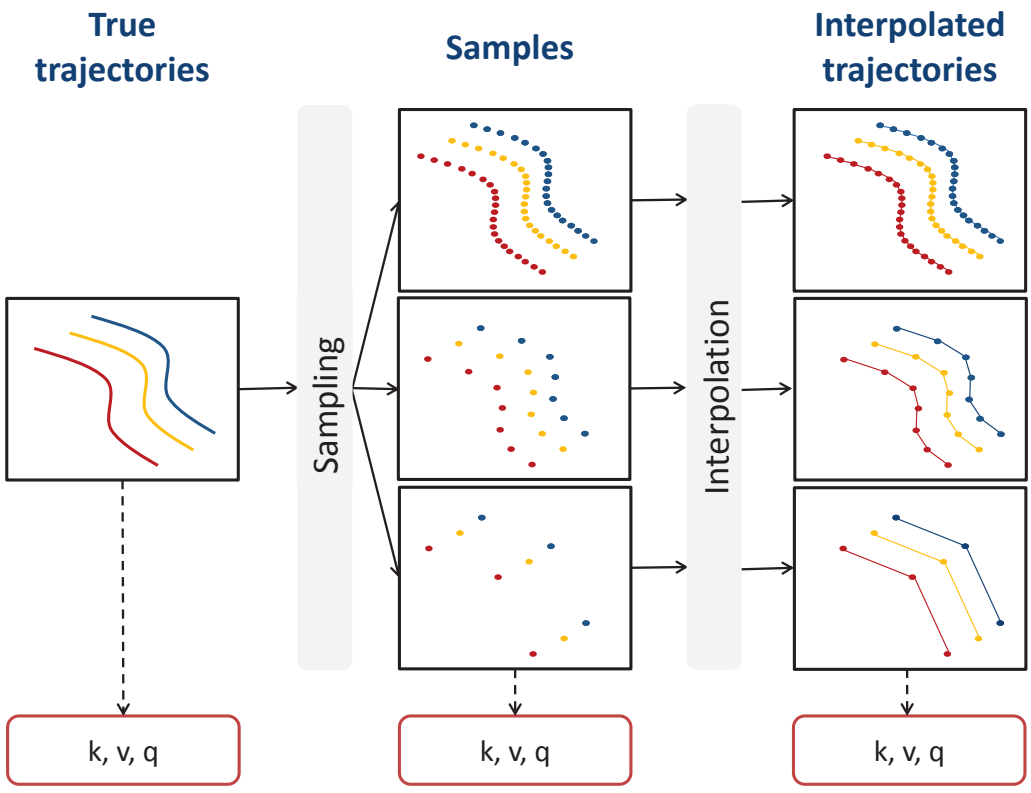


Figure 3.6: Sampling of trajectories and interpolation

4

Probabilistic speed-density relationship for pedestrian traffic

This chapter is based on the article:

Nikolić, M.,Bierlaire, M., Farooq, B. and de Lapparent, M. (2016). Probabilistic speed-density relationship for pedestrian traffic, Transportation Research Part B: Methodological 89: 58 - 81.

The work has been performed by the candidate under the supervision of Prof. Michel Bierlaire, Prof. Bilal Farooq, and Matthieu de Lapparent PhD.

In this chapter, we propose a methodology to represent the speed-density relationship of pedestrian traffic in a probabilistic way. This is motivated by the analysis of data collected from the train station in Lausanne, Switzerland, as well as data collected from a controlled experiment by the Technical University of Delft (Daamen and Hoogendoorn, 2003). The empirical analysis of these pieces of data shows high scatter in the data. The scatter is explicitly represented by relaxing the homogeneity assumption of the equilibrium speed-density relationship. Our approach is not inspired by the physics of the underlying system. Instead, it represents the application of statistical techniques in pedestrian flow modeling.

The structure of the chapter is as follows. In Section 4.1, we first define the variables

involved in the model. We then derive the macroscopic relationships between these indicators, starting from first principles. Based on this theoretical model, we introduce a probabilistic speed-density relationship in Section 4.2. Section 4.3 presents the two case studies from Lausanne and Delft mentioned above. It reports the empirical analysis of the two case studies and emphasizes the limitations of the state of the art approaches on these concrete examples. Section 4.4 and Section 4.5 illustrate the proposed model on the two case studies. Parameter estimation and model validation are discussed in details. Section 4.6 summarizes the outcomes of the proposed methodology.

4.1 Foundation

This section first presents the assumptions related to the quantities involved in our analysis, that is speed and density indicators at the microscopic level. We then present the derivation of the macroscopic relationships between the quantities. It serves as a basis for our modeling framework in Section 4.2.

4.1.1 Density and speed indicators

The *trajectory* of pedestrian i is a curve in space and time, that is

$$p_i(t) = (x_i(t), y_i(t), t), \tag{4.1}$$

where time t spans the horizon of the analysis $[t_0, t_f]$, and $x_i(t)$ and $y_i(t)$ are the coordinates of the position of pedestrian i at time t in a given system of coordinates (typically, we express time in seconds, and use an orthonormal basis for the spatial dimensions).

In practice, the pedestrian trajectory data are collected through an appropriate tracking technology (e.g. Daamen and Hoogendoorn, 2003; Alahi et al., 2011). In this case, the time is discretized and the trajectory is described as a finite collection of triples

$$p_{is} = (x_{is}, y_{is}, t_s), \tag{4.2}$$

where $t_s = (t_0, t_1, \ldots, t_f)$ corresponds to the available sample. We assume that the position of each pedestrian is known at each time t_s of the discretization.

Different measurement methods have been proposed in the literature in order to obtain density and speed indicators from pedestrian trajectories (Chapter 2). Our aim is to be as much independent from the aggregation level as possible and to preserve the heterogeneity of pedestrian population. We rely on a data-driven measurement method inspired by the one proposed by Steffen and Seyfried (2010). This method is based on the spatial discretization that is adjusted to the data through the use of Voronoi

diagrams (Okabe et al., 2000).

The Voronoi space decomposition assigns a personal region to each pedestrian i, in such a way that each point in the personal region is closer to i than to any other pedestrian, with respect of the Euclidean distance. In the presence of sampled data defined by (4.2), for each $s = 0, \ldots, f$ and each pedestrian i, the personal region V_{is} is defined as

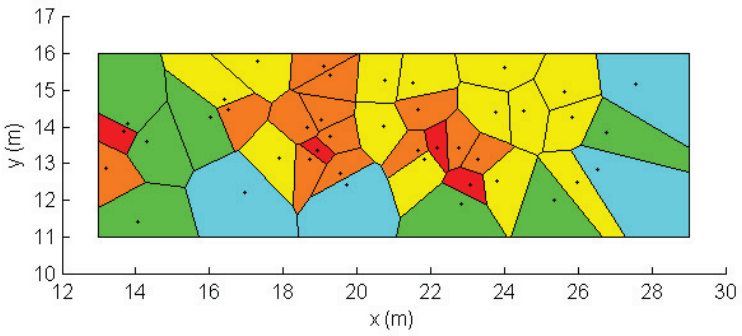
$$V_{is} = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \middle| \left\| \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x_{is} \\ y_{is} \end{pmatrix} \right\|_2 \leq \left\| \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x_{js} \\ y_{js} \end{pmatrix} \right\|_2, \forall j \right\}. \tag{4.3}$$

We assume that each point (x, y) in space is associated with a unique Voronoi region at time t_s, corresponding to the region associated with pedestrian i, that is $V(x, y, t_s) = V_{is}$. Note that if (x, y) is exactly on the border between two or more regions, the unique region associated to it has to be arbitrarily defined.

Given the space discretization specified above, the density of pedestrians at position (x, y) at time t_s is

$$k_{is} = k(x, y, t_s) = \frac{1}{|V_{is}|}, \tag{4.4}$$

where V_{is} is the unique Voronoi region that contains (x, y) at time t_s, and $|V_{is}|$ is the area of V_{is}. The unit is the number of pedestrians per surface unit (typically, square meter). Figure 4.1 illustrates the map of density values obtained through the utilization of (4.4) for a given time instant.



LOS Pedestrian density		
A	< 0.179	[ped/m^2]
B	< 0.270	
C	< 0.455	
D	< 0.714	
E	< 1.333	
F	≥ 1.333	

Table: Pedestrian walkway LOS density treshold values according to NCHRP (in SI units)

Figure 4.1: Voronoi-based density map

The velocity of pedestrian i at time t is given by

$$\vec{v}_i(t) = v_i(t)\vec{d}_i(t), \tag{4.5}$$

where $\vec{d}_i(t)$ is the (normalized) direction of pedestrian i at time t and $v_i(t)$ is the magnitude of the velocity vector, or speed. If the functions $x_i(t)$ and $y_i(t)$ in (4.1) are

differentiable in t, it is defined as

$$v_i(t) = \sqrt{\left(\frac{dx_i(t)}{dt}\right)^2 + \left(\frac{dy_i(t)}{dt}\right)^2}. \tag{4.6}$$

In the presence of discretized data, the speed is approximated using finite differences, for instance

$$v_{is} = \sqrt{\left(\frac{\Delta x_{is}}{\Delta t}\right)^2 + \left(\frac{\Delta y_{is}}{\Delta t}\right)^2}, \tag{4.7}$$

where $\Delta x_{is} = x_{i,s+1} - x_{i,s-1}$, $\Delta y_{is} = y_{i,s+1} - y_{i,s-1}$, and $\Delta t = t_{s+1} - t_{s-1}$.

This definition assumes that the direction of the flow is unique at each point in time and space. It is therefore appropriate for the analysis of speed at an individual level. The definition may not result in the desired outcome if more aggregate characteristics are of interest, in particular if pedestrians in the same area do not walk in the same direction. For instance, if half of pedestrians walk in one direction and the rest in the opposite direction, but both with the same speed, their velocity vectors would cancel out at the aggregate level.

4.1.2 Equilibrium speed-density relationships

In order to derive the relationship between the indicators defined in Section 4.1.1 at the macroscopic level, we start from first principles. Similar to the approach presented in Hoogendoorn et al. (2014), the microscopic Social Force Model (SFM) proposed by Helbing and Molnar (1995) represents the basis of the derivation. The model explains the acceleration of pedestrian i through the influence of neighboring pedestrians j

$$\vec{a}_i = \frac{\vec{v}_i^f - \vec{v}_i}{\tau_i} - C_i \sum_j \exp(-\frac{R_{ij}}{B_i})\vec{n}_{ij}(\lambda_i + (1-\lambda_i)\frac{1+\cos(\phi_{ij})}{2}), \tag{4.8}$$

where $\vec{v}_i^f$ is the desired velocity, $\vec{v}_i$ is the current velocity, τ_i is the relaxation time (the time needed to accelerate/decelerate to the desired velocity), C_i is the interaction strength, R_{ij} is the distance between pedestrians i and j, B_i is a scaling parameter, $\vec{n}_{ij}$ is the unit vector pointing from pedestrian i to j, λ_i denotes the anisotropy parameter and it ranges from 0 to 1 ($\lambda_i = 1$ implies isotropy), and ϕ_{ij} denotes the angle between the direction of i and the position of j.

Consistently with many studies reported in the literature that are focused on the fundamental diagram (Chapter 2), we assume isotropic traffic conditions (identical properties

in all directions). This leads to the model given as

$$a_i = \frac{v_i^f - v_i}{\tau_i} - F_i^r, \tag{4.9}$$

where F_i^r indicates the influence of the interaction between pedestrian i with other pedestrians j for isotropic conditions (the isotropic repulsive force). It is given as

$$F_i^r = C_i \sum_j \exp(-\frac{R_{ij}}{B_i}). \tag{4.10}$$

A pedestrian i experiences a greater repulsive force F_i^r as other pedestrians are closer to i. This can be also reflected through a concept of a personal space associated with a pedestrian. Voronoi diagrams provide a way to derive a personal space "belonging" to a pedestrian, based on the position of i and the positions of neighbors j (see (4.3)). Given that the inverse of a personal space corresponds to density (as defined in (4.4)), the equation (4.9) can be rewritten as

$$a_i = \frac{v_i^f - v_i}{\tau_i} - C_i k_i. \tag{4.11}$$

Under the assumption that traffic conditions are stationary ($a_i = 0$), equation (4.11) results in the relationship between microscopic speed and density indicators

$$v_i = v_i^f - \gamma_i k_i, \tag{4.12}$$

where $\gamma_i = \tau_i C_i$. To establish one-to-one speed-density relationship we further assume that pedestrian population is homogenous. This means that all the pedestrians have the same movement parameters (free-flow speed v_f and sensitivity to congestion γ). In homogenous traffic under stationarity the distances between the pedestrians are also the same, so the densities are. Under these equilibrium conditions the fundamental speed-density relationship is

$$v_e = v_e(k) = v_f - \gamma k, \tag{4.13}$$

where subscript 'e' denotes equilibrium. The resulting relationship correspond to the ones proposed by Older (1968), Navin and Wheeler (1969), Fruin (1971), Tanaboriboon et al. (1986) and Lam et al. (1995) (Table 2.3). It is also in agreement with the model of Hoogendoorn et al. (2014) for isotropic, homogenous and stationary traffic conditions.

Different SFM assumptions can lead to different fundamental relations under the equilibrium assumptions. If the closeness between pedestrians is assumed to affect the free-flow speed (e.g. the free-flow speed is an exponentially decreasing function of density), the

resulting isotropic SFM can be specified as

$$a_i = \frac{v_i^f \exp(-(\frac{k_i}{\theta_i})^{\gamma_i}) - v_i}{\tau_i}, \tag{4.14}$$

where θ_i and γ_i are the pedestrian-specific parameters. For pedestrian traffic at equilibrium (acceleration is zero and all pedestrians possess the same movement parameters) the relationship between speed and density resulting from (4.14) is given as

$$v_e = v_e(k) = v_f \exp(-(\frac{k}{\theta})^{\gamma}). \tag{4.15}$$

This specification coincides with the speed-density relationship proposed by Tregenza (1976) (Table 2.3).

4.2 Methodology

We propose a probabilistic model to characterize speed-density relations. The model is derived by preserving the stationarity assumption and relaxing the homogeneity assumption of equilibrium relations derived in Section 4.1.2. In addition to the general modeling framework, we also present concrete suggestions for the operationalization of the derived model.

4.2.1 Model derivation

We assume that the speed of pedestrians is a random variable (V), such that for each density level there is a distribution of speed values rather than one deterministic value (Figure 4.2). We assume that the following properties characterize this distribution:

P_1: The distribution of speed is continuous with positive support;

P_2: The distribution of speed is unimodal.

The first property is in accordance with the physical characteristic of the speed, being that the speed is a continuous variable whose values cannot be negative. The second property is introduced with a purpose of maintaining the model parsimonious, and it is also motivated by the empirical analysis presented in Section 4.3. Furthermore, we denote by $f_{\text{slow}}(\xi|k)$ the probability density function of the speed values lower than the equilibrium speed $v_e(k)$, and by $f_{\text{fast}}(\xi|k)$ the probability density function of the speed values greater than the equilibrium speed $v_e(k)$, both of which are conditional on density.

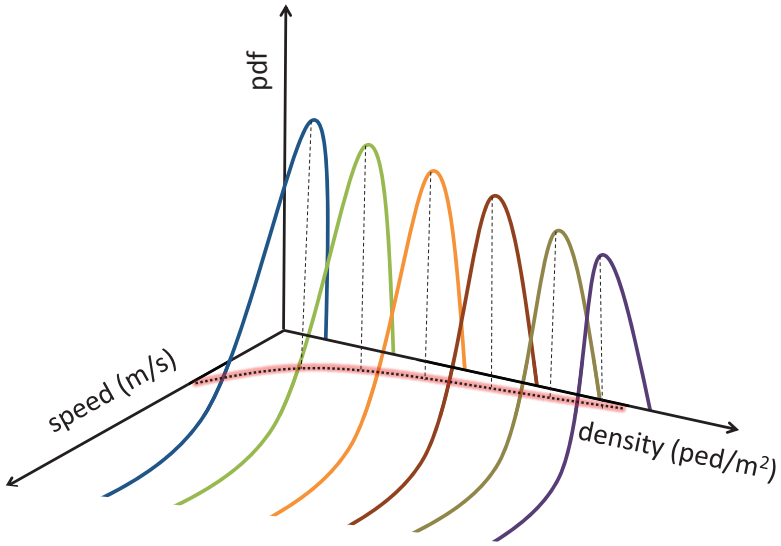


Figure 4.2: Probabilistic speed-density relationship

We define the probability density function of the speed as

$$f_V(\xi|v_e(k),\theta_{slow}(k),\theta_{fast}(k)) = \begin{cases} f_{slow}(\xi|v_e(k),\theta_{slow}(k)), & \xi \leq v_e(k) \\ f_{fast}(\xi|v_e(k),\theta_{fast}(k)), & \xi \geq v_e(k), \end{cases} \tag{4.16}$$

where $\theta_{slow}(k)$ and $\theta_{fast}(k)$ are the density-dependent parameters that characterize slow, respectively fast, regime. The presented modeling assumptions are further supported by the data, as discussed in Section 4.3.

We do not give a specific formulation to individual random effects that may exist. Instead, we assume that individual random effects (i) are independent from other sources of randomness and do not correlate with any of the observed explanatory variable (e.g. density); (ii) do not affect the parameters' estimates and (iii) enter in an additive way the location parameters of the speed distribution.

For the illustration of the proposed methodology we need to specify the exact form of $v_e(k)$, $f_{slow}(\xi|v_e(k),\theta_{slow}(k))$ and $f_{fast}(\xi|v_e(k),\theta_{fast}(k))$. In principle any specification may be chosen, as long as the required properties P_1 and P_2 are satisfied. We give some operational comments and propose concrete examples below.

4.2.2 Operational comments

The proposed specification of the speed-density relationship (4.16) does not satisfy the properties of the C^1 class functions. The maximum likelihood estimation of its parameters may therefore involve complex non-smooth optimization algorithms. Also, the value of $v_e(k)$ is not observed and has to be modeled. To address these issues we adopt a mixing technique, where $v_e(k)$ is assumed to be distributed according to a parametric distribution $f_{v_e(k)}(\zeta;\theta_{v_e}(k))$ with density-dependent parameters $\theta_{v_e}(k)$.

The resulting likelihood function is

$$f_{\text{PedProb-vk}}(\xi|k;\theta_{slow}(k),\theta_{fast}(k),\theta_{v_e}(k)) = \int_{\zeta=0}^{\infty} f_V(\xi|\zeta;\theta_{slow}(k),\theta_{fast}(k)) f_{v_e(k)}(\zeta;\theta_{v_e}(k))d\zeta, \tag{4.17}$$

where $f_V(\xi|\zeta;\theta_{slow}(k),\theta_{fast}(k))$ is defined by (4.16). The model is called *PedProb-vk*, which stands for **Ped**estrian **Prob**abilistic speed **(v)** - density **(k)** relationship.

4.2.3 Exemplary specification

We suggest a linear model for the distribution of the speed of the slow component, since it is well adjusted to impose the lower bound of the distribution at zero

$$f_{\text{slow}}(\xi|v_e(k),\alpha_k,\beta_k) = \frac{\beta_k - \alpha_k}{v_e(k)}\xi + \alpha_k, \tag{4.18}$$

where $0 \leq \xi \leq v_e(k)$, and $\alpha_k \geq 0$ and $\beta_k \geq 0$ are parameters dependent on k. They are such that $f_{\text{slow}}(0|k) = \alpha_k$ and $f_{\text{slow}}(v_e(k)|k) = \beta_k$. We adopt a simple specification for the parameters α_k, β_k. We assume that they depend on the density k in the following way

$$\alpha_k(a_\alpha, b_\alpha) = a_\alpha k + b_\alpha, \tag{4.19}$$

and

$$\beta_k(a_\beta, b_\beta) = a_\beta k + b_\beta \tag{4.20}$$

where $a_\alpha \geq 0$, $b_\alpha \geq 0$, $a_\beta \geq 0$ and $b_\beta \geq 0$. They are such that $\alpha_0 = b_\alpha$ and $\beta_0 = b_\beta$. The specifications (4.19) and (4.20) capture the relationship between the density level and the shape of speed distribution as suggested by the data (Section 4.3). When the density is lower, low speed values are less frequent and speed distribution is more spread. When the density is higher, low speed values are more frequent, speed distribution is less spread and the speed values close to the mode are more frequent.

We propose an exponential model for the distribution of the speed of the fast component

$$f_{\text{fast}}(\xi|v_e(k),\beta_k,\lambda) = \exp(-\lambda\xi + \log(\beta_k) + \lambda v_e(k)), \tag{4.21}$$

where $\xi \geq v_e(k)$, $v_e(k)$ and β_k are defined as above, and $\lambda \geq 0$ is an additional parameter, defining the rate of the exponential distribution. The choice of the exponential specification is well suited to avoid any arbitrariness in imposing the upper bound, while preventing arbitrary high values at the same time. The normalizing constant is in this example equal to $\frac{\alpha_k+\beta_k}{2}v_e(k) + \frac{\beta_k}{\lambda}$.

Note that, the value determined by $v_e(k)$ represents the mode of the distribution ($\frac{df_V(\xi|v_e(k),\alpha_k,\beta_k,\lambda)}{d\xi} = 0$ when $\xi(k) = v_e(k)$). If $\xi = v_e(k)$, the values of $f_{\text{slow}}(\xi|k)$ and $f_{\text{fast}}(\xi|k)$ coincide, and are both equal to β_k. When $k = 0$, then $f_{\text{slow}}(0|k) = b_\alpha$, $f_{\text{slow}}(v_e(k)|k) = f_{\text{fast}}(v_e(k)|k) = b_\beta$ and $f_{\text{fast}}(\xi|k) < b_\beta$ if $\xi > v_e(k)$. When $k \to \infty$, then $f_{\text{slow}}(0|k) = f_{\text{fast}}(0|k) \to \infty$ and $f_{\text{fast}}(\xi|k) \to 0$. The parameters characterizing the distribution of each component are illustrated in Figure 4.3.

We use a symmetric triangular distribution for mixing defined on the interval $[\bar{v}_e(k) - \sigma(k), \bar{v}_e(k) + \sigma(k)]$, where $\bar{v}_e(k)$ is the mean of the distribution and $\sigma(k) = \frac{1}{1+\exp(\eta)}\bar{v}_e(k)$

$$f_{v_e(k)}(\zeta;\bar{v}_e(k),\sigma(k)) = \begin{cases} \frac{\zeta-\bar{v}_e(k)+\sigma(k)}{\sigma(k)^2}, & \bar{v}_e(k) - \sigma(k) \leq \zeta \leq \bar{v}_e(k) \\ \frac{\bar{v}_e(k)+\sigma(k)-\zeta}{\sigma(k)^2}, & \bar{v}_e(k) < \zeta \leq \bar{v}_e(k) + \sigma(k) \\ 0, & \zeta < \bar{v}_e(k) - \sigma(k) \text{ or } \zeta > \bar{v}_e(k) + \sigma(k). \end{cases} \tag{4.22}$$

The choice of this specific distribution is motivated by the simple closed form of its probability density function and cumulative density function. The proposition for the specification of $\sigma(k)$ insures that the property P_1 is satisfied. The specification of $\bar{v}_e(k)$ is typically a deterministic speed-density relationship such as those presented in Section 4.1.2.

Putting everything together, the distribution of the speed in our example is given by

$$f_{\text{PedProb-vk}}(\xi|k;\alpha_k,\beta_k,\lambda,\bar{v}_e(k),\sigma(k)) = \int_{\zeta=\bar{v}_e(k)-\sigma(k)}^{\bar{v}_e(k)+\sigma(k)} f_V(\xi|\zeta;\alpha_k,\beta_k,\lambda) f_{v_e(k)}(\zeta;\bar{v}_e(k),\sigma(k))d\zeta, \tag{4.23}$$

where the distributions of the slow and the fast component are defined by (4.18), respectively (4.21), and $f_{v_e(k)}(\zeta;\bar{v}_e(k),\sigma(k))$ is defined by (4.22). For this specification, the integral (4.23) does not have a closed-form, and numerical integration or Monte-Carlo simulation is required. However, other specifications of $f_{v_e(k)}(\zeta;\theta_{v_e}(k))$,

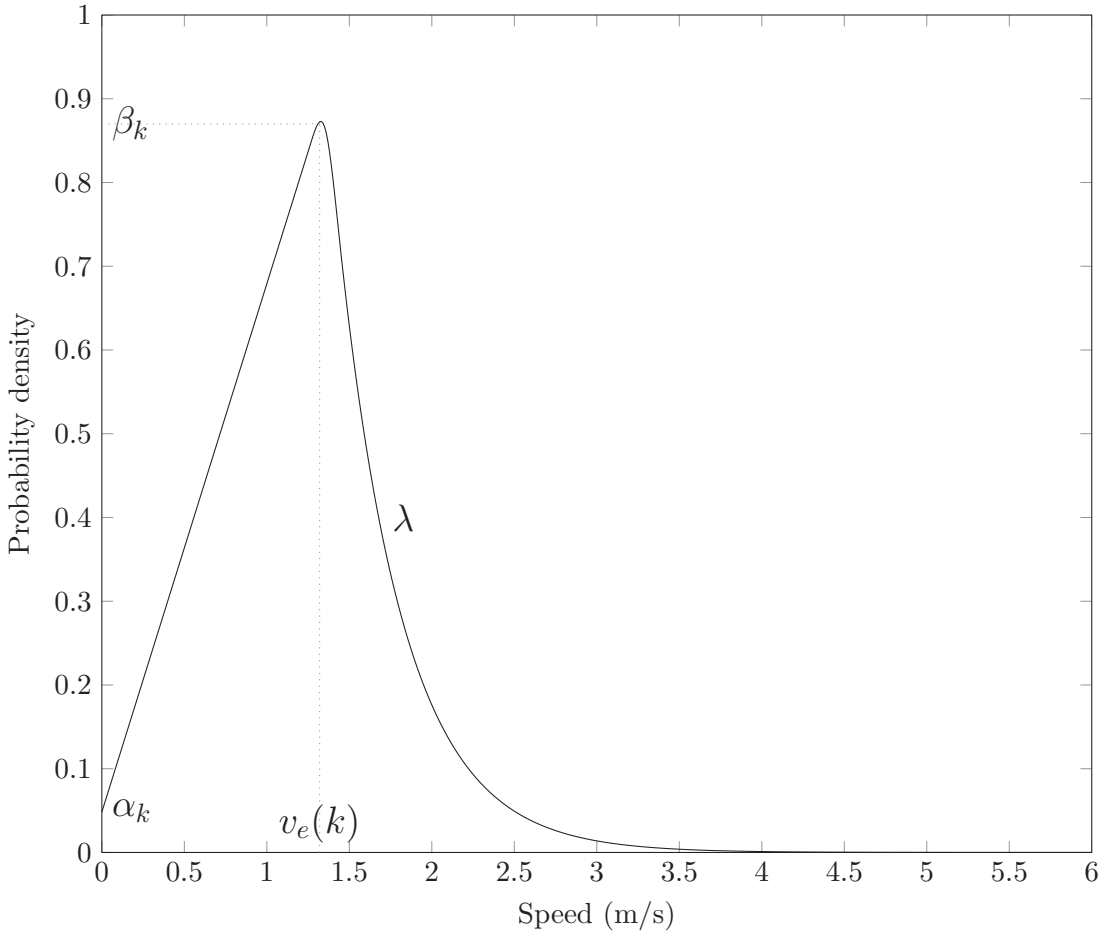


Figure 4.3: Illustration of the model - one density level

$f_{slow}(\xi|v_e(k), \theta_{slow}(k))$ and $f_{fast}(\xi|v_e(k), \theta_{fast}(k))$ may lead to a closed-form formulation, as illustrated in Appendix B.

We emphasize that the proposed specification is merely one possibility, and the framework is general and other specifications are possible (e.g. different functional forms of $v_e(k)$, $f_{v_e(k)}(\zeta; \theta_{v_e}(k))$, $f_{slow}(\xi|v_e(k), \theta_{slow}(k))$ and $f_{fast}(\xi|v_e(k), \theta_{fast}(k))$, different smoothing strategies, etc). To illustrate and validate our approach, we perform next the analysis on two case studies (Section 4.4, Section 4.5).

4.2.4 Estimation procedure

For the estimation of the model parameters, we use quasi-maximum likelihood estimation.[1] As each pedestrian generates several piece of data, there is serial correlation

[1] Quasi-maximum likelihood estimation differs from full information maximum likelihood estimation in the sense that it is necessary to explicitly specify the covariance structure for the latter.

among these observation. If ignored, the unobserved/ unmodeled covariance among the observations of a same individual in the estimation procedure leads to consistent but inefficient estimates (Gourieroux et al., 1984; Trognon, 1987; Wooldridge, 2010). This holds as long as the first order moment of the distribution of the endogenous variable is not affected by its higher order moments (Gourieroux et al., 1984), i.e. independence between mean and higher moments. In this case, the Rao-Cramer bound is not reached, and a block bootstrap method must be used for the calculation of the standard errors of the estimates (Hall et al., 1995; Davison and Hinkley, 1997; Davidson and MacKinnon, 2004). It is also convenient to assume that there is no correlation across pedestrians, and that the observations of one and the same pedestrian are serially correlated due to unobserved time invariant specific effects. Bootstrap replications are thus based on the sampling over blocks, where each block contains the series of the observations of one individual. We adopt these assumptions in the case studies presented below.

4.3 Case studies and empirical investigation

The motivation of this research comes from the analysis of two real datasets, that we use below to illustrate and validate our approach.

4.3.1 Lausanne train station

The first dataset is collected in a pedestrian underpass of the train station of Lausanne, Switzerland. Figure 4.4 shows the layout of the studied area. It covers approximately 685 m^2. The underpass is frequently used especially during the morning and afternoon peak hours since it connects the exterior of the train station to the main platforms. It also acts as a connection between mostly residential south and the center of the city in the north.

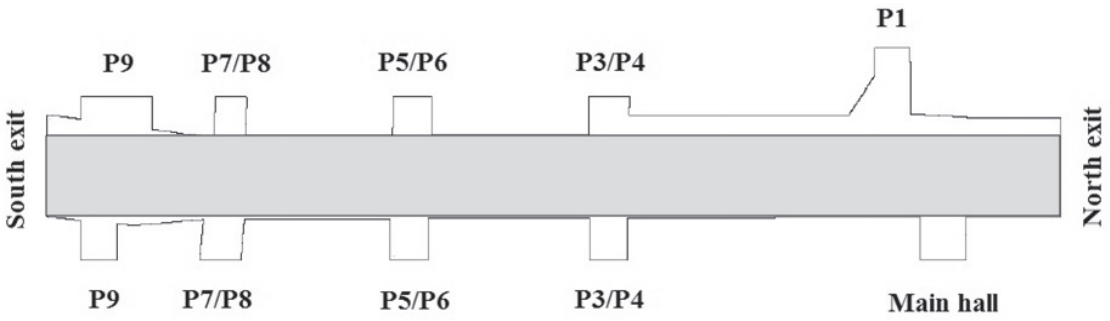


Figure 4.4: Lausanne train station - pedestrian underpass West

To collect the raw data, a large-scale network of smart sensors has been deployed in the station. The underlying technology is based on infrared and depth sensors that detect silhouettes and track each pedestrian in the scene covered by the network. The tracking engine uses a sparsity driven framework (Alahi et al., 2011, 2014) to link detected pedestrians over the network of sensors.

It results in a dataset of 25,603 trajectories, collected during a time period between 07:00 and 08:00 on February 12, 13, 14, 15 and 18, 2013. The temporal resolution of every trajectory ranges from 10 to 25 points per second and it has been processed to obtain the position of every pedestrian in the scene at every second. The average length of the trajectories is 78 meters and the duration of a pedestrians' stay in the underpass ranges from 15 seconds to 2.2 minutes.

Note that we have selected only trajectories collected in the shaded area shown in Figure 4.4, referring to a corridor. The trajectories from the ramps and stairs (denoted as P1-P9) are not considered in this study. Indeed, as explained by Daamen (2004) and Weidmann (1993), the walking behavior and, therefore, the speed-density relationship, varies with the type of infrastructure.

In the rest of the thesis, we refer to this case study as the Lausanne case study.

4.3.2 Controlled experiment

The second set of data has been collected during a controlled experiment at the Technical University of Delft in the Netherlands (Daamen and Hoogendoorn, 2003). The individuals participating in the experiment were instructed to walk along a corridor that is 10 meters long and 4 meters wide, at a normal speed, and to pass through a bottleneck of 1 meter in width (see Figure 4.5, where individuals walk from right to left).

The scene was filmed from the top by digital cameras. The individual trajectories were extracted from the digital video sequences.

The experiment lasted about 15 minutes. A total of 1,123 trajectories were collected, where the position of each individual is available every 0.1 second. The average length of the trajectories is similar inside and upstream of the bottleneck and it is approximately 5 meters. The average travel time of the trajectories upstream of the bottleneck is 10 seconds, whereas inside the bottleneck it is lower (approximately 5 seconds).

Note that we have selected trajectories collected in the rectangular area (5 meters long and 4 meters wide) upstream of the bottleneck (Figure 4.5). As explained by Duives et al. (2014b, Figure 4), this is where the variability is observed.

In the rest of the thesis, we refer to this case study as the Delft case study.

Figure 4.5: Narrow bottleneck experiment (Daamen and Hoogendoorn, 2003)

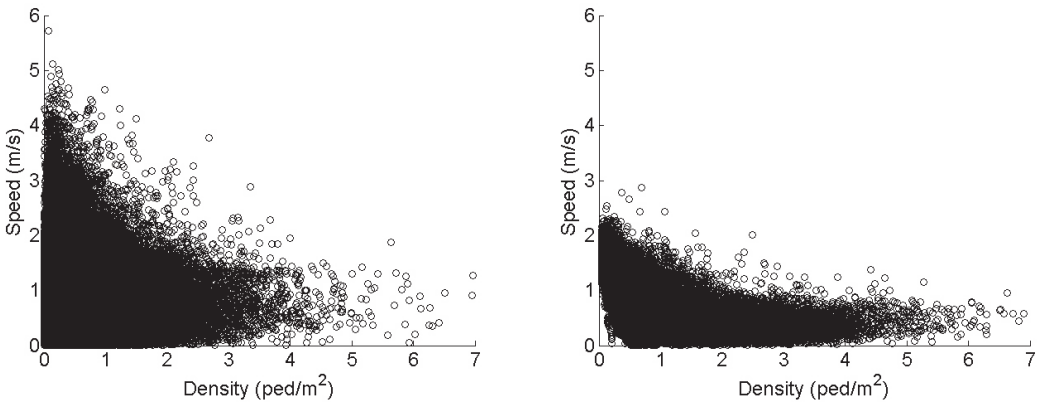
4.3.3 Empirical analysis

The speed-density profiles corresponding to the Lausanne and the Delft case studies are obtained from the measurement method presented in Section 4.1.1. In Figure 4.6, each circle corresponds to one observation, that is, one pedestrian at one specific time in the horizon. The x coordinate of the circle corresponds to the density, calculated from (4.4), and its y coordinate corresponds to the speed calculated from (4.7).

Figure 4.6a plots 270,291 observations corresponding to the peak hour of February 12, 2013 for the Lausanne case study. The same pattern was observed on any weekday. Figure 4.6b plots 119,156 observations for the Delft case study.

A high scattering is observed in both cases (Figure 4.6). The density ranges from 0 to approximately 7 pedestrians per square meter. In the Lausanne case, the speed ranges from 0 to 5.72 meters per second (that is about 21 km/h), and 99% of the observations are between 0 and 2.42 meters per second (that is about 9 km/h). In the Delft case, the speed ranges from 0 to 2.87 meters per second (that is about 10 km/h). The difference in the speed distribution is attributed to the controlled nature of the experiment in Delft, where individuals where instructed to walk at normal speed, resulting to a lower variance compared to Lausanne, where no instruction was given. For the same reason, low speeds where not observed at low density in Delft, contrarily to Lausanne.

To investigate this data in more details, the speed distributions at various density levels are presented in Figures 4.7 and 4.8 for the Lausanne and Delft case study, respectively. In both cases, a higher level of variability is noticeable at lower densities, compared to



(a) Lausanne case study (b) Delft case study

Figure 4.6: Speed-density profiles

higher densities where the distribution of speed is less spread and shifted towards lower values.

The deterministic models for the speed-density relationship proposed in the literature (Chapter 2) appear to be inadequate for representing the observed patterns. Clearly, density is not the only factor influencing pedestrians' speed. We emphasize that there are different sources of uncertainty that can contribute to the observed pattern: (i) the variability due to the nature of the system (e.g. heterogeneity of behavior, interaction of behaviors such as people walking in groups, the existence of multi-directional flows that are more chaotic) and (ii) measurement errors (e.g. discretization and measurement noise, technological issues). The Voronoi based approach is completely data-driven and designed to minimize the spatial discretization errors. Given the type of the available data, these sources are not separable from each other. Instead, the proposed probabilistic model for the speed-density relationship accounts for the whole uncertainty at once.

4.4 Case study: Lausanne

We illustrate and validate now the model on the Lausanne dataset, introduced in Section 4.3.1. The process involves analyzing the goodness of fit (Section 4.4.1), qualitative and quantitative comparison of the distributions predicted by the model and empirical ones (Section 4.4.2). We also analyze the predictive performance of the model when applied to data that were not used in model estimation (Section 4.4.3).

4.4.1 Model Estimation

The dataset used for the estimation consists of 1,269,393 pairwise speed-density observations corresponding to the peak hour of February 12, 13, 14, 15 and 18, 2013. The descriptive statistics of the estimation dataset are presented in Table 4.1. The dataset is categorized according to six levels of service (LoS) proposed by Fruin (1971) for pedestrian facilities, labeled from A to F. Table 4.1 shows that the largest part of the observations falls below the LoS F. Actually, 99% of the observations are below 2.06 ped/m^2.

LoS	Number of observations
A (k $\leq$ 0.31 ped/m^2)	644546
B (k $\in$ (0.31 - 0.43 ped/m^2])	174116
C (k $\in$ (0.43 - 0.71 ped/m^2])	229808
D (k $\in$ (0.71 - 1.11 ped/m^2])	133812
E (k $\in$ (1.11 - 2.17 ped/m^2])	76725
F (k $>$ 2.17 ped/m^2)	10386

Table 4.1: Estimation data classified according to LoS (Fruin, 1971) - Lausanne case study

The estimation results for the model (4.23) presented in Section 4.2.3 are shown in Table 4.2. The number of iterations (200) for the block bootstrap has been determined by computing bootstrap approximation of the mean square error of the estimates (Ross, 2013b, Section 8.3). The parameter $\bar{v}_e(k)$ is specified in the model (4.13).

All estimates have the expected sign and value. The results also show the low standard errors of all parameters and their statistical significance (t-test) at a usual significance level (0.05).[2]

The positive sign of the parameter a_α shows that α_k, that is the likelihood of low speeds, increases with density. Similarly, the positive sign of the parameter a_β shows that β_k, that is the likelihood of the mode of the speed distribution, also increases with density.

The signs and the estimated values of the parameters v_f and γ are consistent with the ones reported in the literature and with the trend observed in the data. The low variance of the parameter η is as expected, since the mixing distribution (4.22) is introduced for the smoothing purposes.

The estimated equilibrium speed-density relationship is shown in Figure 4.9. Note that it suggests the absence of the decreasing part in the corresponding flow-density relationship. This is due to the fact the density in the considered system is below its critical

[2]The high significance of the parameters can also be attributed to the large number of observations used for the model estimation.

Parameter	Value	Std err (block bootstrap)	t-test
a_α	0.0393	$4.78e^{-03}$	8.52
b_α	0.00708	$8.85e^{-04}$	7.87
a_β	0.00487	$2.41e^{-03}$	2.10
b_β	0.142	$3.71e^{-03}$	38.4
λ	3.53	$1.70e^{-03}$	$2.07e^{03}$
v_f	1.29	$2.85e^{-03}$	$4.57e^{02}$
γ	0.0512	$3.85e^{-03}$	13.4
η	3.48	$1.25e^{-03}$	$2.77e^{03}$
$\log \mathcal{L}$	-783942.897		
Number of parameters	8		
Number of observations	1269393		

Table 4.2: Estimation results - Lausanne case study

value.

Various alternative specifications of the model have been investigated in order to represent the speed-density relationship. The first alternative specification (denoted as the specification A), assumes the Kumaraswamy distribution of speed values (e.g. Nikolić et al. (2014)). In the second specification (denoted as the specification B), the speed values are assumed to be generated from the Rayleigh distribution. The third specification (denoted as the specification C), assumes the Weibull distribution of speed values. In all cases the parameters of the distributions are allowed to vary with density. Also, the alternative specification of *PedProb-vk* is considered (denoted as the specification D), where we tried for $\bar{v}_e(k)$ the specification proposed by Weidmann (1993) (Table 2.3). All alternative specifications resulted in a poorer fit with respect to *PedProb-vk*, when compared by the means of the Bayesian information criterion - BIC (Wasserman, 2000):

$$BIC = -2 \cdot \log \mathcal{L} + r \cdot \log(n) \tag{4.24}$$

where n is the number of the observations, r is the number of parameters to be estimated and $\mathcal{L}$ is the value of the likelihood function of the model. The model with the lowest BIC is preferred. Indeed, the BIC values shown in Table 4.3 suggest that the *PedProb-vk* represents the best compromise between the model accuracy and simplicity among all evaluated specifications.

4.4.2 Kolmogorov-Smirnov validation

The validation is performed by comparing the distribution functions of the estimated model, and the empirical distributions from the dataset. The analysis is carried out

at different density levels. Figure 4.10 shows the probability density functions of the model (model pdf) and the data (empirical histogram) at the same density level. The corresponding cumulative density functions (model cdf and empirical cdf) are plotted in Figure 4.11. Qualitatively, the match between the two is pretty satisfactory.

For the quantitative analysis, we use the Kolmogorov-Smirnov probability distance metric (Massey, 1951)

$$D_k = \max_v \left| F_{\text{model}}(v|k) - F_{\text{data}}(v|k) \right|, \tag{4.25}$$

where $F_{\text{model}}(v|k)$ corresponds to the model cdf and $F_{\text{data}}(v|k)$ to the empirical cdf. This metric represents the maximum value of the absolute vertical difference between the two cumulative distribution functions. It is reported in Figure 4.12a.

We calculate the p-value of the Kolmogorov-Smirnov statistic using simulation (Ross, 2013a, p. 257), with 100 simulation runs. The model (4.23) is simulated using the rejection method (Ross, 2013b, Section 5.2) on draws from a Rayleigh distribution. The results are shown in Figure 4.12b. They suggest that there is no evidence in the data to reject *PedProb-vk* at significance level 0.05 for all levels of density except maybe for the one corresponding to densities close to zero.

4.4.3 Specification test

In order to test the robustness of the proposed specification, we have performed the validation that consists in splitting the dataset into two subsets. The model is re-estimated on one subset and the remaining data, unused for estimation, is used for validation purposes. The procedure consisting of the following steps is repeated 100 times:

1. A sample of 80% of pairwise speed-density observations is selected using simple random selection.
2. The parameters of the model are estimated using the generated sample.
3. The Kolmogorov-Smirnov statistic D_k (4.25) is calculated to compare the estimated model and the data on the remaining 20% of the dataset.

In Figure 4.13a, we compare the value of D_k calculated on the full dataset (in dashed line) with the values calculated with the above mentioned procedure. The 100 values are summarized using a box plot at each level of density. These results are satisfactory. The specification is robust and no over-fitting is detected.

To be more precise, we also calculate the p-value for each value of D_k calculated with the above mentioned procedure. For this purpose we use simulation (Ross, 2013a, p. 252), with 100 simulation runs. The box plot of the estimated p-values are shown in Figure 4.13b. The results do not allow to reject the hypothesis that the data and the model follow the same distribution, at a usual level of significance.

4.5 Case study: Delft

In this section we illustrate and validate the model on the Delft dataset, described in Section 4.3.2. The process involves qualitative and quantitative comparison of the distributions predicted by the model and empirical ones (Section 4.5.2), and the analysis of the predictive performance of the model when applied to unseen data (Section 4.5.3).

4.5.1 Model Estimation

The dataset used for estimation consist of 119,156 pairwise speed-density values observed upstream of the bottleneck. The data has been classified according to the LoS standard of Fruin (1971), showing that now the majority of the observations corresponds to the LoS C, D, E and F (Table 4.4). Consequently 99% of the data is below the density value of 3.9 ped/m^2. This is as expected, given the existence of flow constraint (in the form of a narrow bottleneck) that in this case causes congestion upstream of the bottleneck.

The parameters of the model (4.23) have been estimated, where $\bar{v}_e(k)$ is specified by the model (4.15). The estimation results are shown in Table 4.5. The sign and the magnitude of the parameters are as expected. The standard errors of the parameters were computed via block bootstrap method (see Section 4.2.4). The results also indicate a high significance of the estimated values.

The positive sign of the parameter a_α shows that α_k, that is the likelihood of low speeds, increases with density. Similarly, the positive sign of the parameter a_β shows that β_k, that is the likelihood of the mode of the speed distribution, also increases with density.

The signs and the estimated values of the parameters v_f, θ and γ of the model (4.15) are in accordance with the trend observed in the data.

The value of the parameter λ is higher than the one for the Lausanne case study, which is consistent with the reduced range of speed values in the data. Finally, the value of the parameter η is slightly lower than that of Lausanne case study.

The estimated equilibrium speed-density relationship is shown in Figure 4.14.

4.5.2 Kolmogorov-Smirnov validation

The agreement between the model predictions and the observations from the estimation dataset is illustrated in Figure 4.15 and Figure 4.16. The Kolmogorov-Smirnov distances (4.25) between the model cdf and the empirical cdf are illustrated in the Figure 4.17a.

The agreement between the model predictions with data appears to be less satisfactory for lower density levels, as we have fewer data with low speed at low density levels. This is an artefact of the experimental nature of the data (see discussion in Section 4.3.3). The quality of the fit for higher density levels (which is of greater interest for applications anyway) is satisfactory. Figure 4.17a shows that the smallest Kolmogorov-Smirnov distances correspond to the density levels which are characterized by the largest number of observations (Table 4.4).

The p-values are estimated using the procedure described in Section 4.4.2 and shown in Figure 4.17b. Again, there is no evidence in the data to reject *PedProb-vk* at significance level 0.05 for most of the density levels. The p-values less than 0.05 are observed for lower density levels, up to 0.3 ped/m^2, and for density levels greater than 3.5 ped/m^2. In the former case low p-values are caused by the experimental nature of the data (as discussed above), while in the latter a low number of observations (0.02% of the data) is insufficient to reach any conclusion.

4.5.3 Specification test

We test the robustness of the model specification by performing the validation using 80% of the data for estimation and the remaining 20% for validation (see Section 4.4.3). The Kolmogorov-Smirnov statistics for different density levels from 100 simulation runs and corresponding p-values are shown using box plot representation in Figure 4.18a and Figure 4.18b, respectively. The above results validate the model also for the Delft case study.

4.6 Summary

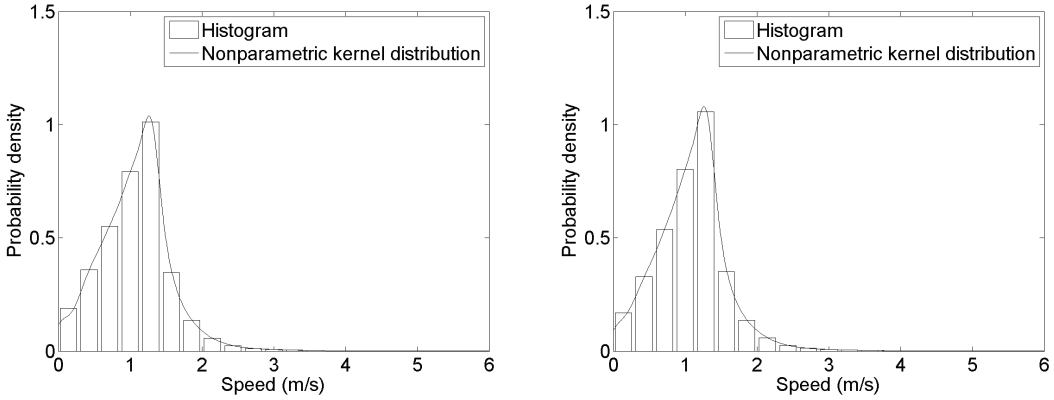
This chapter points out a major shortcoming in the existing literature in terms of explaining the heterogeneity in the high resolution datasets on pedestrian flows. These datasets are available more and more thanks to advancements in computer vision and sensors technology, and their power needs to be utilized.

The chapter contributes by developing a probabilistic model of speed-density relationship for pedestrians to tackle this heterogeneity. The approach is inspired by the analyzed

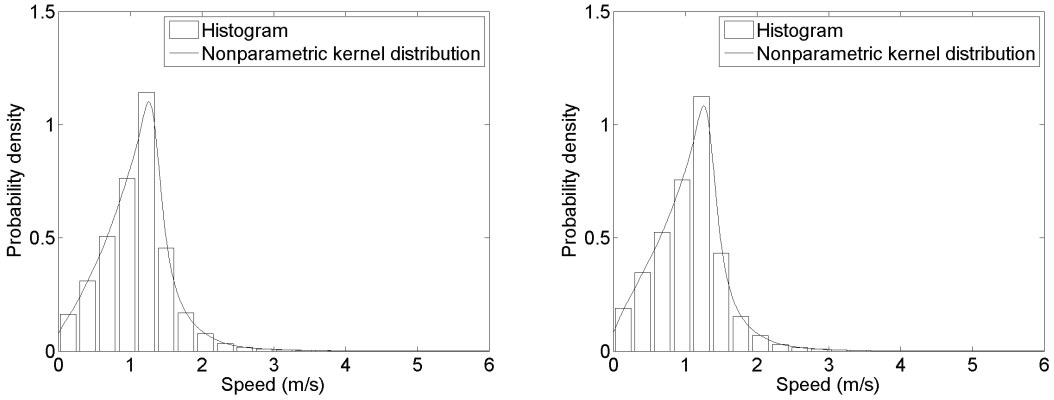
data and it results in the model that is parsimonious and flexible. It relies on relatively few assumptions, and can be efficiently adjusted to accommodate different behavioral situations and different types of infrastructure.

Model estimation and validation performed in this chapter are based on two different and extensive case studies (data from a real scene and from controlled experiments). Various statistical tests empirically validate the proposed model specification and indicate its high performance across case studies. Results also show satisfactory predictive capabilities of the model.

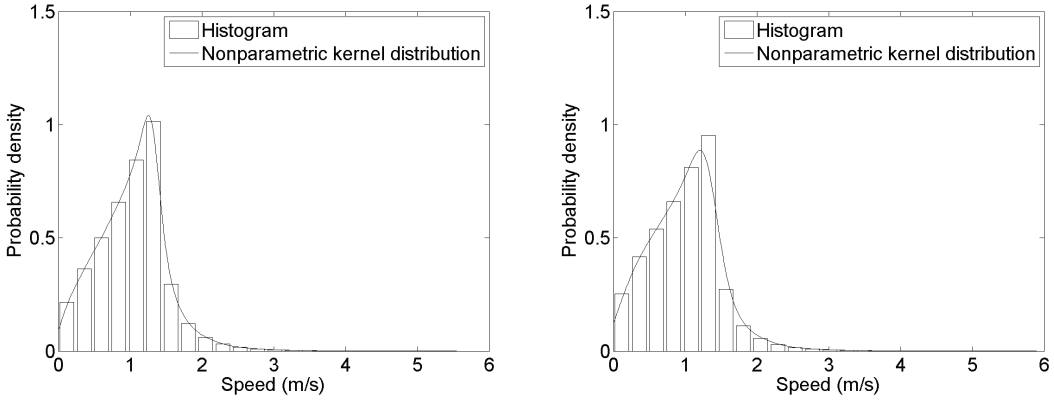
The proposed model implicitly captures the impact of pedestrian heterogeneity on the form of macroscopic relationships of pedestrian flow. Thus, the model does not provide the way to explore what are the microscopic factors influencing pedestrian walking behavior at macroscopic level, and to what extent. We address this matter in the next chapter (Chapter 5).



(a) k < 0.1 ped/m^2 (b) k $= 0.1$ ped/m^2

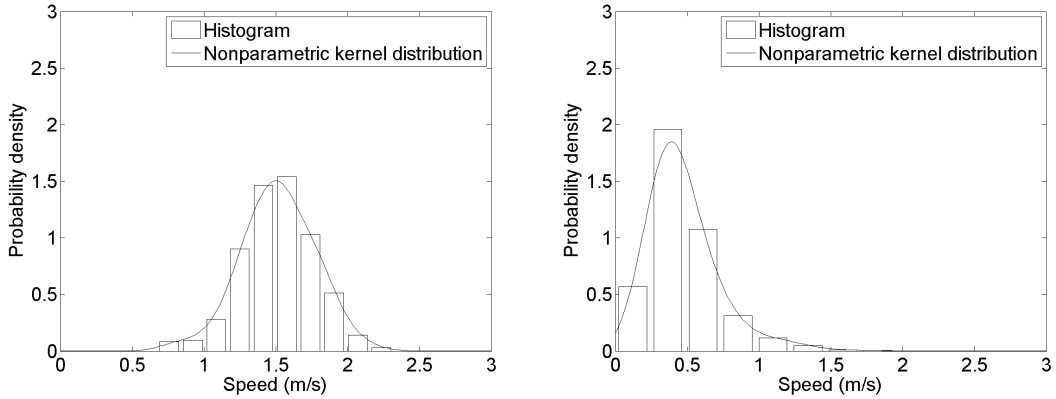


(c) k $= 0.2$ ped/m^2 (d) k $= 0.3$ ped/m^2



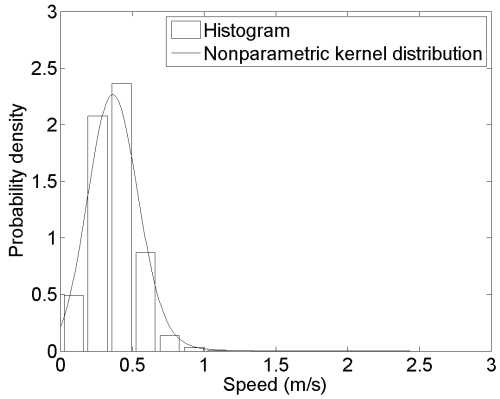
(e) k $=0.4$ ped/m^2 (f) k $= 0.5$ ped/m^2

Figure 4.7: Speed distributions for different density levels - Lausanne case study

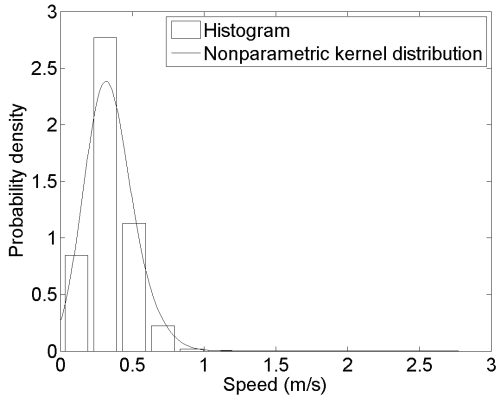


(a) k < 0.1 ped/m^2

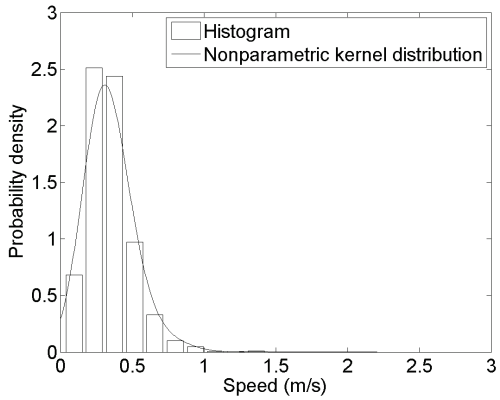
(b) k = 1 ped/m^2



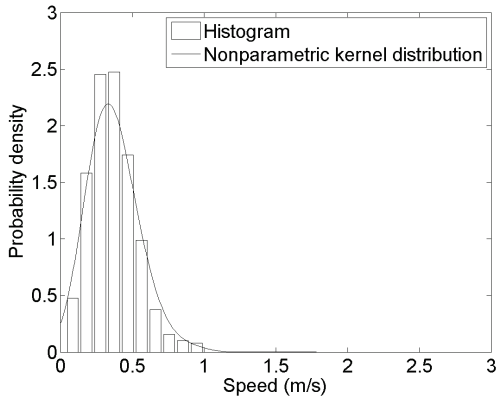
(c) k = 2 ped/m^2



(d) k = 2.5 ped/m^2



(e) k = 3 ped/m^2



(f) k = 3.5 ped/m^2

Figure 4.8: Speed distributions for different density levels - Delft case study

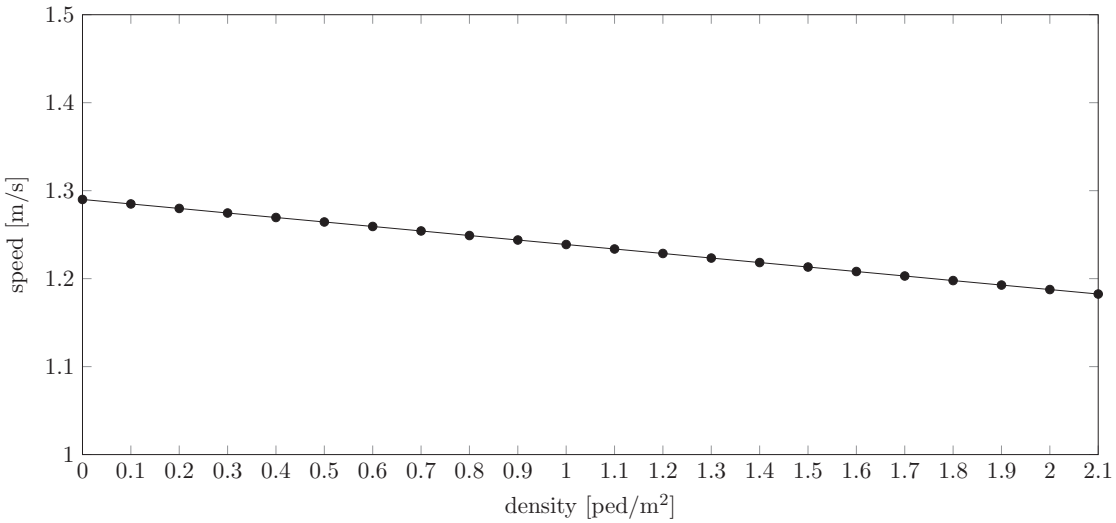


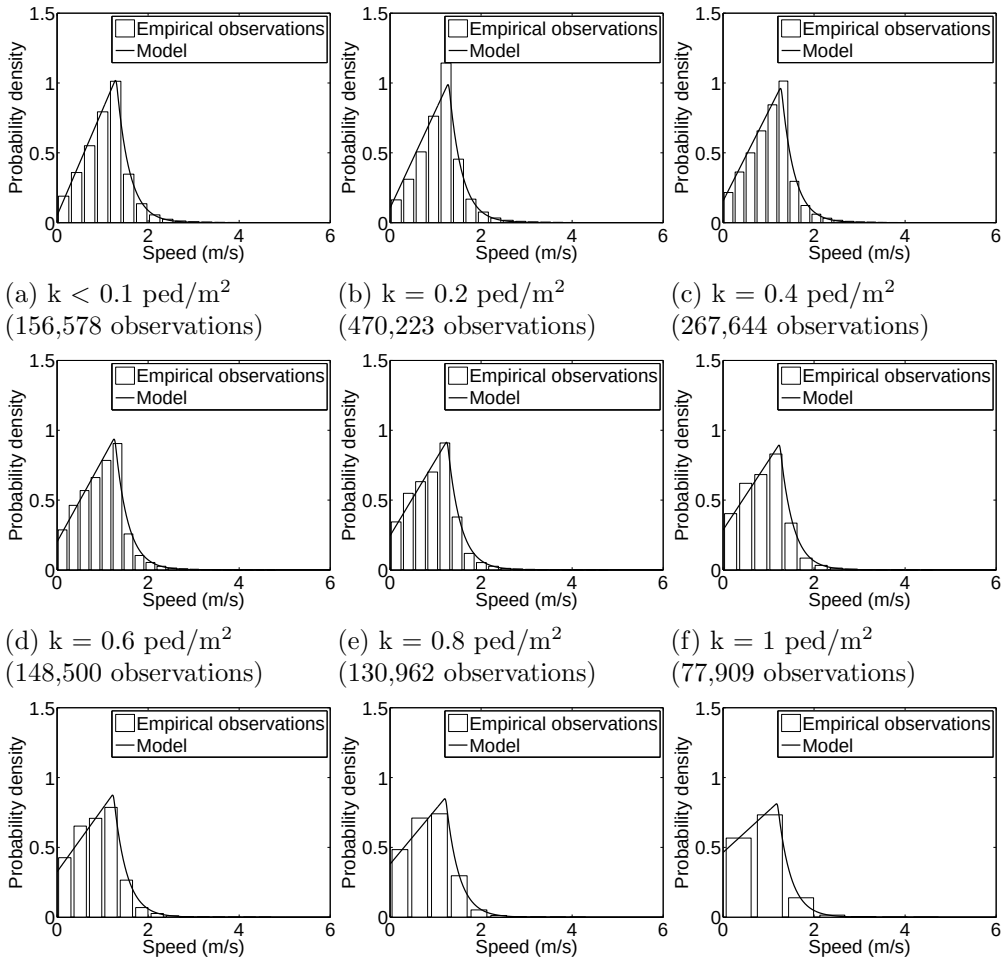
Figure 4.9: Equilibrium relationship - Lausanne case study

Model specification	A	B	C	D	*PedProb-vk*
$\log \mathcal{L}$	-891878.142	-897336.414	-895420.690	-845079.532	-783942.897
# parameters	8	3	4	9	8
# observations	1269393	1269393	1269393	1269393	1269393
BIC	1783868.715	1794678.828	1790897.596	1690285.550	1567998.226

Table 4.3: Goodness of Fit (BIC) - Lausanne case study

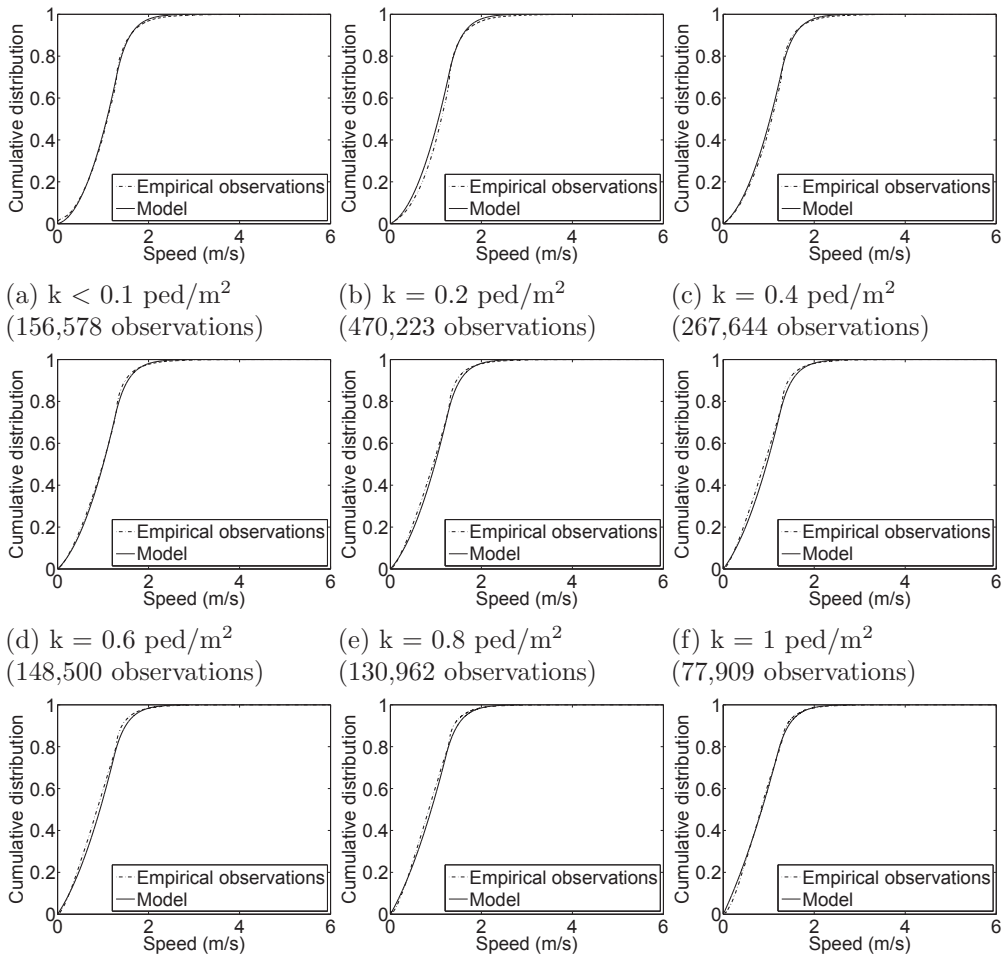
LoS	Number of observations
A ($k \leq 0.31$ ped/m^2)	9288
B ($k \in (0.31$ - 0.43 ped/m$^2]$)	6967
C ($k \in (0.43$ - 0.71 ped/m$^2]$)	20497
D ($k \in (0.71$ - 1.11 ped/m$^2]$)	21540
E ($k \in (1.11$ - 2.17 ped/m$^2]$)	37114
F ($k > 2.17$ ped/m^2)	23750

Table 4.4: Estimation data classified according to LoS (Fruin, 1971) - Delft case study



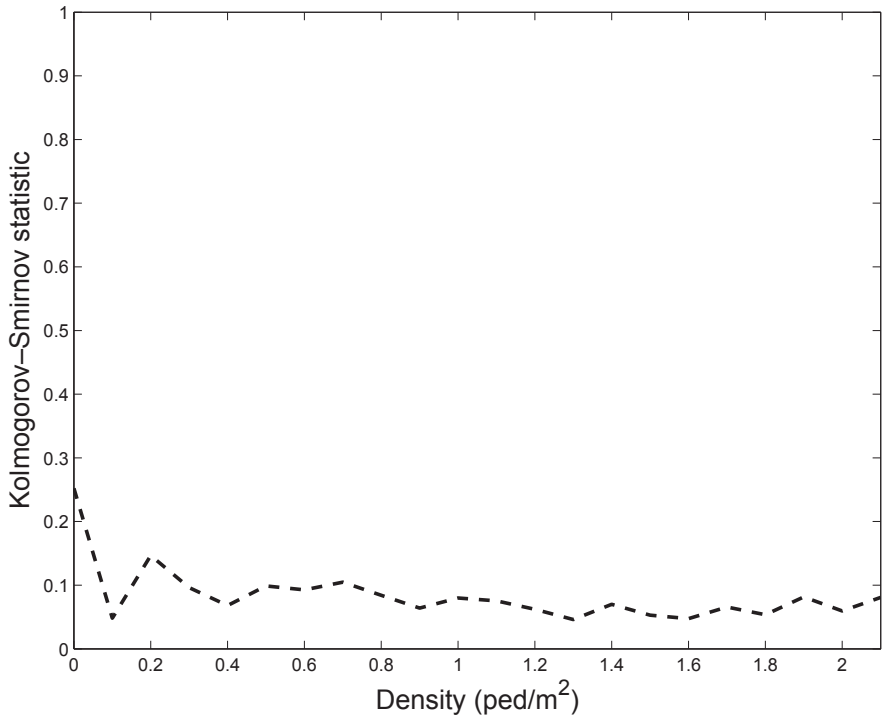
(a) k < 0.1 ped/m^2 (156,578 observations)
(b) k $= 0.2$ ped/m^2 (470,223 observations)
(c) k $= 0.4$ ped/m^2 (267,644 observations)
(d) k $= 0.6$ ped/m^2 (148,500 observations)
(e) k $= 0.8$ ped/m^2 (130,962 observations)
(f) k $= 1$ ped/m^2 (77,909 observations)
(g) k $= 1.2$ ped/m^2 (47,727 observations)
(h) k $= 1.5$ ped/m^2 (23,828 observations)
(i) k $= 2$ ped/m^2 (11,148 observations)

Figure 4.10: Comparison between model predictions (probability density) and empirical observations - Lausanne case study

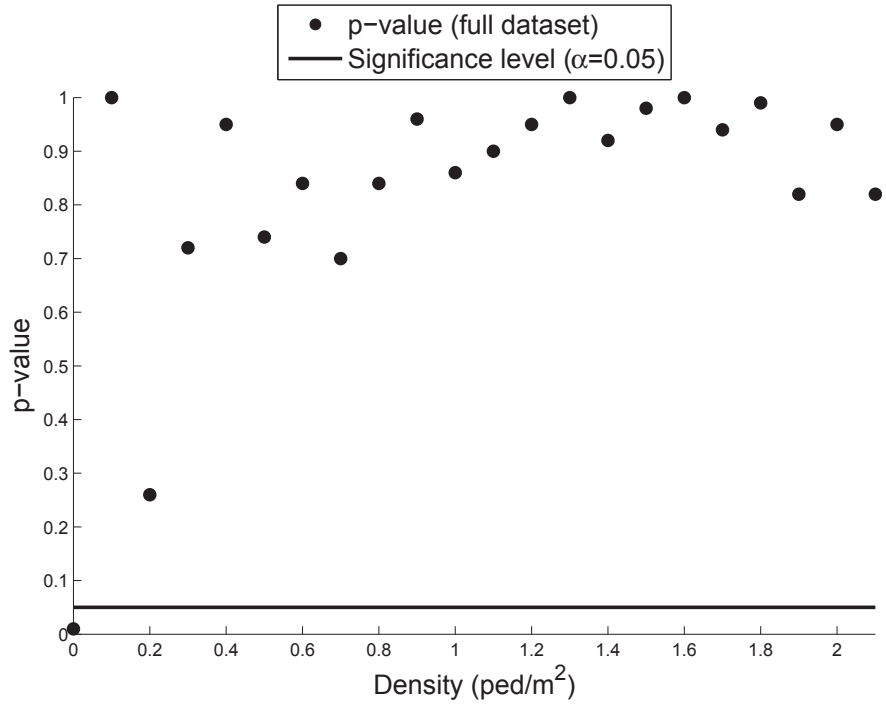


(a) k < 0.1 ped/m^2 (156,578 observations)
(b) k = 0.2 ped/m^2 (470,223 observations)
(c) k = 0.4 ped/m^2 (267,644 observations)
(d) k = 0.6 ped/m^2 (148,500 observations)
(e) k = 0.8 ped/m^2 (130,962 observations)
(f) k = 1 ped/m^2 (77,909 observations)
(g) k = 1.2 ped/m^2 (47,727 observations)
(h) k = 1.5 ped/m^2 (23,828 observations)
(i) k = 2 ped/m^2 (11,148 observations)

Figure 4.11: Comparison between model predictions (cumulative density) and empirical observations - Lausanne case study

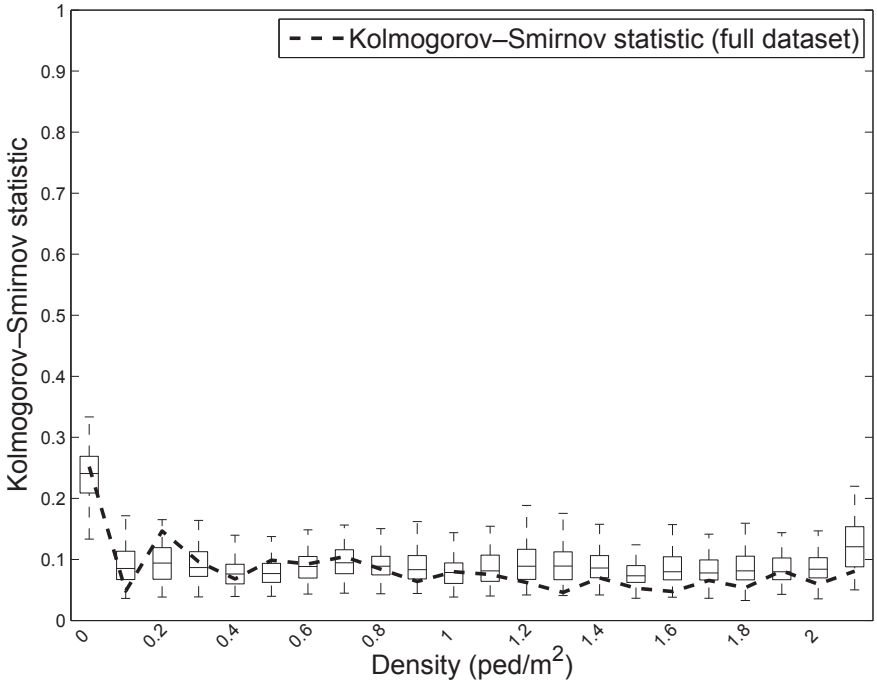


(a) Kolmogorov-Smirnov distance

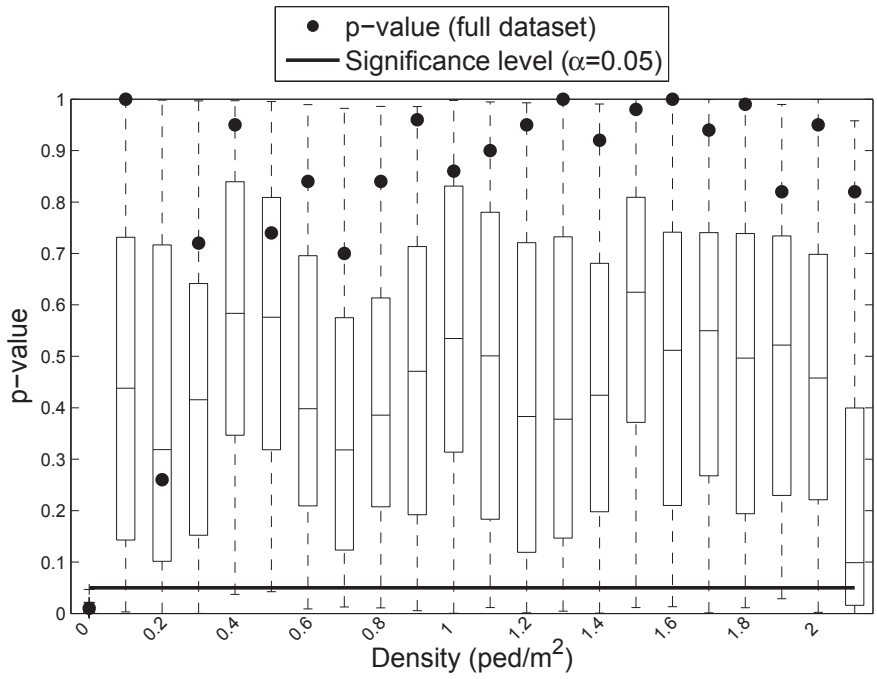


(b) p-value of Kolmogorov-Smirnov statistic

Figure 4.12: Kolmogorov-Smirnov validation - Lausanne case study



(a) Kolmogorov-Smirnov statistic



(b) p-value of Kolmogorov-Smirnov statistic

Figure 4.13: Specification test - Lausanne case study

Parameter	Value	Std err (block bootstrap)	t-test
a_α	0.165	$6.62e^{-02}$	2.08
b_α	0.244	$6.97e^{-02}$	3.11
a_β	0.166	$6.85e^{-02}$	2.49
b_β	0.965	$5.97e^{-02}$	16.4
λ	6.90	$1.41e^{-02}$	$4.89e^{02}$
v_f	1.87	$4.92e^{-02}$	38.1
θ	1.13	$6.03e^{-02}$	18.3
γ	0.545	$4.31e^{-02}$	13.1
η	3.29	$3.28e^{-02}$	$1.00e^{02}$
$\log \mathcal{L}$	-3642.944		
Number of parameters	9		
Number of observations	119156		

Table 4.5: Estimation results - Delft case study

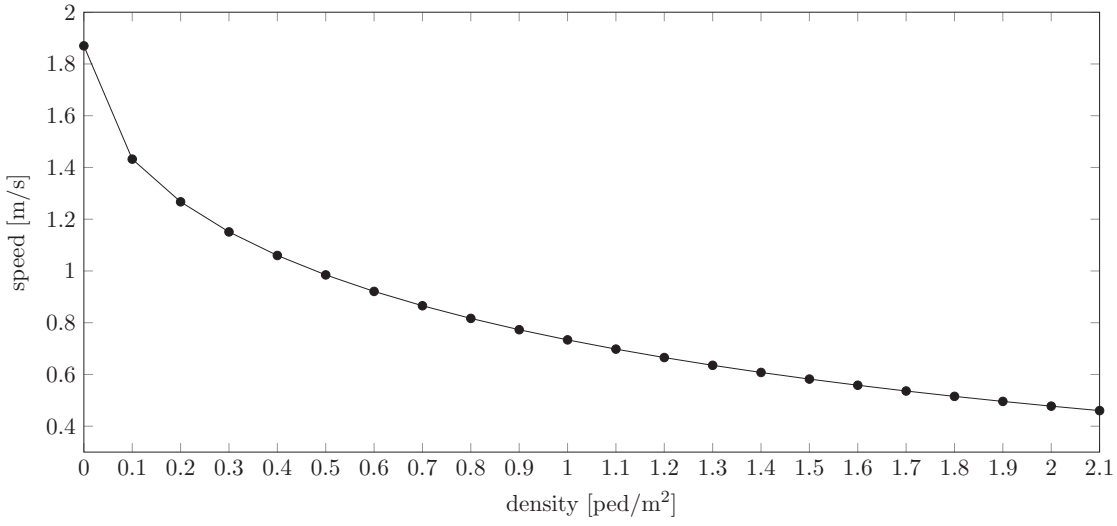
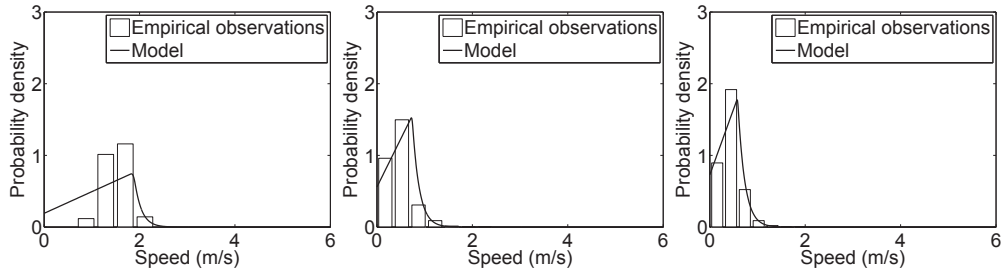


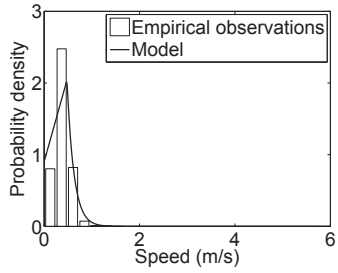
Figure 4.14: Equilibrium relationship - Delft case study



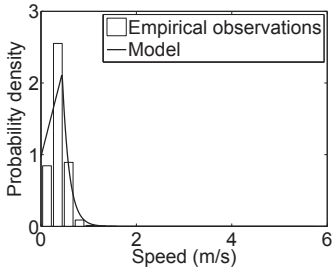
(a) k < 0.1 ped/m^2 (1,545 observations)

(b) k = 1 ped/m^2 (9,181 observations)

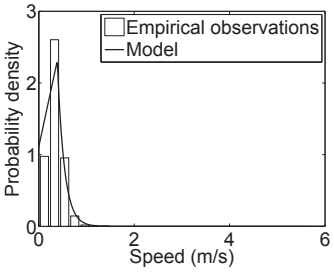
(c) k = 1.5 ped/m^2 (6,426 observations)



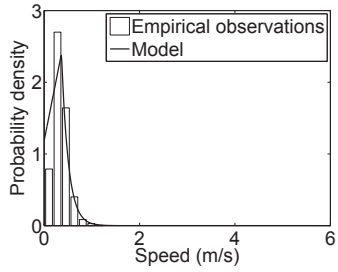
(d) k = 2 ped/m^2 (7,343 observations)



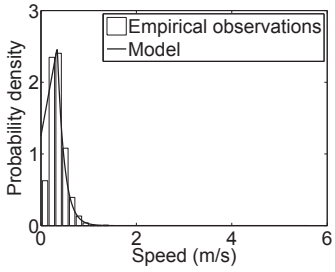
(e) k = 2.2 ped/m^2 (6,628 observations)



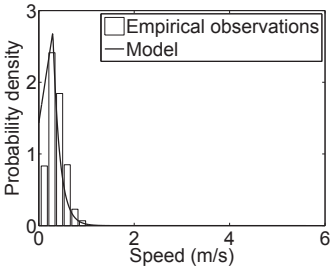
(f) k = 2.6 ped/m^2 (4,377 observations)



(g) k = 2.8 ped/m^2 (2,939 observations)

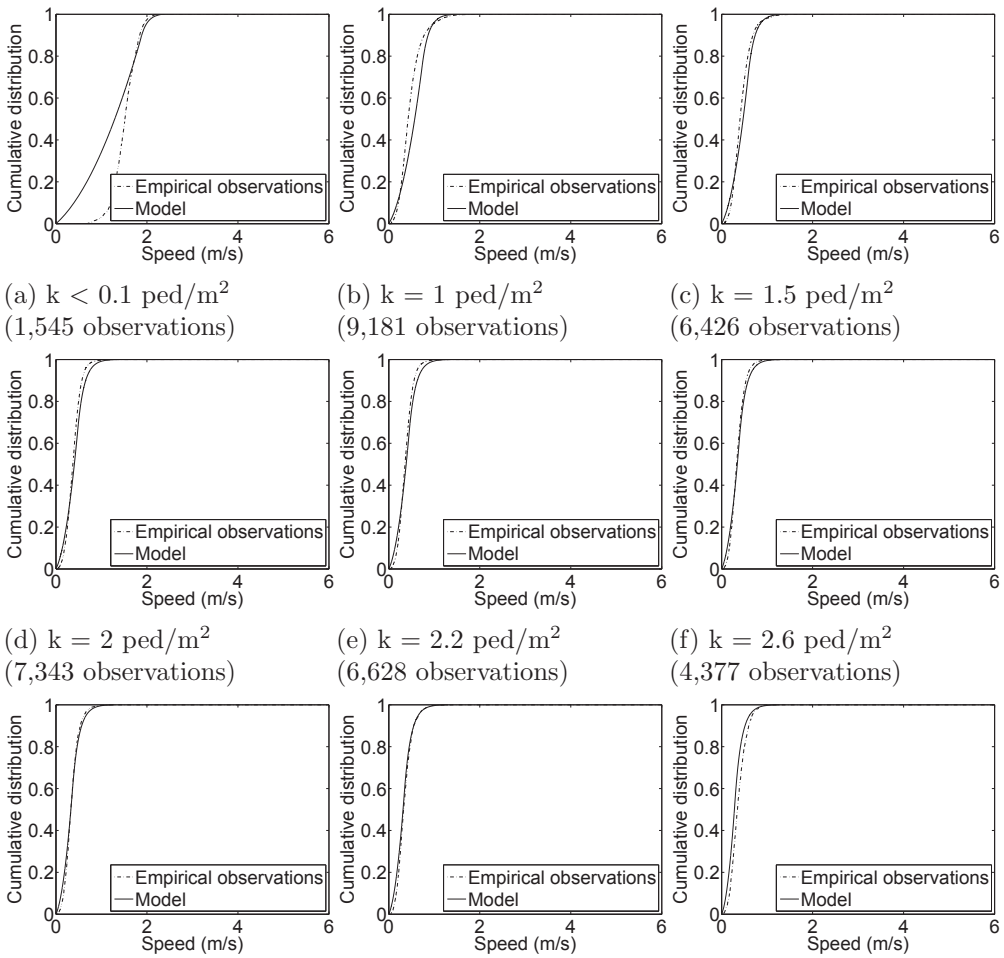


(h) k = 3 ped/m^2 (2,001 observations)



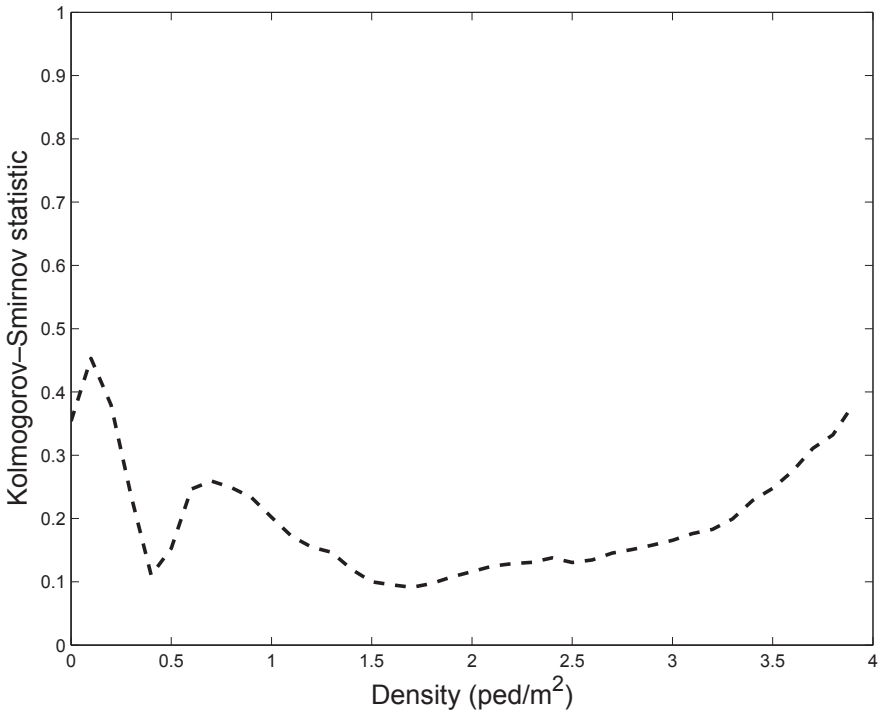
(i) k = 3.5 ped/m^2 (737 observations)

Figure 4.15: Comparison between model predictions (probability density) and empirical observations - Delft case study

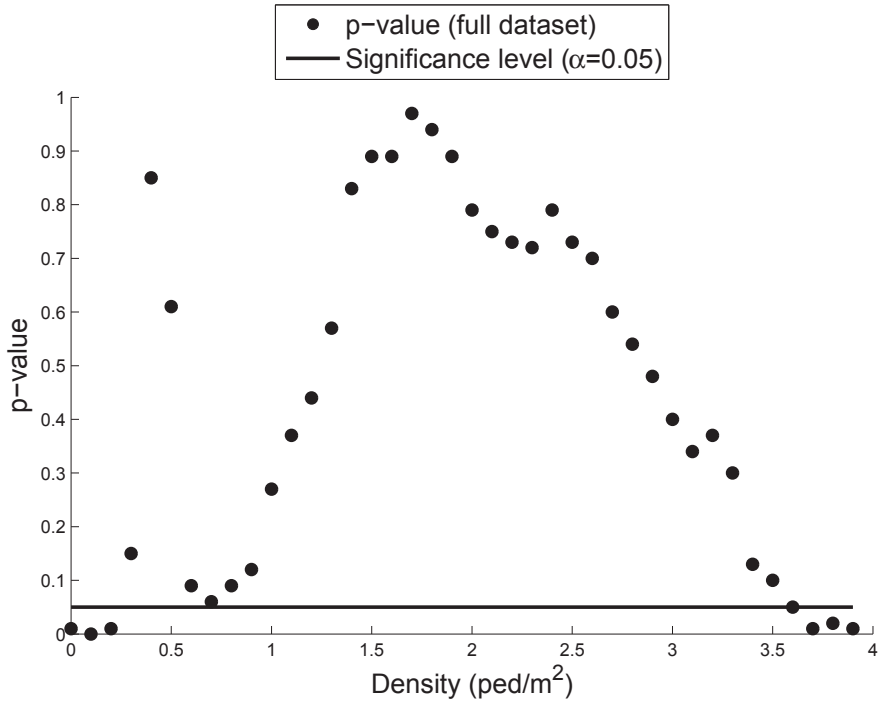


(a) k < 0.1 ped/m^2 (1,545 observations)
(b) k $= 1$ ped/m^2 (9,181 observations)
(c) k $= 1.5$ ped/m^2 (6,426 observations)
(d) k $= 2$ ped/m^2 (7,343 observations)
(e) k $= 2.2$ ped/m^2 (6,628 observations)
(f) k $= 2.6$ ped/m^2 (4,377 observations)
(g) k $= 2.8$ ped/m^2 (2,939 observations)
(h) k $= 3$ ped/m^2 (2,001 observations)
(i) k $= 3.5$ ped/m^2 (737 observations)

Figure 4.16: Comparison between model predictions (cumulative density) and empirical observations - Delft case study

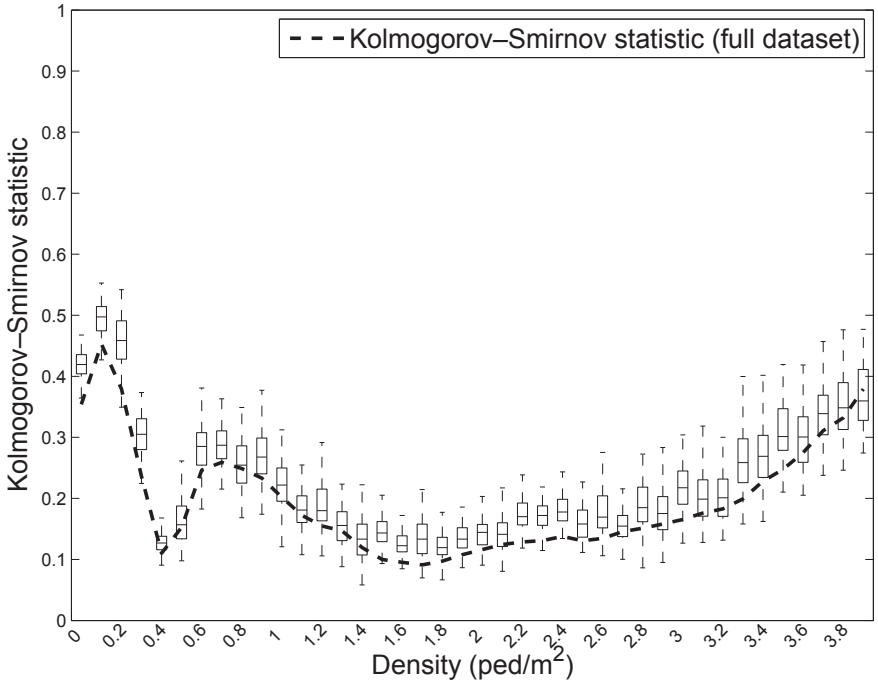


(a) Kolmogorov-Smirnov distance

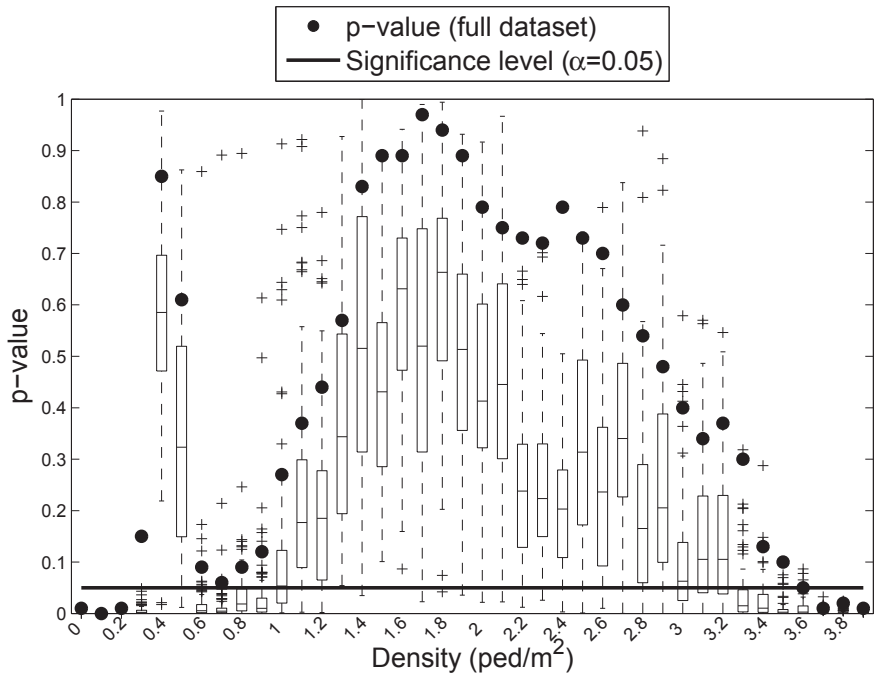


(b) p-value of Kolmogorov-Smirnov statistic

Figure 4.17: Kolmogorov-Smirnov validation - Delft case study



(a) Kolmogorov-Smirnov statistic



(b) p-value of Kolmogorov-Smirnov statistic

Figure 4.18: Specification test - Delft case study

5

Multi-class speed-density relationship for pedestrian traffic

This chapter is based on the article:

Nikolić, M., Bierlaire, M., Lapparent, M. and Scarinci, R. (2017). Multi-class speed-density relationship for pedestrian traffic. Technical report TRANSP-OR 170115.

The work has been performed by the candidate under the supervision of Prof. Michel Bierlaire, Matthieu de Lapparent PhD., and Riccardo Scarinci PhD.

The objective of this chapter is the derivation of multi-class model for pedestrian speed-density relationship. It is motivated by (i) a high scatter in real data that precludes the use of traditional equilibrium relationships, and (ii) the lack of behavior-oriented explanatory power of *PedProb-vk* (Chapter 4) when describing the observed phenomenon. The model is derived based on the latent class methodology. In addition to the general modeling framework, we present some concrete model specifications. Real data is utilized to test the performance of the approach. The approach is able to reveal fundamental properties causing the heterogeneity in population and describe their impact on pedestrian movement. We also show the advantages of the proposed approach compared to approaches from the literature. The proposed model is flexible, and it provides richer information than traditional models.

The structure of the chapter is as follows. Section 5.1 describes the proposed methodological framework for the derivation of the multi-class speed-density relationship. Section 5.2 presents the empirical analysis based on the data collected in Lausanne (Switzerland) train station. In Section 5.3 we empirically illustrate the performance of the approach. Section 5.4 summarizes the outcomes of the proposed methodology.

5.1 Methodology

This section first describes the general modeling framework and presents concrete suggestions for the model specification. Then, data requirements and model estimation method are discussed.

5.1.1 Modeling framework

Let $(v_i,\ k_i,\ X_i)$ be a triplet representing the speed v_i, the density k_i and the vector of observable characteristics X_i (such as age, trip purpose, etc.) associated with individual i. We assume that the population is partitioned into J classes (sub-populations) of pedestrians. In this framework, the individual speeds v_i are random variables. It is assumed that the speed is influenced by the prevailing density, and that this relationship varies across classes. Therefore, the distribution of v_i is characterized by its probability density function conditional on the density k_i experienced by individual i and the class j

$$f_j(v_i|k_i, j; \theta_j(k_i)), \tag{5.1}$$

where $\theta_j(k_i)$ are parameters. We refer to this distribution as a class-specific model (CSM). Each class may be characterized by a different probability distribution of the individual speeds. However, in most applications it is assumed that all class densities arise from the same parametric distribution family (Frühwirth-Schnatter, 2006). We assume that this distribution is continuous with positive support. This property is in accordance with the physical characteristic of the speed, being that the speed is a continuous variable whose values cannot be negative. Note that the density k_i is not class dependent, as all pedestrians contribute to the density, irrespectively of the class they belong to. The specification of the parameters $\theta_j(k_i)$, which characterize the class-specific speed distribution, is assumed to vary with the density of pedestrians. The assumption is motivated by empirical observations (Cheung and Lam, 1998) that suggest different trend of speed distribution for different density levels (e.g. the mean and the spread of the speed distribution decrease with increase in the density of pedestrians). For instance, this dependency can be represented using deterministic speed-density relationships, such as those presented in Table 2.3. Some concrete examples are shown in Section 5.1.2.

The class of pedestrian i cannot be directly observed. Therefore, we propose a class membership model (CMM) that provides the probability that a pedestrian i, characterized by her socio-economic characteristics X_i, belongs to class j

$$\Pr(j|X_i;\beta_j), \tag{5.2}$$

where β_j are parameters. The CMM can take a number of forms. A typical assumption is based on a fitness function, that is a continuous variable measuring how much individual i fits into class j. For example, a linear formulation would consist in

$$U_{i,j} = V_{i,j} + \varepsilon_{i,j} = CSC_j + \beta_j X_i + \varepsilon_{i,j}, \tag{5.3}$$

where CSC_j and β_j are unknown parameters to be estimated from data, and $\varepsilon_{i,j}$ is a random term. The assumption is that the individual belongs to the class with the highest value of the fitness function. A specific distribution assumption for $\varepsilon_{i,j}$ leads to a specific probability model. The exact specification of $V_{i,j}$, and in particular the exact list of characteristics involved in X_i, is application dependent. We show some examples in Section 5.1.2.

The *multi-class speed-density model* (MC-vk) is obtained by combining the CSM and the CMM as follows

$$f_{MC-vk}(v_i|k_i, X_i; \theta_j(k_i), \beta_j) = \sum_{j=1}^{J} \underbrace{f_j(v_i|k_i, j; \theta_j(k_i))}_{\text{CSM}} \underbrace{\Pr(j|X_i;\beta_j)}_{\text{CMM}}. \tag{5.4}$$

The described framework is illustrated in Figure 5.1.

The latent classes can be assumed a priori or interpreted a posteriori. The a priori specification of classes can be based on the information from the literature about different pedestrian sub-populations (e.g business and leisure travelers, or children, adults and seniors), and their preferred walking speed and the attitude towards congestion (Daamen, 2004; Weidmann, 1993). The a posteriori interpretation of latent classes should be supported by the estimation results.

5.1.2 Exemplary specification

For the illustration of the model, we need to assume the number of classes (J), and to specify the exact form of the CMM and the CSM. We assume the existence of two classes ($J = 2$), denoted as class C_1 and class C_2, in order to keep the model parsimonious and to avoid potential over-fitting. Note that other models with higher number of classes are estimated for the train station context and compared using statistical tests in Section 5.3.1.3. The results of this comparison show that the two-class model is

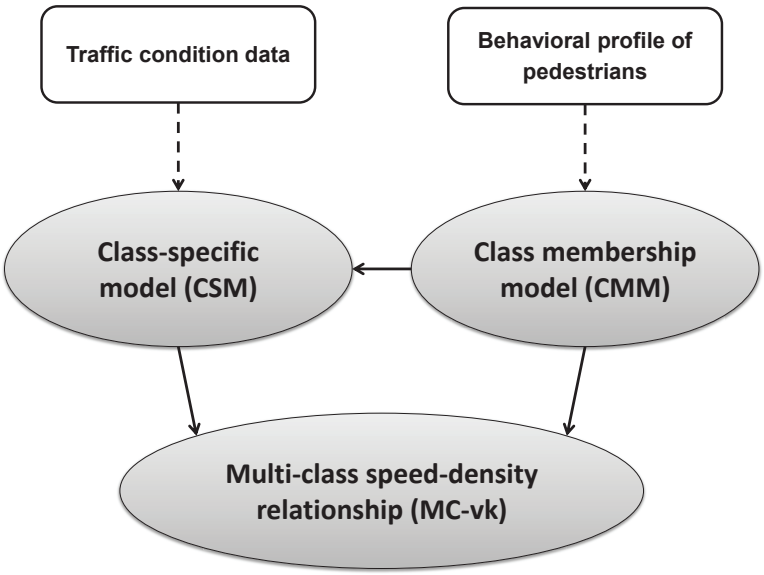


Figure 5.1: Modeling framework

superior for this case study.

We suggest the Rayleigh model for the distribution of the speed in each class

$$f_j(v_i|k_i, j, \theta_j(k_i)) = \frac{v_i}{\theta_j^2(k_i)} \exp(-\frac{v_i^2}{2\theta_j^2(k_i)}), \tag{5.5}$$

which is driven by only one parameter, the scale $\theta_j(k_i)$. The choice of the Rayleigh distribution is motivated by its properties (continuous distribution that is defined on the positive support) that are in accordance with the physical properties of the speed. The mean of a Rayleigh random variable is expressed as $\mu_j(k_i) = \theta_j(k_i)\sqrt{\pi/2}$. We model the mean as the linear class-specific equilibrium speed-density relationship for all classes

$$\mu_j(k_i) = v_{f,j} - \gamma_j k_i, \tag{5.6}$$

where $v_{f,j}$ and γ_j are class-specific parameters, referring to the free-flow speed, respectively the sensitivity to congestion. These parameters are expected to vary across classes such that they reflect the class-specific behavior. The equilibrium relationship in (5.6) is derived in Chapter 4, Section 4.1.2 from the social force model proposed by Helbing and Molnar (1995), for isotropic, homogenous and stationary traffic conditions. It also corresponds to the relationships proposed by Older (1968), Navin and Wheeler (1969), Fruin (1971), Tanaboriboon et al. (1986) and Lam et al. (1995) (Table 2.3).

Each pedestrian is assumed to belong to one class only (pedestrians do not switch among classes over time). This is consistent with the literature on vehicular traffic (van Wageningen-Kessels, 2013). We assume that the error term $\varepsilon_{i,j}$ in (5.3) is i.i.d. type 1 Extreme Value (EV1(0,1)) across classes and individuals. This assumption yields the binary logit CMM, defined as

$$\Pr(j|X_i;\beta_j) = \frac{e^{V_{i,j}}}{\sum_{j=1}^{2} e^{V_{i,j}}}. \tag{5.7}$$

The quality of the CMM depends on the available information about the behavioral profiles of pedestrians, that constitutes the deterministic part of the model ($V_{i,j}$). For instance, various studies have shown that, in general, the age (children, adults, seniors), the gender (female, male) and the trip purpose (leisure, commuters, shoppers, business) of pedestrians have an effect on their movement behavior (Chapter 2, Section 2.3). If this information is available, the deterministic parts of the fitness function for each class and pedestrian can be defined as

$$\begin{aligned} V_{i,1} &= CSC_1 + \beta_{CHILD,1} CHILD_i + \beta_{ADULT,1} ADULT_i + \beta_{FEMALE,1} FEMALE_i \\ V_{i,2} &= \beta_{LEISURE,2} LEISURE_i + \beta_{COMMUTERS,2} COMMUTERS_i + \\ &\quad \beta_{SHOPPERS,2} SHOPPERS_i, \end{aligned} \tag{5.8}$$

where $SENIOR$, $MALE$ and $BUSINESS$ are considered as the reference levels of the corresponding discrete variables. Depending on the context, some additional information may be also useful. For instance, in transportation hubs, such as train stations, metro stations or airports, the type of passenger (arriving, departing, transferring), the time to departure, the distance that pedestrians need to traverse, the presence of luggage and walking in groups appear as relevant factors. Similarly, the opening hours of shops in commercial centers, or highlights in museums, interacted with the cultural background of pedestrians, can be valuable in explaining the class membership.

Note that the framework described in Section 5.1.1 is general and allows for different specifications to be tested. To assess the performance of our approach, we present the analysis on a real case study in Section 5.3.

5.1.3 Data requirements

In order for the presented framework to be applied, the following data must be available. The key type of data, like for any such model, is the traffic condition data. It includes individual speed and density observations, necessary for the estimation of the model. They can be extracted from the individual trajectory data, that is the data provided in the form of individual-specific pairs of consecutive time and location observations.

Ideally, trajectories are collected using precise pedestrian tracking systems with high temporal resolution. For instance, the systems based on optical, thermal and depth sensors (Alahi et al., 2011, 2014) or digital cameras (Daamen and Hoogendoorn, 2003). Pedestrian trajectories can be also obtained using wireless technologies such as WiFi or Bluetooth. The issue with these technologies is the low temporal resolution and strong sample bias (Danalet, 2015). In this case, the combination of WiFi or Bluetooth traces with count data may provide better understanding about prevailing traffic conditions (Hänseler, 2016).

Although the class membership model could be specified without any explanatory variable, the quality of the model would benefit from characteristics of individuals. This means that the more information is available about pedestrian characteristics, the easier it will be to obtain a good class membership model. This information includes typical socio-economic and demographic data, such as age, gender, health conditions, culture, trip purpose, etc. As mentioned in Section 5.1.2, depending on the type and the purpose of pedestrian facilities, additional types of data would be useful. For instance, the apparent types of data are (i) the type of passenger, the time to departure, the presence of luggage and walking in groups in public transport facilities, (ii) the points of pedestrians' interest in museums and the attractivity of these points, (iii) opening hours of shops and restaurants in commercial centers, (iv) concert schedules and toilets' locations in music festivals, (v) pollution and noise levels experienced by pedestrians in urban streets, etc. To collect the mentioned characteristics various recall methods may be used, including paper-based surveys distributed to individuals (Bachu et al., 2001; Kalakou and Moura, 2014), smartphone-based applications (Ohmori et al., 2005; Cottrill et al., 2013; Ball et al., 2014; Zhao et al., 2015), and web-based methods (Bohte and Maat, 2009).

In case data collection using recall methods is not performed, it is possible to compensate, to some extent, by relying on the available information from pedestrian trajectories and some other sources (e.g. attributes of a facility, timetables and schedules in transportation hubs, etc.). This is demonstrated in Section 5.2.2.

5.1.4 Estimation procedure

With respect to traffic condition data, we assume the availability of individual observations collected over multiple time instants. This means that traffic condition data has panel nature, and the speed observations of a single individual are likely to be correlated across time. This is usually referred to as serial correlation. The issue arises due to unobserved individual factors that persist over time. If ignored it leads to consistent but inefficient estimators (Gourieroux et al., 1984). To address the issue of serial correlation we introduce an agent effect, α_i, in the mean of the speed distribution that drives the

CSM (Wooldridge, 2005). We assume that it is exponentially distributed

$$f(\alpha_i;\mu_{\alpha_i}) = \frac{1}{\mu_{\alpha_i}} \exp(-\frac{\alpha_i}{\mu_{\alpha_i}}), \tag{5.9}$$

where μ_{α_i} is the mean. The agent effect is assumed to be independently and identically distributed across pedestrians, but remains constant within the observations of a given pedestrian. The choice of the exponential specification is motivated by its positive support. Also, it is well suited to avoid any arbitrariness in imposing the upper bound, while preventing arbitrary high values at the same time. The likelihood conditional on class j for the observations of pedestrian i in the described panel setting is as follows

$$f(v_{i1},...,v_{iT_i}|k_{i1},...,k_{iT_i},j;\theta_j(k_i)) = \int_{\alpha_i} \prod_{t=1}^{T_i} f_j(v_{it}|k_{it},j,\alpha_i;\theta_j(k_{it}))f(\alpha_i;\mu_{\alpha_i})d\alpha_i, \tag{5.10}$$

where T_i is the number of observations of individual i, and $v_{i1},...,v_{iT_i}$ and $k_{i1},...,k_{iT_i}$ are the sequences of speed, respectively density observations associated with this individual.

Contrary to traffic condition data, pedestrian characteristics (e.g. age, gender, etc.) are assumed to be time independent during the observation period. Each pedestrian is associated with one observation of the considered characteristics. For instance, the age category of a pedestrian is *adult* and her gender is *female*. This means that the class membership probability is calculated only once, and it remains constant across all speed-density observations for a given pedestrian (the error terms $\varepsilon_{i,j}$ (5.3) are not time dependent). Therefore, the issue of serial correlation does not appear at the CMM level. We also assume that the error terms of the CSM and the CMM are not correlated, as in Walker and Li (2007).

Combining the CSM and the CMM, the contribution of individual i to the likelihood is given as

$$\begin{aligned} &f(v_{i1},...,v_{iT_i}|k_{i1},...,k_{iT_i},X_i;\theta_j(k_i),\beta_j) = \\ &\sum_{j=1}^{2} \left\{ \int_{\alpha_i} \prod_{t=1}^{T_i} f_j(v_{it}|k_{it},j,\alpha_i;\theta_j(k_{it}))f(\alpha_i;\mu_{\alpha_i})d\alpha_i \right\} \Pr(j|X_i;\beta_j). \end{aligned} \tag{5.11}$$

The integral in (5.11) is approximated via simulation as

$$\begin{aligned} &f(v_{i1},...,v_{iT_i}|k_{i1},...,k_{iT_i},X_i;\theta_j(k_i),\beta_j) = \\ &\sum_{j=1}^{2} \left\{ \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T_i} f_j(v_{it}|k_{it},j,\alpha_{i,r};\theta_j(k_{it})) \right\} \Pr(j|X_i;\beta_j), \end{aligned} \tag{5.12}$$

where R refers to the number of draws from $f(\alpha_i;\mu_{\alpha_i})$. The likelihood function for the

sample of N individuals ($i = 1, ..., N$) is given by

$$\mathcal{L} = \prod_{i=1}^{N} f(v_{i1}, ..., v_{iT_i} | k_{i1}, ..., k_{iT_i}, X_i; \theta_j(k_i), \beta_j) = \prod_{i=1}^{N} \left\{ \sum_{j=1}^{2} \left\{ \frac{1}{R} \sum_{r=1}^{R} \prod_{t=1}^{T_i} f_j(v_{it} | k_{it}, j, \alpha_{i,r}; \theta_j(k_{it})) \right\} \Pr(j | X_i; \beta_j) \right\}, \tag{5.13}$$

that is to be maximized.

5.2 Case study and empirical analysis

We illustrate the methodology and analyze its performance using the Lausanne case study, described in Chapter 4, Section 4.3.1. In the underpass (Figure 4.4), 33 depth sensors are installed. The placement of the sensors is such that the major part of the underpass is monitored (Alahi et al., 2013). However, blind areas exist (the areas that are not covered by the sensors), where missing data are completed using an inter-sensor tracking algorithm (Alahi et al., 2011). In this chapter we make a distinction and call the observations form the covered areas the "detected" observations, and those from the uncovered areas the "imputed" observations. This distinction is made in order to reduce the effect of errors due to technological issues in the calculation of the density indicator, as elaborated in Section 5.2.1.

In addition to detailed pedestrian trajectories, the infrastructure data, that is detailed plans containing the locations and dimensions of all relevant parts of the monitored system, is also available. We also have access to the train timetable for the period under study. The arrival and departure times and the assigned tracks are thus known for all trains.

We first present the analysis of traffic condition data, that is speed and density observations. We then discuss the factors used to explain the class membership of individuals.

5.2.1 Speed and density observations

The speed-density profile corresponding to the Lausanne case study (Figure 5.3) is obtained from the Voronoi-based measurement method presented in Chapter 4, Section 4.1. As described, the Voronoi space decomposition assigns a personal region to each pedestrian i, based on the positions of pedestrians. The position of "detected" observations is considered to be accurate, while the position of "imputed" observations might be subject to inter-sensor tracking algorithm errors (Alahi et al., 2011). Therefore, the Voronoi spatial discretization is in this chapter performed based on "detected" observations only

(Figure 5.2). We, however, need to account for the existence of the pedestrians whose observations are marked as "imputed". To do so, the density at each point $p = (x, y, t_s)$ is computed as

$$k(x, y, t_s) = \frac{n_{V_{is}}^{detected} + n_{V_{is}}^{imputed}}{|V_{is}|}, (x, y) \in V_{is}, \tag{5.14}$$

where $t_s = (t_0, ..., t_f)$ corresponds to the available sample, $|V_{is}|$ is the area of Voronoi cell V_{is} assigned to "detected" pedestrian i at time t_s, and $n_{V_{is}}^{detected}$ and $n_{V_{is}}^{imputed}$ refer to the number of "detected", respectively "imputed" observations within the cell V_{is}. The speed is approximated using finite differences, based on the "detected" observations (Chapter 4, Section 4.1).

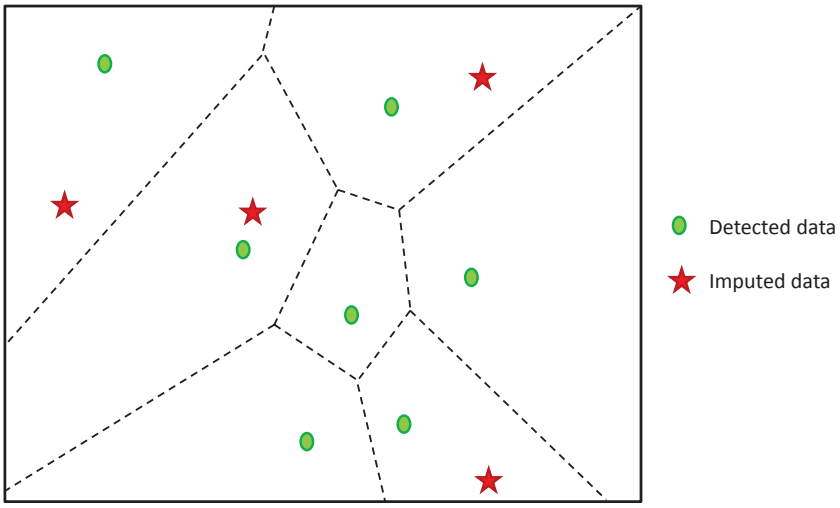


Figure 5.2: Spatial discretization based on "detected" observations

In Figure 5.3, each circle corresponds to one observation, that is, one pedestrian at one specific time. The x coordinate of the circle corresponds to the density, and its y coordinate corresponds to the speed. The figure plots 154,417 observations corresponding to the peak hour of February 13, 2013. The same pattern was observed on any weekday.

A high scattering is observed in Figure 5.3. The density ranges from 0 to approximately 7 pedestrians per square meter, and 99% of the observations are below 1.4 pedestrians per square meter. The speed ranges from 0 to approximately 6 meters per second, and 99% of the observations are below 2.42 meters per second.

The deterministic models for the speed-density relationship proposed in the literature

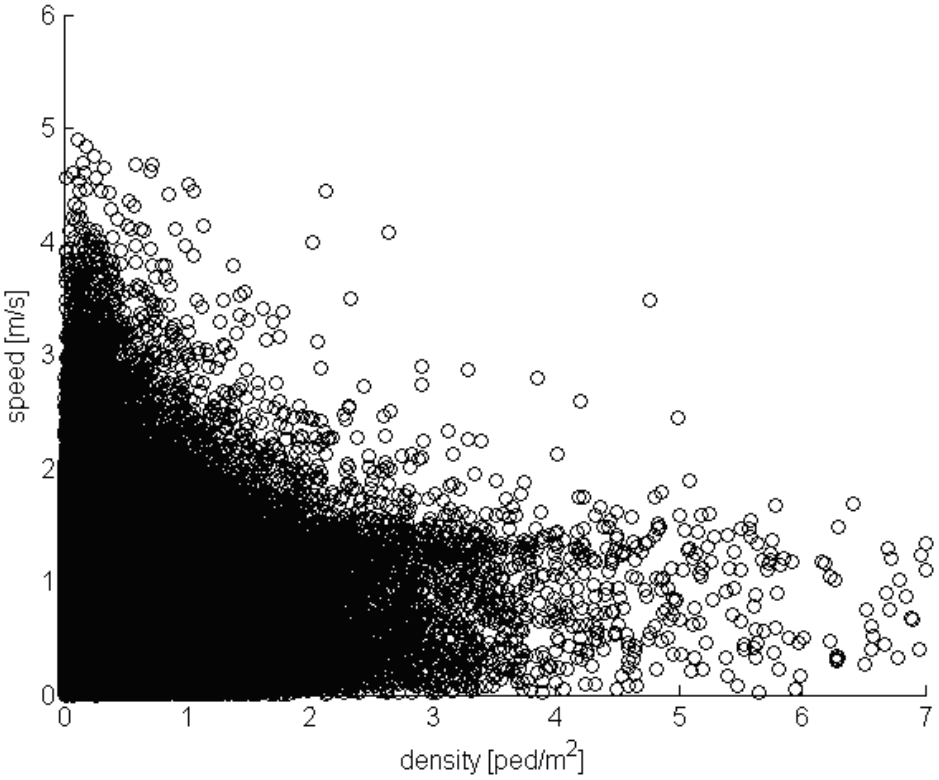


Figure 5.3: Speed-density profile

(Chapter 2 , Section 2.3.1) appear to be inadequate for representing the observed pattern.

5.2.2 Characteristics associated with pedestrians

In addition to the speeds and densities, we extract additional variables that are used in the model: the pedestrian type, the OD distance, the peak and off-peak periods and the time to departure. In the following, we explain each factor separately.

The pedestrian type refers to the classification of pedestrians based on their OD pairs. The definition of OD areas depends on the layout of a train station, and it includes the stairs and ramps to the platforms (denoted as P1-P9 in Figure 4.4) and the entrance/exit areas. We consider four pedestrian types

1. arriving passenger (AP) - pedestrians originating from a platform and exiting the station,
2. departing passenger (DP) - pedestrians walking to a platform to embark on their trains,

3. transferring passenger (TP) - pedestrians whose origin and destination are different platforms,

4. non-passenger (NP) - pedestrians whose origin and destination are not a platform (e.g. pedestrians that go shopping in the station, or use the underpass to reach the other side of the city).

Between 8% and 9% of pedestrians each day are not classified, due to the mismatch of their initial and/or final observations with any of the predefined zones that indicate origins and destinations. These observations are considered for the calculation of density, but are not taken into account at the level of the CMM. Figure 5.4 shows the percentage of pedestrians belonging to each of the four types across days. The pattern is relatively stable over days, and it indicates that the majority of pedestrians are arriving and departing. Figure 5.5 illustrates the speed distributions for these two types of pedestrians,

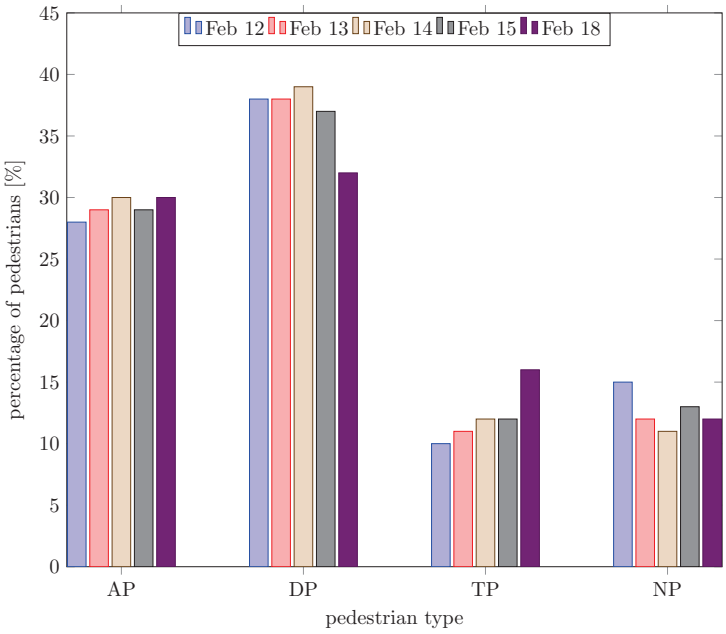


Figure 5.4: Number of pedestrians per pedestrian type

for a particular density level. It shows that arriving passengers tend to walk with lower speeds, compared to departing passengers. The shift of the distribution for departing passengers towards higher values can be explained by the fact that they are more likely to be under time pressure (to embark on the trains) than the arriving ones.

The OD distance (OD) is associated with each pedestrian based on the corresponding OD pair. The OD distances are calculated as the shortest Euclidean physical distance between each origin and destination. The distances range from approximately 3 meters to 80 meters, as shown in Figure 5.6.

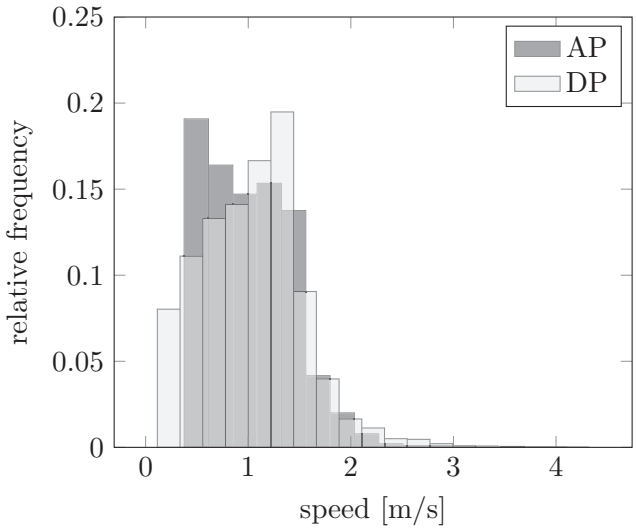


Figure 5.5: Speed distribution per pedestrian type (k=0.6 ped/m^2)

The peak and off-peak periods are defined based on the number of people observed over time for each day (Figure 5.7). For the temporal aggregation we consider the intervals that are 5 minutes long. Peak periods (PP) refer to the intervals where the local maxima of the number of people are observed, and off-peak periods (OPP) refer to all other intervals. For February 13, 2013, for instance, the peak periods are 07:15-07:20 and 07:40-07:45. We expect that pedestrians are characterized with different walking behavior, depending on the period when they are observed.

The time to departure (TTD) is obtained by exploiting the information contained in the train timetable. We define the time to departure as the difference between the departure time of the next train from the platform that the pedestrian is going to and the time at which the pedestrian is first observed in the underpass. The distribution of time to departure is shown in Figure 5.8. It ranges from a few seconds to approximately 50 minutes. The distribution suggests that most of the people arrive to the train station approximately 3 minutes (the mode of the distribution) before the train departure. It is natural to assume that people that have more time to the departure of their trains behave differently from those rushing to catch their trains.

We also explored the impact of the group behavior on the speeds of pedestrians. We adopted spatial clustering together with temporal frequent patterns analysis to identify the pedestrians walking in groups. In the spatial clustering step, we used the values proposed by McPhail and Wohlstein (1982) for the features characterizing pedestrians that walk in groups (e.g. distances pedestrians keep between each other, differences in speeds and directions). However, less than one percent of the population was identified to walk in groups, and the analysis with respect to this factor showed no significant effect on the speed at which pedestrians move. The factor is therefore excluded from

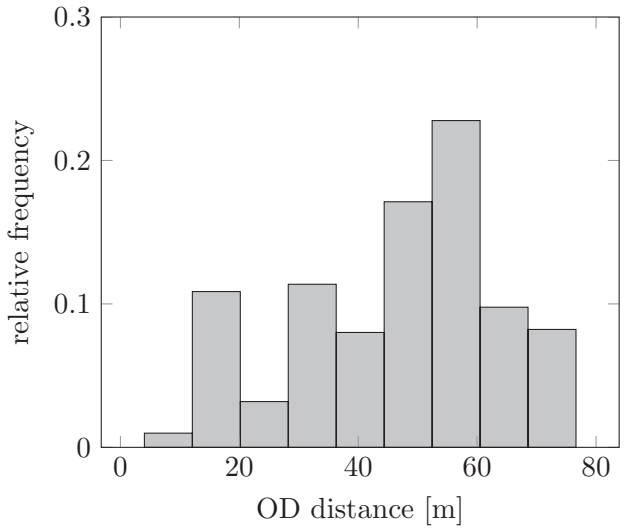


Figure 5.6: OD distance distribution

further analysis.

To investigate the case study data in more details, we consider the speed distributions with respect to the mentioned factors at various density levels. The analysis is presented in Appendix C, Figure C.1 - Figure C.4 and Table C.2. The findings are in agreement with our assumption on the pedestrian heterogeneity reflected through different walking speeds. They suggest that density is not the only factor influencing pedestrians' speed and that the observed heterogeneity in speed values might come from multiple factors.

5.3 Applying the framework

The framework proposed in Section 5.1 is illustrated on the Lausanne case study. The CSM is specified in the model (5.5). The CMM is specified as the logit model given in (5.7), with the deterministic parts of the fitness function defined as

$$\begin{aligned} V_{i,1} &= CSC_1 + \beta_{DP,1} DP_i + \beta_{TP,1} TP_i + \beta_{NP,1} NP_i \\ V_{i,2} &= \beta_{TTD,2} TTD_i + \beta_{PP,2} PP_i + \beta_{OD,2} OD_i, \end{aligned} \tag{5.15}$$

where the variables are described in Section 5.2.2. This CMM specification is motivated by the analysis presented in Section 5.2. We present below the model estimation results and detailed examination of each component of the model. Also, comparisons with the existing models and practical implications are discussed.

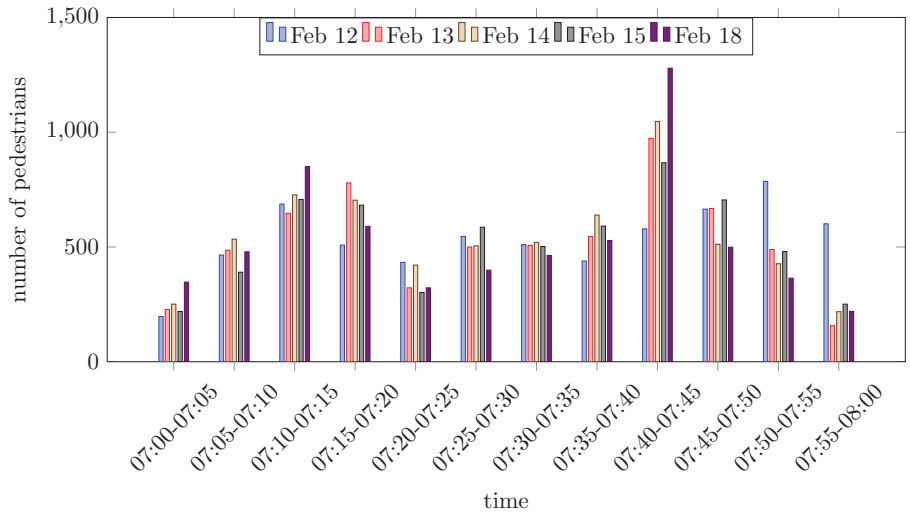


Figure 5.7: Number of pedestrians grouped into intervals of 5 minutes

5.3.1 Model estimation

For the estimation of the model parameters, we use maximum likelihood procedure. The estimation results of the presented model are shown in Table 5.1. We use $R = 200$ draws from (5.9) for the simulation of the integral in (5.13). The standard errors are calculated using bootstrapping (Chapter 4). All estimates have the expected sign and value. The results also show the low standard errors of the parameters and their statistical significance at a usual significance level (0.05).

5.3.1.1 Class-specific model

The parameter estimates of the class-specific models show how movement behavior vary across classes.

The signs and the estimated values of the parameters $v_{f,1}$,γ_1,$v_{f,2}$ and γ_2 are consistent with the ones reported in the literature. The corresponding class-specific speed-density relationships are shown in Figure 5.9.

The parameters μ_{α_1} and μ_{α_2} show that agent effect distributions are characterized with similar mean in both classes.

We make inferences on the movement behavior of each class based on the estimated parameters of the CSM. The C_1 class pedestrians are characterized with a higher free-flow walking speed and significantly lower sensitivity to congestion, compared to pedestrians belonging to C_2 class. We thus call C_1 the class of "pedestrians less sensitive to congestion" and C_2 the class of "pedestrians more sensitive to congestion". The classes are

Parameter	Value	Std err	t-test
CSC_1	-0.682	$2.82e^{-02}$	$-2.35e^{01}$
$\beta_{DP,1}$	2.02	$7.50e^{-03}$	$2.69e^{02}$
$\beta_{TP,1}$	1.46	$2.90e^{-02}$	$4.99e^{01}$
$\beta_{NP,1}$	7.16	$2.63e^{-02}$	$2.72e^{02}$
$\beta_{TTD,2}$	0.00125	$2.61e^{-04}$	4.95^{00}
$\beta_{PP,2}$	1.31	$2.00e^{-02}$	$6.58e^{01}$
$\beta_{OD,2}$	0.0207	$2.65e^{-03}$	$8.24e^{00}$
$v_{f,1}$	1.08	$2.24e^{-03}$	$4.84e^{02}$
γ_1	0.0478	$3.05e^{-03}$	$1.52e^{01}$
$v_{f,2}$	0.984	$3.26e^{-02}$	$2.90e^{01}$
γ_2	0.186	$2.51e^{-03}$	$7.38e^{01}$
μ_{α_1}	0.0985	$4.09e^{-03}$	$2.32e^{01}$
μ_{α_2}	0.121	$1.87e^{-02}$	$5.98e^{00}$
$\log \mathcal{L}$	-519050.63		
Number of parameters	13		
Number of observations	747385		

Table 5.1: Estimation results

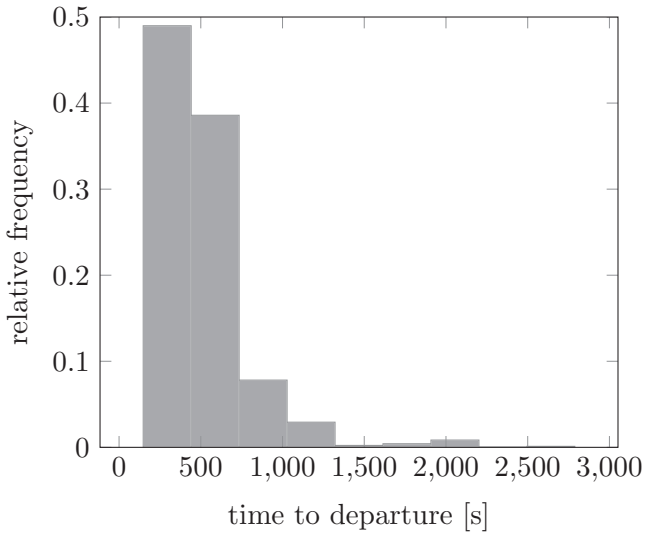


Figure 5.8: Distribution of time to departure

relevant for the context of pedestrian movement in train stations.

5.3.1.2 Class membership model

The parameter estimates of the class membership model show what are the underlying factors leading to different movement behavior.

The class-specific constant CSC_1 has negative sign denoting that, the rest of fitness functions being equal, pedestrians in the train station are less likely to belong to C_1 class than to C_2.

The positive sign of the parameters $\beta_{DP,1}$, $\beta_{TP,1}$ and $\beta_{NP,1}$ indicates that departing, transferring and non-passengers are more likely to be in class C_1, compared to arriving passengers.

The positive sign of the parameters $\beta_{TTD,2}$ and $\beta_{OD,2}$ shows that as pedestrians have more time to departure and longer distances to traverse, they are more likely to belong to C_2. Similarly, the sign of the parameter $\beta_{PP,2}$ shows that pedestrians are more likely to be of C_2 during peak periods with respect to off-peak periods.

Figure 5.10a shows the average probability over the sample of belonging to each class

$$w_j = \frac{\sum_{i=1}^{N} \Pr(j|X_i;\beta_j)}{N}, \tag{5.16}$$

where N refers to the number of individuals. According to the predictions of the model,

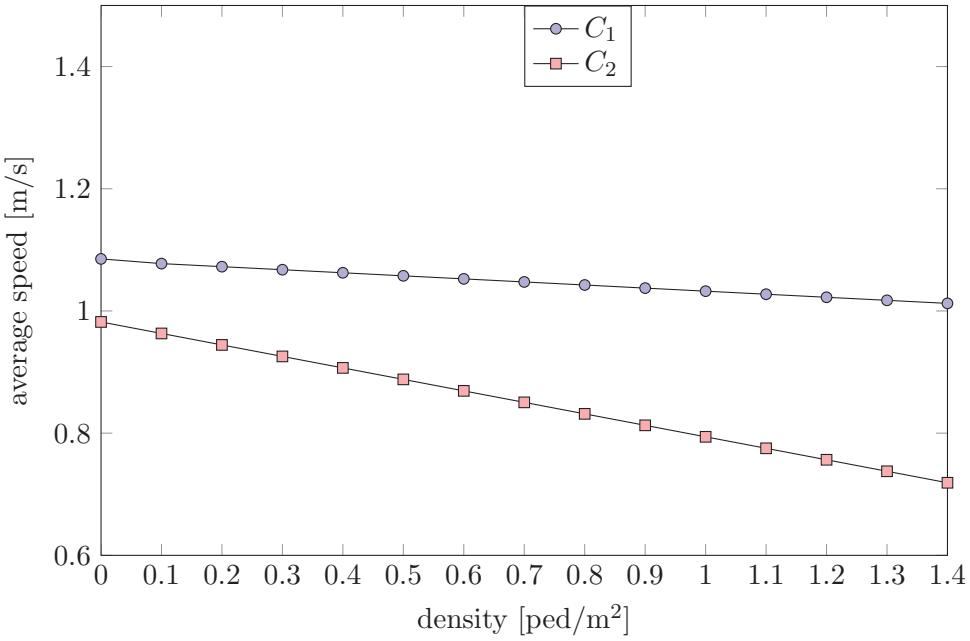
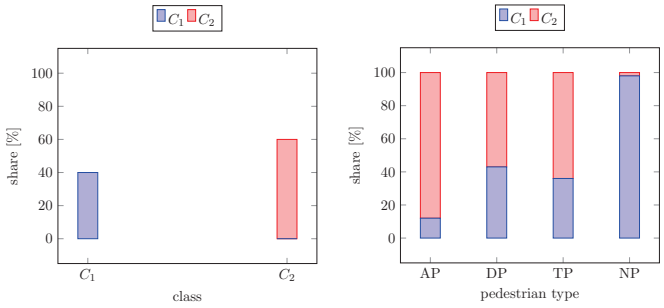


Figure 5.9: Class-specific speed-density relationships

39.77% of pedestrians are "pedestrians less sensitive to congestion" (C_1) and 60.23% are "pedestrians more sensitive to congestion" (C_2). Figure 5.10b indicates the higher share of C_1 class in the case of departing and transferring passengers, compared to arriving passengers. This is expected, given that departing and transferring passengers are usually associated with higher time pressure that drives their more "aggressive" behavior. It is however not clear whether the high share of C_1 class in the case of non-passengers can be related to some behavioral aspects, or to the fact that OD pairs of non-passengers correspond to the main flow direction in the underpass. It is documented in the literature that pedestrians constituting the main flow are characterized by higher walking speeds (Wong et al., 2010).

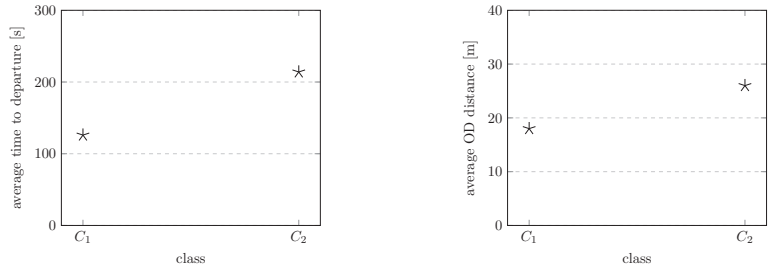


(a) Overal (b) Per pedestrian type

Figure 5.10: Shares per class

Figure 5.11 shows the predicted value of time to departure and OD distance for pedestri-

ans in each class. The predicted values represent the weighted average of each variable, where the weight is the probability of being in a particular class. As expected, pedestrians in class 1 (less sensitive to congestion) on average have less time to departure and shorter distances to traverse, compared to pedestrians in class 2 (more sensitive to congestion).



(a) Avg. time to departure (b) Avg. OD distance

Figure 5.11: Class profiling

The results indicate the existence of relatively strong profiling. However, the segmentation of a population is not deterministically determined. Even though the parameter estimates in Table 5.1 suggest that the characteristics associated with pedestrians significantly influence the segmentation of considered population, a probabilistic model is necessary.

5.3.1.3 Alternative specifications

We have also performed a sensitivity analysis to investigate the potential existence of one and three sub-populations. The suggested model represents a form of a clustering analysis of data. Compared to the standard clustering methods, it is a probabilistic model-based approach. The issue of determining the correct number of classes is therefore reduced to model selection problem in the probabilistic framework. We asses the goodness of the model by the means of the Bayesian information criterion (Wasserman, 2000)

$$BIC = -2 \cdot \log \mathcal{L} + m \cdot \log(n), \tag{5.17}$$

where n is the number of observations, m is the number of unknown parameters and $\mathcal{L}$ is the value of the likelihood function of the model. The statistics are reported in Table 5.2, indicating that models with multiple classes are preferred over a single-class model. Although the best value of the $\log \mathcal{L}$ is achieved for the three-class model, the BIC statistic suggests that the two-class model represents the best compromise between the model accuracy and simplicity among all evaluated specifications. It also provides the most satisfactory behavioral interpretation of the results.

Various alternative specifications of the model with two classes have been investigated, and the results are listed in Table 5.3. They differ from the proposed model in terms of assumed class-specific speed-density relationships and agent effect distributions. Linear speed-density relationships in Table 5.3 correspond to those proposed by Older (1968), Navin and Wheeler (1969), Fruin (1971), Tanaboriboon et al. (1986) and Lam et al. (1995) (Table 2.3). Exponential speed-density relationships refer to the relationship proposed by Rastogi et al. (2013) (Table 2.3). All alternative specifications resulted in a poorer fit with respect to the proposed model (BIC=1038277.07, Table 5.2), according to the BIC.

5.3.2 Qualitative analysis

In this section we analyze in a qualitative way how well the model represents the data. The analysis considers the distribution functions of the estimated model, and the empirical distributions from the dataset. It is carried out at different density levels. Figure 5.12 shows the probability density functions of the model (model pdf) and the data (empirical histogram) at the same density level. Qualitatively, the match between the two is pretty satisfactory. In particular this is the case for the density levels which are characterized by the largest number of observations (Table 5.4). Note that unimodality of distribution in the probabilistic model proposed in Chapter 4, Section 4.2.1 is also consistent with the *MC-vk* (Figure 5.12).

Figure 5.13 shows the predictions of the models proposed in the literature (Table 2.3) at the aggregate level. The density and space-mean speed are in this case measured in a cell A using (2.19), respectively (2.20). The cell is placed in the middle of the corridor (Figure 4.4) and its area is $|A| = 1m^2$. Note that the scattered pattern also appears when more aggregate measures are considered (Figure 5.13). We have estimated the parameters of the deterministic models from Table 2.3 using linear regression on the measured values. All the considered models appear to be inadequate for representing the observed phenomena. The goodness of fit across the models is similar and shows significant discrepancy between observed values and the values predicted by the models. For instance, the sum of squared errors of prediction ranges from 271.8 to 273.2, the adjusted coefficient of determination is around -9.835e^{-05} and the root-mean-square error is in the range from 0.3339 to 0.3348.

The *MC-vk* is more complex than traditional speed-density relationships. Our exemplary specification has 13 parameters (Table 5.1), while the deterministic models proposed in the literature have from two to three parameters (Table 2.3). Yet, the *MC-vk* relies on relatively few assumptions and it is easy to analyze by using simulation. Moreover, the presented analysis suggests that it results in the predictions that are more consistent with empirical observations.

5.3.3 Posterior analysis

The estimated model allows for the calculation of posterior probabilities of class-membership for each observation

$$\Pr(j|v_i, k_i, X_i; \theta_j(k_i), \beta_j) = \frac{\Pr(j|X_i; \beta_j) f_j(v_i|k_i, j; \theta_j(k_i))}{\sum_{j=1}^{2} \Pr(j|X_i; \beta_j) f_j(v_i|k_i, j; \theta_j(k_i))}. \tag{5.18}$$

The assessment of posterior probabilities allows to assign an observation to multiple classes with different degrees of membership, or to perform the so-called "soft" clustering. This form of clustering is desirable when the information about probabilities is useful. It provides more information than classical methods, where each observation belongs to only one class ("hard" clustering).

Figure 5.14 shows the posterior class-membership probabilities of the observations in the speed-density plot for the class of "pedestrians less sensitive to congestion" (C_1). They indicate how strongly each observation belongs to this class. Darker gray color indicates higher posterior probability, and lighter gray color corresponds to lower posterior probability. The results indicate relatively strong separation of the observations. It can be observed that higher speed values belong to C_1 class more strongly than lower speed values, for all density levels. The results of the analysis are in line with our expectations. They show the capability of the model to capture the structure of the data better than the traditional models.

5.3.4 Practical implications

The proposed model can be used by the operators as an instrument for policy making (e.g. long run planning) and daily operations (e.g. to control the flow). For instance, using the estimated CMM it is possible to examine the influence that the modification of the explanatory variables values has on the split of pedestrians among the classes. We look at the effect of the reduction of time to departure for departing and transferring passengers, assuming all the rest remains unchanged. Such a reduction is typically the result of the modification of the train time table in the train station. The reduction of 10%, 20%, 30%, 40% and 50% is considered. Figure 5.15 indicates the increase in the share of "pedestrians less sensitive to congestion" class, and decrease in the share of "pedestrians more sensitive to congestion" class, with the decrease in time to departure. This observation is in accordance with our expectations.

The presented analysis suggests that the model can be utilized for the analysis of the effects of such scenarios on the movement behavior of pedestrians. This is important for a variety of applications, such as the impact of different scenarios on the resulting LoS within train stations. Also, the evaluation can be further augmented by posterior

analysis of the class membership probabilities (Section 5.3.3). Moreover, the proposed methodology is not case specific. It is general and can be simply adopted to other specific applications.

5.4 Summary

In this chapter, a latent class model for speed-density relationship for pedestrian traffic is proposed. Differently from approaches in the literature, it is a multi-class model designed to account for the heterogeneity of speed observed in the data. The model uses the latent classes to relax the homogeneity assumption of equilibrium speed-density relationships.

To illustrate the proposed methodology, we use the data collected in the train station in Lausanne, Switzerland. The analysis confirms the existence of multiple classes of pedestrians characterized by different movement behavior. The resulting behavior is relevant for the studied situation. In addition to density, the observed pattern is explained by factors related to pedestrian type, train timetable and infrastructure. The proposed model thus represents a flexible tool for (i) recognizing the main factors driving the heterogeneity in population and (ii) describing the impact of heterogeneity on pedestrian movement. Being conceptually insightful, the model can be further used to support the derivation of pedestrian flow models from the first principles.

We present various analyses using empirical data that illustrate the performance of the model. In contrast to the existing approaches, the suggested model features a behavior-oriented explanatory power. As such, it provides more realistic representation of the observed phenomena, and it is better suited for forecasting analysis.

Model	1 class	2 classes	3 classes
$\log \mathcal{L}$	-529000.76	-519050.63	-519007.51
# parameters	3	13	23
# observations	747385	747385	747385
BIC	1058042.09	1038277.07	1038326.07

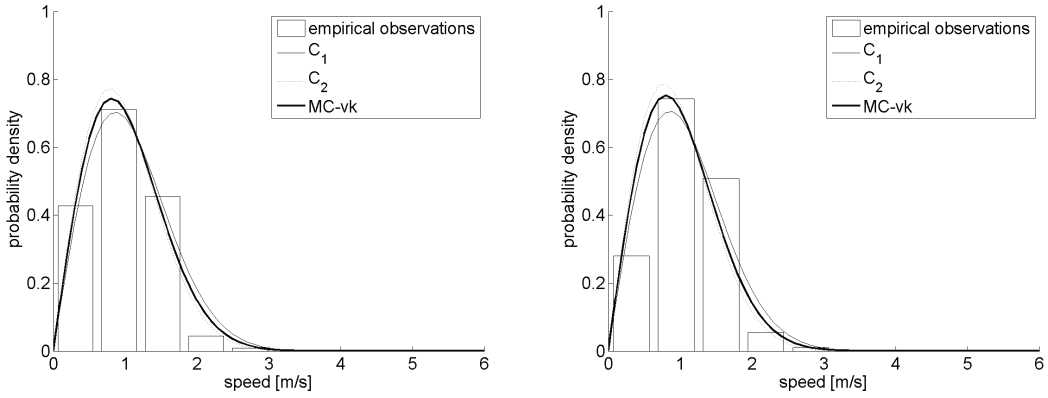
Table 5.2: Goodness of fit - BIC

Model	M_a	M_b	M_c	M_d	M_e
Speed-density rel. C_1	Linear	Linear	Exponential	Linear	Linear
Speed-density rel. C_2	Linear	Exponential	Exponential	Exponential	Exponential
Agent effect distribution C_1	Rayleigh	Rayleigh	Rayleigh	Log-normal	Exponential
Agent effect distribution C_2	Rayleigh	Rayleigh	Rayleigh	Log-normal	Exponential
$\log \mathcal{L}$	-519952.55	-525143.19	-525064.29	-529320.93	-519070.04
# parameters	13	13	13	13	13
# observations	747385	747385	747385	747385	747385
BIC	1040080.92	1050462.20	1050304.40	1058817.69	1038315.91

Table 5.3: Goodness of fit for alternative specifications - BIC

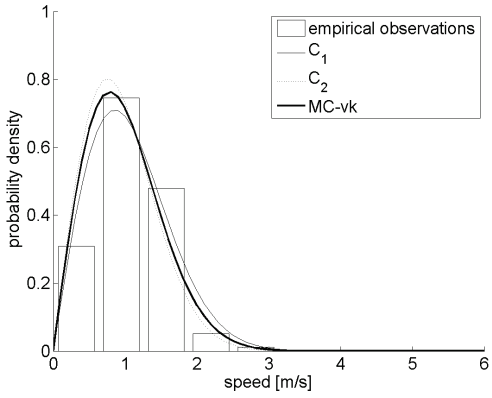
LoS	Number of observations
A ($k \leq 0.31$ ped/m^2)	417032 (56%)
B ($k \in (0.31$ - 0.43 ped/m^2])	98964 (13%)
C ($k \in (0.43$ - 0.71 ped/m^2])	128335 (17%)
D ($k \in (0.71$ - 1.11 ped/m^2])	76918 (10%)
E ($k \in (1.11$ - 2.17 ped/m^2])	26136 (4%)

Table 5.4: Estimation data classified according to LoS (Fruin, 1971)

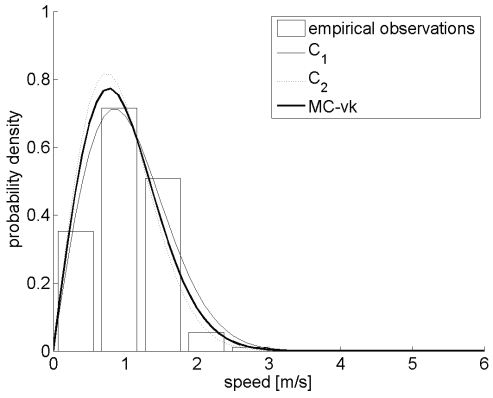


(a) k < 0.1 ped/m^2

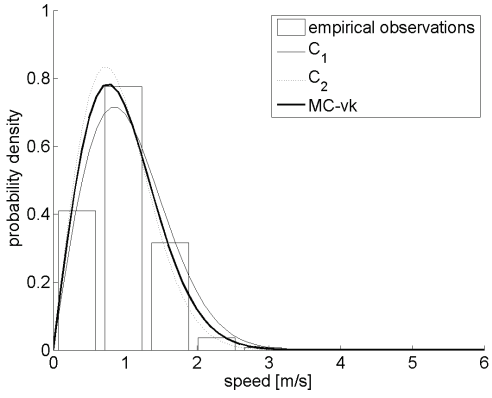
(b) k $= 0.1$ ped/m^2



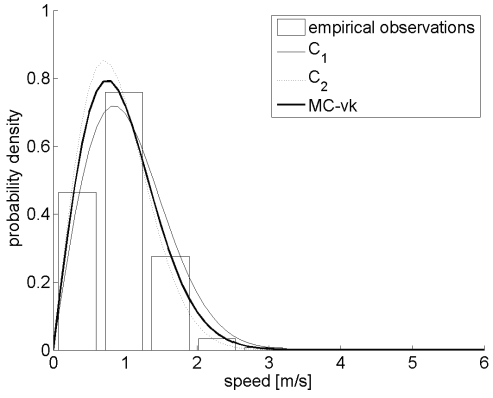
(c) k $= 0.2$ ped/m^2



(d) k $= 0.3$ ped/m^2



(e) k $= 0.4$ ped/m^2



(f) k $= 0.5$ ped/m^2

Figure 5.12: Comparison between model predictions and empirical observations

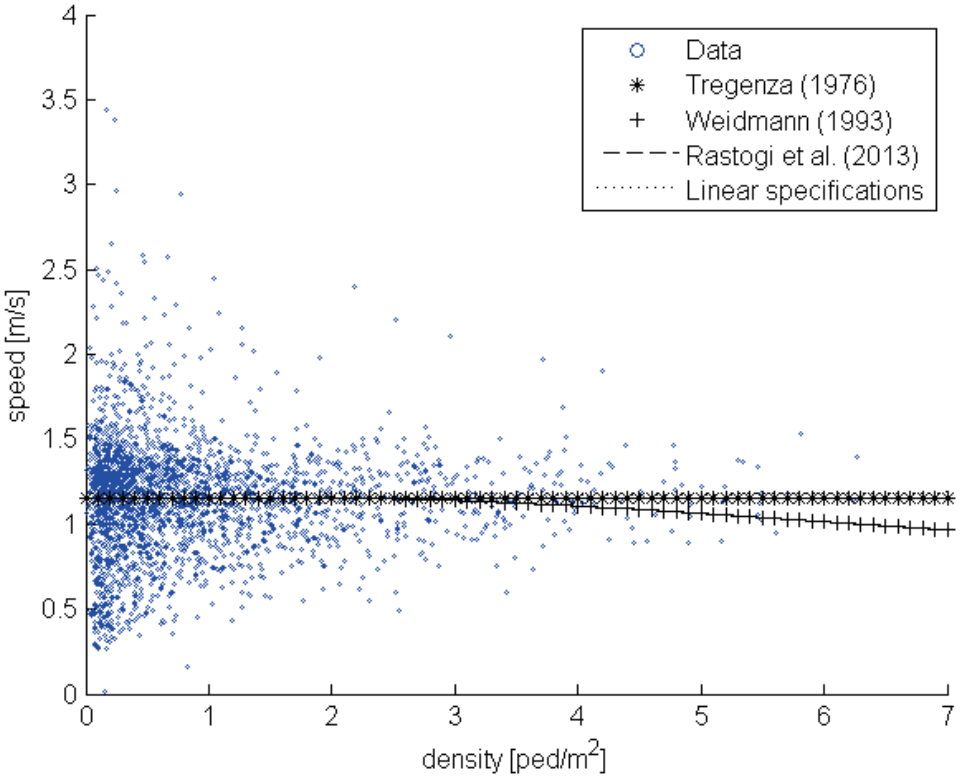


Figure 5.13: Deterministic models

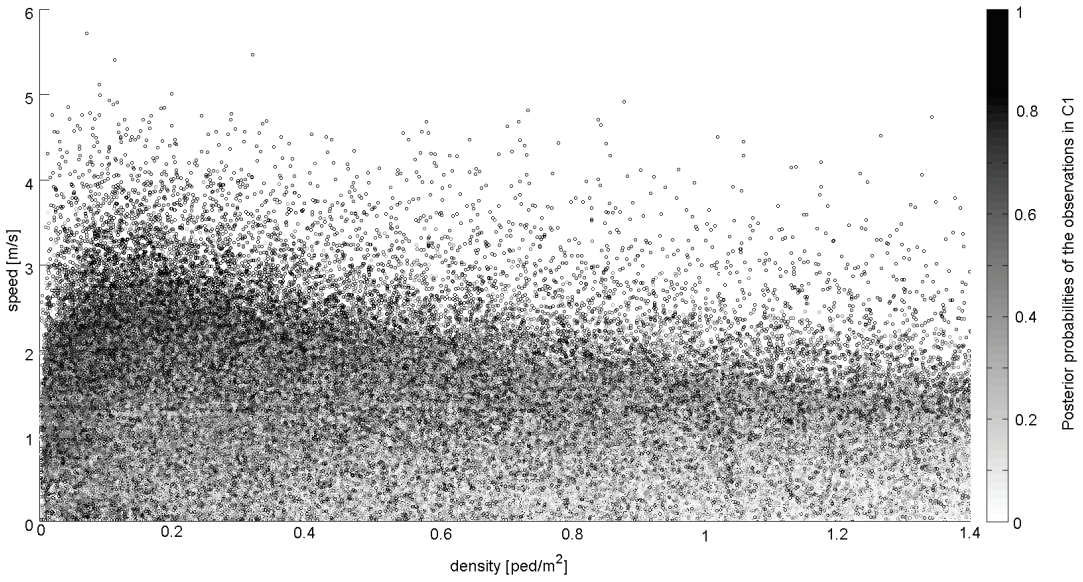


Figure 5.14: Posterior probabilities of the observations for C_1

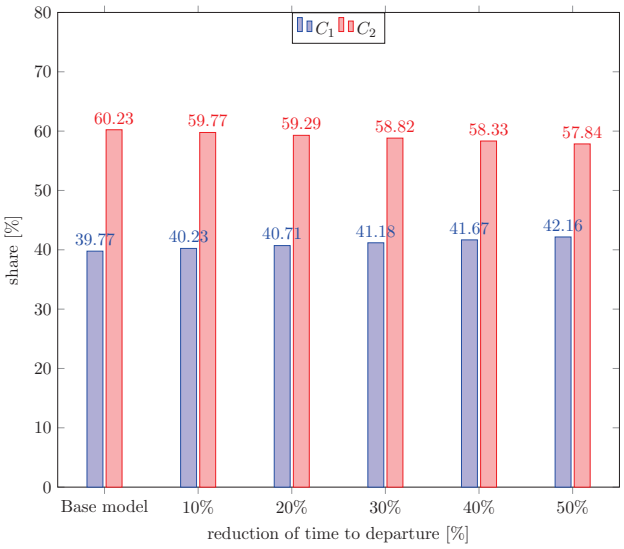


Figure 5.15: Scenario analysis - train timetable modification

6

Conclusion

In this concluding chapter, we first summarize the main findings of the presented research (Section 6.1). Then, the implications of the research for both science and practice are described (Section 6.2). Finally, we discuss some aspects that require further investigation and outline future research directions (Section 6.3).

6.1 Main findings

The main conclusion of the research presented in this thesis is that the questions related to complex phenomena of pedestrian movements can be answered engaging the latest revolutions in modern digital technologies. Data collected through innovative sensors opens the door to fresh insights and ideas in this field, and thus stimulate a paradigm shift. One aspect involves data-driven approach. The data-driven approach that we have put forward has led to a substantial refinement of fundamental theories and models in relation to pedestrian movements. The usefulness and applicability of this approach is demonstrated in this thesis using different case studies. The principal contributions of new methodologies that we propose are listed in the following.

We have proposed a novel methodology for pedestrian traffic characterization. The definitions of pedestrian traffic variables that we develop are based on data-driven partitioning in space and time. As such, they resolve the issue of arbitrary selection of spatial discretization units or spatial influence functions typical for the methods existing in the

literature. Also, our approach for the first time provides the way for data-driven discretization in temporal domain. Therefore, it allows for the characterization that is independent from arbitrarily chosen discretization intervals, and subsequently leads to more reliable results.

The discretization framework is based on spatio-temporal Voronoi diagrams designed through the utilization of pedestrian trajectories. It is applicable to pedestrian trajectories described analytically, but also as a sample of points. This is important for practical implementations, when the analytical description of a trajectory is hardly available and one often deals with discrete observations. On the other hand, the existing methods can be applied on trajectories described either analytically or as a synchronized sample of points. Otherwise, they might lead to the underestimation of the indicators.

The proposed definitions of pedestrian traffic variables are also microscopic. Thus, they address the issue of the existing methods that rely on macroscopic approach, which does not always comply with the nature of the underlying system. Thanks to macroscopic nature, the characterization of pedestrian flows that we put forward is by design appropriate for the multi-directional composition of pedestrian traffic. It is also able to reflect the heterogeneity of the pedestrian population.

Synthetic data is used to empirically investigate the performance of our pedestrian flow characterization and to illustrate its advantages. Our approach outperforms the considered approaches from the literature, in terms of the smoothness of the results, and in terms of the robustness with respect to the simulation noise and with respect to the sampling frequency. As such, the approach is advantageous for observing and modeling pedestrian movements in a realistic and robust manner.

An explorative data analysis, using the proposed definitions of fundamental traffic characteristics, has revealed a pattern in speed-density relationship that is not possible to capture using traditional models. We propose a fundamentally different approach, and revise the underlying assumptions of traditional models to account for the nature of the data. Unlike deterministic approaches in the literature, our approach is probabilistic. The traditional equilibrium assumption with respect to macroscopic relationships is relaxed to account for the heterogeneity of pedestrian population. The first model (*PedProb-vk*) is a probabilistic relationship that is able to implicitly account for the heterogeneity of pedestrian flows. The second model (*MC-vk*) combines a data-inspired and a behavioral approach using probabilistic latent class methodology. The heterogeneity of speeds is, in this case, explicitly explained using information on train timetables, pedestrian type and infrastructure data. The models provide the way to explore the impact of pedestrian heterogeneity on macroscopic relationships of pedestrian flow. The *PedProb-vk* model is rather descriptive than explanatory. On the other hand, the *MC-vk* model represents a tool useful for recognizing the main factors driving the heterogeneity in the population. It also allows statistical inference about latent classes.

Various tests validate empirically the specifications of the models representing speed-density relationships for pedestrians. For this purpose, two different and extensive case studies, data from a real scene and from controlled experiments, are employed. The ability of the models to reproduce real-world data is also illustrated. In contrast to the existing approaches, the suggested models provide more realistic representation of the observed phenomena.

The presented methodologies, for both pedestrian flow characterization and modeling, are flexible and fairly general. The exact specifications can be adapted to specific situations with little effort.

6.2 Theoretical and practical implications

Detailed observations on pedestrian movements allow for new empirical findings. These findings consequently support the development of solid, empirically supported theoretical and conceptual basis, and finally the development of powerful models. This is achievable only if the methods used for the measurement and analysis are in compliance with the nature and the resolution of the data. Otherwise, the potential of the data is reduced or completely eliminated.

Our approach to pedestrian flow characterization utilizes the gain obtained using new data collection technologies. Being entirely adjusted to the data, it allows for the revelation of new empirical phenomena. This has a direct effect on pedestrian flow modeling, by influencing the modeling approach and underlying assumptions.

The proposed probabilistic approach to model macroscopic flow relationships (MC-vk) allows investigating and explaining observed phenomena through the impact of pedestrian heterogeneity and the operations in pedestrian infrastructures. As such, it allows formulating new theories from first principles, based on empirical evidence. These models can be combined with a conservation principle in dynamic continuum and discrete models, leading to probabilistic conservation laws for the representation of the pedestrian motion. This would allow the detailed analysis of the effects of heterogeneity on pedestrian dynamics.

Moreover, the presented work has also practical implications. Proposed methodologies may be utilized by practitioners as such or jointly with other models of pedestrian movements. The variables speed, density and flow represent the main level of service indicators, and the definitions proposed in this thesis allow for the level of service analysis at a resolution down to individual level, and the utilization of the gain obtained by the new tracking technologies. Contrasted with existing approaches, the proposed probabilistic models yield a more realistic representation of the empirically observed phenomena. They are useful for testing scenarios when planning, designing and modifying

infrastructures or designing public transit timetables. This is beneficial in practice so as to ensure that (i) infrastructures' operations are efficient and smooth, and (ii) people enjoy their stay in infrastructures.

The proposed data-driven framework can be applied to various pedestrian facilities. In addition to train stations, the suggested models can be useful to support defining the layout of other infrastructures (e.g. shopping malls, museums or even urban walking areas), the location of stands and other facilities during festivals and similar events, or the location of ticket machines and the optimization of schedules in other transportation hubs (e.g. metro stations, airports). The proposed underlying modeling principles would remain the same, and potential adaptations concern the assumptions related to a specific context.

6.3 Future research directions

In this section we discuss some aspects that can be subject of further investigation. Also, new research directions, triggered by the research presented in this thesis, are established.

The set of distances proposed in this thesis to characterize the Voronoi diagrams can be extended. In particular, an interesting research topic would be to relate the definition of the distance with some behavioral assumptions about the pedestrian movements. For example, it has been recognized that pedestrians are affected much stronger by stimuli that appear within their vision field (Johansson et al., 2007). To account for the anisotropy of pedestrian movements, weighted versions of the proposed distances could be further studied. The weights would be modeled based on the movement direction of pedestrians and their vision field.

Future research can also be directed towards the characterization in the presence of obstacles. One possibility would be to extend a generator set from pedestrian trajectories to trajectories and areas, where areas represent obstacles. This would result in three-dimensional discretization where each pedestrian and each obstacle are associated with their own, non-overlapping, spatio-temporal units.

Another aspect that could be pursued is the consideration of speed dynamics (e.g. state dependence) in both models, *PedProb-vk* and *MC-vk*. In this case the data generating process would be different from the one assumed in this thesis, and all the above statements and estimation method would need to be reconsidered. In the presence of lagged speeds and random individual effects there is correlation between unobservables and the lagged dependent variable. The current quasi-maximum likelihood estimator would lead to inconsistent estimates, while the full information maximum likelihood estimator would involve more simulation-based inference and working on the initial value problem.

Therefore conditional maximum likelihood estimation using a correction as proposed by Wooldridge (2005) would be a good starting methodological point to extent the model.

Also, the *PedProb-vk* and *MC-vk* as such are insufficient to explain the multi-directional nature of pedestrian flows. As further steps, the possibility of addressing this issue by adapting the approach proposed in Nikolić and Bierlaire (2014) can be considered. In this case, pedestrian traffic would be considered as being composed of different streams that interact within the same space. This assumption would lead to stream-based macroscopic relationships, similar to Xie and Wong (2015). A stream definition is typically direction-based and assumed to be exogenous. For instance, it might be designed depending on the type of problem or learned from sample sets of data by utilizing the principal component analysis. This way it would be possible to capture the existence of multiple directions, which constitute the pedestrian flow pattern, at macroscopic level.

Our probabilistic macroscopic relations are well suited for real-time flow estimation and traffic forecasting applications (Jabari et al., 2014). They also form the basis for more advanced applications, such as pedestrian flow simulation and the management of pedestrian traffic (e.g. route guidance, flow control, etc.), which can be tested in future research.

In our analysis, no socio-economic information was available. Therefore, collection of pedestrian characteristics, such as age, gender, health conditions, trip purpose, etc., and their combination with already available information can be considered as well. In particular, the examination of the influence of these additional explanatory factors on the performance of the *MC-vk* would be an interesting research aspect.

One interesting direction is also the consideration of a sequential classification approach and its comparison with the proposed integrated approach employed in the derivation of the *MC-vk*. It would involve cluster analysis first (e.g. feature-based, shape-based), followed by the estimation of separate, cluster-specific traffic relationships.

In order to develop models of good performance, strong theoretical basis has to be built. Different theories have been proposed in the past, and most of their concepts are derived independently from empirical analysis (Daamen, 2004). The *MC-vk* model can be further used to support the derivation of pedestrian flow models from the first principles. Being conceptually insightful, it encourages the formulation of theoretical concepts that are empirically supported.

Finally, future research can also be directed towards the examination of the effectiveness of proposed methodologies in other behavioral situations, and in other types of infrastructure.

This thesis has explored what we can do with the data expansion within the context of fundamental models of pedestrian movements. However, there are studies concerned

about what we should do with data. In those cases the emphasis is placed on moral, ethical, and potentially legal issues created by the data growth and data availability. These issues, in addition to technical or infrastructural restrictions, are the main causes of still insufficient data from real sites. The lack of data concerns both movement and characteristics of pedestrians. The "big data" revolution is expected to surpass this problem in the near future and make heterogeneous and massive amount of information on pedestrian movement increasingly available. From an academic standpoint, data availability is essential in achieving the goals related to the convenience and safety of pedestrians.

A

3DVoro: Additional analysis

In this chapter the robustness with respect to the sampling frequency of the velocity and flow indicators is illustrated. As discussed in Chapter 3, the 3DVoro method in general leads to the smaller differences between the indicators calculated based on sample of points or interpolated trajectories and the benchmark values than the XY-T method.

It is interesting to notice that the XY-T method results in the most satisfactory estimation of speed only in $Uni_{LD-HomoPop}$ for high sampling frequency (Table A.1a). This can be explained by the simulated homogenous speed conditions that the fixed-grid discretization is able to reflect. When the sampling frequency is lower (Table A.1b), or when traffic conditions are more complex (Table A.2), the performance of this method deteriorates, in regard to the speed estimation.

Method	Mean		Mode		Median		90% quantile	
	IT	SoP	IT	SoP	IT	SoP	IT	SoP
XY-T	$4.30e^{-03}$	/	0	/	$3.40e^{-03}$	/	$1.16e^{-02}$	/
E-3DVoro	$1.55e^{-01}$	/	0	/	$3.56e^{-02}$	/	$4.99e^{-01}$	/
TT_1-3DVoro	$9.60e^{-03}$	$2.31e^{-02}$	0	0	$2.20e^{-03}$	$9.38e^{-03}$	$2.79e^{-02}$	$4.85e^{-02}$
TT_2-3DVoro	$2.04e^{-02}$	$7.66e^{-02}$	0	$4.10e^{-03}$	$5.80e^{-03}$	$4.48e^{-02}$	$6.48e^{-02}$	$1.68e^{-01}$
TT_3-3DVoro	$1.81e^{-02}$	$9.15e^{-02}$	0	$8.00e^{-04}$	$5.70e^{-03}$	$4.51e^{-02}$	$5.42e^{-02}$	$2.15e^{-01}$
P-3DVoro	$2.98e^{-02}$	$1.38e^{-01}$	0	$5.90e^{-03}$	$1.41e^{-02}$	$7.90e^{-02}$	$5.75e^{-02}$	$2.92e^{-01}$
M-3DVoro	$1.88e^{-02}$	$1.46e^{-01}$	0	$2.00e^{-04}$	$5.90e^{-03}$	$1.04e^{-01}$	$5.95e^{-02}$	$3.22e^{-01}$

(a) Sampling frequency: 3 s^{-1}

Method	Mean		Mode		Median		90% quantile	
	IT	SoP	IT	SoP	IT	SoP	IT	SoP
XY-T	$5.80e^{-01}$	/	1.02	/	$3.26e^{-01}$	/	1.42	/
E-3DVoro	1.77	/	$4.36e^{-02}$	/	$7.11e^{-01}$	/	1.27	/
TT_1-3DVoro	$5.42e^{-01}$	$5.40e^{-01}$	$2.28e^{-02}$	$2.10e^{-03}$	$3.43e^{-01}$	$3.02e^{-01}$	1.04	$9.66e^{-01}$
TT_2-3DVoro	$5.11e^{-01}$	$5.56e^{-01}$	$1.39e^{-01}$	$8.20e^{-03}$	$3.15e^{-01}$	$3.17e^{-01}$	1.07	1.04
TT_3-3DVoro	$6.08e^{-01}$	$5.52e^{-01}$	$3.72e^{-02}$	$7.50e^{-03}$	$3.29e^{-01}$	$3.18e^{-01}$	1.05	1.05
P-3DVoro	$5.60e^{-01}$	$5.41e^{-01}$	$8.75e^{-02}$	$1.30e^{-03}$	$3.32e^{-01}$	$3.04e^{-01}$	$9.76e^{-01}$	$9.82e^{-01}$
M-3DVoro	$5.03e^{-01}$	$5.43e^{-01}$	$3.93e^{-02}$	$6.91e^{-02}$	$3.76e^{-01}$	$3.15e^{-01}$	1.08	$9.52e^{-01}$

(b) Sampling frequency: 0.5 s^{-1}

Table A.1: Robustness to the sampling frequency of velocity indicator - $Uni_{LD-HomoPop}$

Method	Mean		Mode		Median		90% quantile	
	IT	SoP	IT	SoP	IT	SoP	IT	SoP
XY-T	$1.92e^{-02}$	/	$9.60e^{-03}$	/	$6.20e^{-03}$	/	$3.42e^{-02}$	/
E-3DVoro	$3.17e^{-02}$	/	0	/	$6.30e^{-03}$	/	$3.86e^{-02}$	/
TT_1-3DVoro	$1.57e^{-02}$	$6.18e^{-02}$	0	0	$6.10e^{-03}$	$1.87e^{-02}$	$3.23e^{-02}$	$1.30e^{-01}$
TT_2-3DVoro	$1.83e^{-02}$	$1.38e^{-01}$	0	$1.73e^{-02}$	$7.90e^{-03}$	$4.27e^{-02}$	$3.82e^{-02}$	$3.88e^{-01}$
TT_3-3DVoro	$1.85e^{-02}$	$1.88e^{-01}$	0	$1.00e^{-01}$	$8.00e^{-03}$	$6.46e^{-02}$	$4.08e^{-02}$	$4.87e^{-01}$
P-3DVoro	$2.93e^{-02}$	$2.05e^{-01}$	0	$7.96e^{-02}$	$9.00e^{-03}$	$9.82e^{-02}$	$6.49e^{-02}$	$5.29e^{-01}$
M-3DVoro	$2.14e^{-02}$	$3.16e^{-01}$	0	$5.10e^{-03}$	$8.00e^{-03}$	$1.47e^{-01}$	$4.37e^{-02}$	$8.21e^{-01}$

(a) Sampling frequency: 3 s^{-1}

Method	Mean		Mode		Median		90% quantile	
	IT	SoP	IT	SoP	IT	SoP	IT	SoP
XY-T	$5.73e^{-01}$	/	1.15	/	$3.51e^{-01}$	/	1.58	/
E-3DVoro	1.01	/	$8.57e^{-01}$	/	$3.85e^{-01}$	/	1.67	/
TT_1-3DVoro	$5.82e^{-01}$	$5.80e^{-01}$	$8.69e^{-01}$	$5.85e^{-02}$	$4.51e^{-01}$	$3.13e^{-01}$	1.40	1.28
TT_2-3DVoro	$5.76e^{-01}$	$5.67e^{-01}$	$9.40e^{-01}$	$1.02e^{-01}$	$3.75e^{-01}$	$2.64e^{-01}$	1.54	1.16
TT_3-3DVoro	$5.79e^{-01}$	$5.94e^{-01}$	$8.50e^{-01}$	$5.73e^{-02}$	$3.70e^{-01}$	$2.77e^{-01}$	1.46	1.29
P-3DVoro	$5.66e^{-01}$	$5.62e^{-01}$	$8.92e^{-01}$	$4.61e^{-02}$	$3.83e^{-01}$	$2.95e^{-01}$	1.38	1.26
M-3DVoro	$6.27e^{-01}$	$7.11e^{-01}$	$9.13e^{-01}$	$1.43e^{-02}$	$5.05e^{-01}$	$2.86e^{-01}$	1.55	1.49

(b) Sampling frequency: 0.5 s^{-1}

Table A.2: Robustness to the sampling frequency of velocity indicator - $Uni_{HD-HeteroPop}$

Method	Mean		Mode		Median		90% quantile	
	IT	SoP	IT	SoP	IT	SoP	IT	SoP
XY-T	$1.93e^{-02}$	/	0	/	$1.77e^{-02}$	/	$7.73e^{-02}$	/
E-3DVoro	$1.65e^{-02}$	/	0	/	$5.60e^{-03}$	/	$3.75e^{-02}$	/
TT_1-3DVoro	$3.00e^{-04}$	$7.60e^{-03}$	0	0	0	$2.60e^{-03}$	$8.00e^{-04}$	$1.74e^{-02}$
TT_2-3DVoro	$1.40e^{-03}$	$4.16e^{-02}$	0	0	0	$3.17e^{-02}$	$3.60e^{-03}$	$8.99e^{-02}$
TT_3-3DVoro	$1.30e^{-03}$	$4.65e^{-02}$	0	$4.32e^{-02}$	0	$3.48e^{-02}$	$3.90e^{-03}$	$1.14e^{-01}$
P-3DVoro	$2.70e^{-03}$	$4.69e^{-02}$	0	$1.41e^{-02}$	$8.00e^{-04}$	$2.27e^{-02}$	$5.50e^{-03}$	$1.29e^{-01}$
M-3DVoro	$1.20e^{-03}$	$5.09e^{-02}$	0	$4.75e^{-02}$	0	$3.54e^{-02}$	$2.50e^{-03}$	$1.23e^{-01}$

(a) Sampling frequency: 3 s^{-1}

Method	Mean		Mode		Median		90% quantile	
	IT	SoP	IT	SoP	IT	SoP	IT	SoP
XY-T	$2.55e^{-01}$	/	$1.45e^{-01}$	/	$2.45e^{-01}$	/	$5.06e^{-01}$	/
E-3DVoro	$4.17e^{-01}$	/	$6.50e^{-02}$	/	$1.27e^{-01}$	/	$3.83e^{-01}$	/
3DVoro-δ_{TT_1}	$1.74e^{-01}$	$1.50e^{-01}$	$1.79e^{-01}$	$8.00e^{-04}$	$1.13e^{-01}$	$8.77e^{-02}$	$3.21e^{-01}$	$2.98e^{-01}$
TT_1-3DVoro	$2.07e^{-01}$	$1.53e^{-01}$	$1.92e^{-01}$	$1.00e^{-04}$	$1.39e^{-01}$	$8.52e^{-02}$	$3.71e^{-01}$	$3.29e^{-01}$
TT_2-3DVoro	$2.33e^{-01}$	$1.52e^{-01}$	$2.05e^{-01}$	$3.00e^{-04}$	$1.48e^{-01}$	$8.46e^{-02}$	$3.63e^{-01}$	$3.27e^{-01}$
TT_2-3DVoro	$2.17e^{-01}$	$1.43e^{-01}$	$1.53e^{-01}$	$1.40e^{-03}$	$1.34e^{-01}$	$8.49e^{-02}$	$3.01e^{-01}$	$2.98e^{-01}$
M-3DVoro	$1.75e^{-01}$	$1.48e^{-01}$	$1.83e^{-01}$	$1.00e^{-04}$	$1.36e^{-01}$	$9.11e^{-02}$	$3.43e^{-01}$	$3.22e^{-01}$

(b) Sampling frequency: 0.5 s^{-1}

Table A.3: Robustness to the sampling frequency of flow indicator - $Uni_{LD-HomoPop}$

Method	Mean		Mode		Median		90% quantile	
	IT	SoP	IT	SoP	IT	SoP	IT	SoP
XY-T	$2.75e^{-02}$	/	$2.30e^{-03}$	/	$1.75e^{-02}$	/	$7.21e^{-02}$	/
E-3DVoro	$1.09e^{-02}$	/	0	/	$8.70e^{-04}$	/	$2.83e^{-02}$	/
TT_1-3DVoro	$7.80e^{-03}$	$6.06e^{-02}$	0	0	$7.00e^{-04}$	$1.21e^{-02}$	$2.22e^{-02}$	$1.58e^{-01}$
TT_2-3DVoro	$1.05e^{-02}$	$1.45e^{-01}$	0	0	$1.10e^{-03}$	$6.08e^{-02}$	$2.78e^{-02}$	$3.11e^{-01}$
TT_3-3DVoro	$1.06e^{-02}$	$2.03e^{-01}$	0	0	$1.00e^{-03}$	$8.27e^{-02}$	$2.19e^{-02}$	$4.64e^{-01}$
P-3DVoro	$1.62e^{-02}$	$1.95e^{-01}$	0	$4.86e^{-02}$	$1.80e^{-03}$	$8.54e^{-02}$	$3.70e^{-02}$	$4.90e^{-01}$
M-3DVoro	$1.29e^{-02}$	$3.06e^{-01}$	0	0	$1.60e^{-03}$	$1.48e^{-01}$	$2.92e^{-02}$	$8.95e^{-01}$

(a) Sampling frequency: 3 s^{-1}

Method	Mean		Mode		Median		90% quantile	
	IT	SoP	IT	SoP	IT	SoP	IT	SoP
XY-T	$5.18e^{-01}$	/	$3.50e^{-01}$	/	$4.48e^{-01}$	/	1.09	/
E-3DVoro	$6.54e^{-01}$	/	$3.69e^{-01}$	/	$2.03e^{-01}$	/	1.54	/
TT_1-3DVoro	$4.99e^{-01}$	$4.02e^{-01}$	$1.06e^{-01}$	$6.49e^{-02}$	$3.24e^{-01}$	$1.81e^{-01}$	1.35	$9.43e^{-01}$
TT_2-3DVoro	$5.66e^{-01}$	$4.16e^{-01}$	$1.47e^{-01}$	$5.55e^{-02}$	$2.73e^{-01}$	$1.73e^{-01}$	1.57	1.21
TT_3-3DVoro	$5.91e^{-01}$	$4.45e^{-01}$	$1.53e^{-01}$	$1.57e^{-01}$	$2.94e^{-01}$	$1.71e^{-01}$	1.68	1.31
P-3DVoro	$4.81e^{-01}$	$4.28e^{-01}$	$5.53e^{-02}$	$3.98e^{-02}$	$2.22e^{-01}$	$1.89e^{-01}$	1.34	1.12
M-3DVoro	$6.41e^{-01}$	$4.47e^{-01}$	$9.07e^{-02}$	$4.55e^{-02}$	$3.97e^{-01}$	$1.73e^{-01}$	1.66	1.24

(b) Sampling frequency: 0.5 s^{-1}

Table A.4: Robustness to the sampling frequency of flow indicator - $Uni_{HD-HeteroPop}$

B

PedProb-vk: Alternative specifications

The integral (4.17) can be simplified in some circumstances and can be solved in a closed form. Consider the following probability density function of the speed that is defined on the interval $[v_e(k) - \omega(k), v_e(k) + \omega(k)]$

$$f_V(\xi|v_e(k), \omega(k)) = \begin{cases} \frac{\xi - v_e(k) + \omega(k)}{\omega(k)^2}, & v_e(k) - \omega(k) \leq \xi \leq v_e(k) \\ \frac{v_e(k) + \omega(k) - \xi}{\omega(k)^2}, & v_e(k) \leq \xi \leq v_e(k) + \omega(k) \\ 0, & \xi < v_e(k) - \omega(k) \text{ or } \xi > v_e(k) + \omega(k). \end{cases} \tag{B.1}$$

If the symmetric triangular distribution (4.22) is the assumed mixing distribution, the resulting probabilistic speed-density relationship is given as

$$f_{\text{PedProb-vk}}(\xi|k; \bar{v}_e(k), \omega(k), \sigma(k)) =$$

$$\begin{cases} \frac{\omega(k) + \xi - \bar{v}_e(k)}{\omega^2(k)}, & \xi \leq \bar{v}_e(k) - \sigma(k) \\ \frac{\omega(k) - \frac{\sigma(k)}{3}}{\omega(k)^2} - \frac{(\xi - \bar{v}_e(k))^2}{\sigma(k)\omega(k)^2} - \frac{(\xi - \bar{v}_e(k))^3}{3\sigma(k)^2\omega(k)^2}, & \bar{v}_e(k) - \sigma(k) < \xi \leq \bar{v}_e(k) \\ \frac{\omega(k) - \frac{\sigma(k)}{3}}{\omega(k)^2} - \frac{(\xi - \bar{v}_e(k))^2}{\sigma(k)\omega(k)^2} + \frac{(\xi - \bar{v}_e(k))^3}{3\sigma(k)^2\omega(k)^2}, & \bar{v}_e(k) < \xi \leq \bar{v}_e(k) + \sigma(k) \\ \frac{\omega(k) - \xi + \bar{v}_e(k)}{\omega^2(k)}, & \xi > \bar{v}_e(k) + \sigma(k). \end{cases} \tag{B.2}$$

Similarly, if we assume the uniform mixing distribution defined on the interval $[\bar{v}_e(k) -$

$\alpha(k), \bar{v}_e(k) + \alpha(k)]$

$$f_{v_e(k)}(\zeta; \bar{v}_e(k), \alpha(k)) = \begin{cases} \frac{1}{2\alpha(k)}, & \bar{v}_e(k) - \alpha(k) \leq \zeta \leq \bar{v}_e(k) + \alpha(k) \\ 0, & \zeta < \bar{v}_e(k) - \alpha(k) \text{ or } \zeta > \bar{v}_e(k) + \alpha(k), \end{cases} \tag{B.3}$$

the resulting probabilistic speed-density relationship is given as

$$f_{\text{PedProb-vk}}(\xi|k; \bar{v}_e(k), \omega(k), \alpha(k)) =$$

$$\begin{cases} \frac{\xi + \omega(k) - \bar{v}_e(k)}{\omega(k)^2}, & \xi \leq \bar{v}_e(k) - \alpha(k) \\ -\frac{\alpha(k) - 2\omega(k)}{2\omega(k)^2} - \frac{(\xi - \bar{v}_e(k))^2}{2\alpha(k)\omega(k)^2}, & \bar{v}_e(k) - \alpha(k) < \xi \leq \bar{v}_e(k) + \alpha(k) \\ \frac{-\xi + \omega(k) + \bar{v}_e(k)}{\omega(k)^2}, & \xi > \bar{v}_e(k) + \alpha(k). \end{cases} \tag{B.4}$$

C

MC-vk: Additional analysis

The speed observations (Chapter 5, Section 5.2) are aggregated based on Levels of Service (LoS) standard for pedestrian facilities proposed by Fruin (1971). The levels are labeled from A to F, as shown in Table C.1. We consider the levels A-E, given that the largest

LoS	Density range [ped/m^2]
A	k $\leq$ 0.31
B	k $\in$ (0.31 - 0.43]
C	k $\in$ (0.43 - 0.71]
D	k $\in$ (0.71 - 1.11]
E	k $\in$ (1.11 - 2.17]
F	k $>$ 2.17

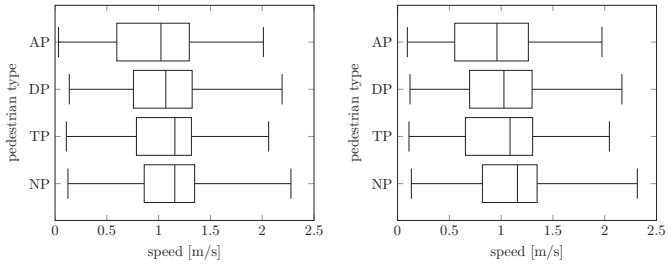
Table C.1: LoS (Fruin, 1971)

part of the observations falls below the LoS F (99%) of the observations are below 1.4 pedestrians per square meter.

The box-plots of the distributions are shown in Figure C.1, Figure C.2, Figure C.3 and Figure C.4 for different pedestrian types, respectively OD distances, periods and time to departure.

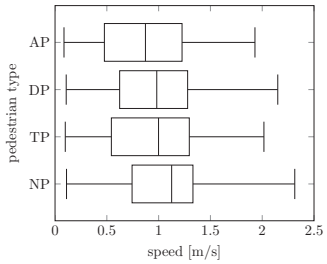
For the purpose of the analysis we discretize the range of OD distances and time to departure into segments. We consider distances shorter than 10 meters (short distances - SD), distances between 10 meters and 30 meters (medium distances - MD), and distances longer than 30 meters (long distances - LD). Time to departure interval is segmented

into time intervals shorter than 5 minutes (short intervals - SI), and time intervals longer than 5 minutes (long intervals - LI). The bottom and top of the boxes in Figure C.1 - Figure C.4 are the first and third quartiles, and the line inside the box is the median. The ends of the whiskers represent the 2^{nd} and the 98^{th} percentiles. The analysis indicates dissimilar trends, in particular with respect to speed distributions for different pedestrian types (Figure C.1) and different ranges of OD distances (Figure C.2).

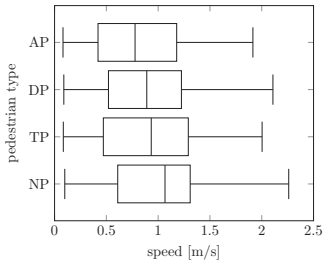


(a) LoS A

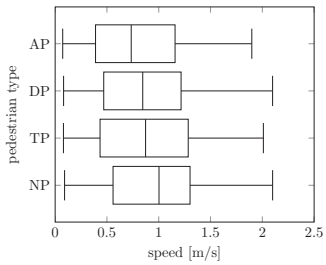
(b) LoS B



(c) LoS C



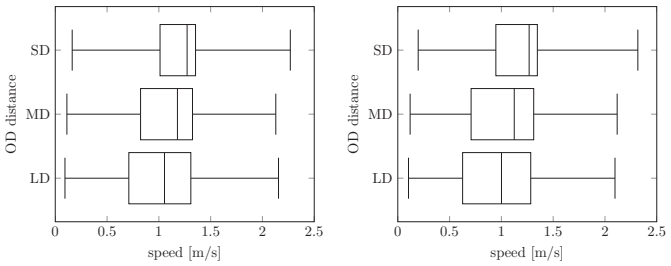
(d) LoS D



(e) LoS E

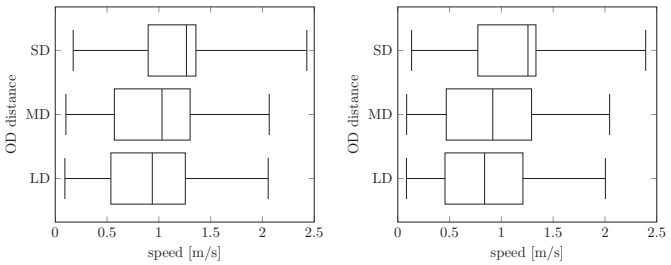
Figure C.1: Speed distribution for different pedestrian types and different LoS

For the quantitative analysis, we use Kolmogorov-Smirnov test (Massey, 1951). For each LoS and each factor we test the hypothesis that speed distributions for different factor values represent the same population. The results are shown in Table C.2. The test rejects the hypothesis in all cases, at the 5% significance level.



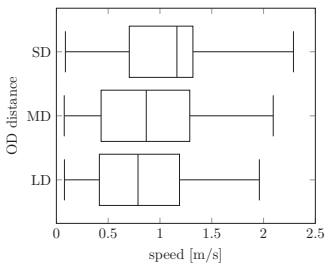
(a) LoS A

(b) LoS B



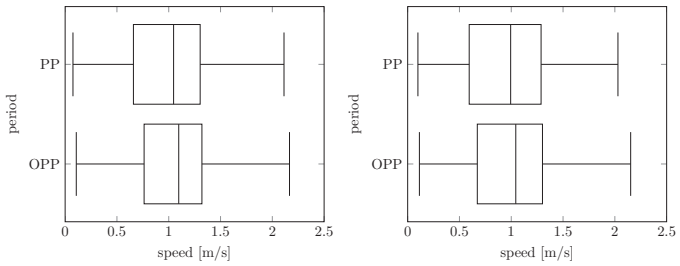
(c) LoS C

(d) LoS D



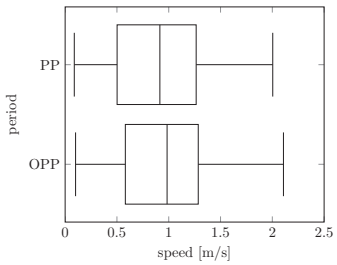
(e) LoS E

Figure C.2: Speed distribution for different OD distances and different LoS

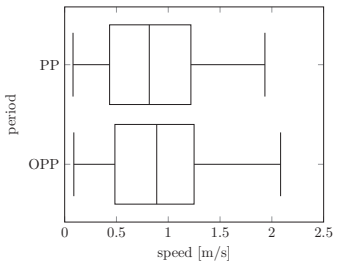


(a) LoS A

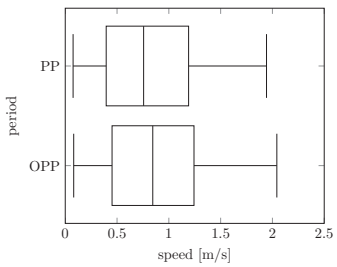
(b) LoS B



(c) LoS C

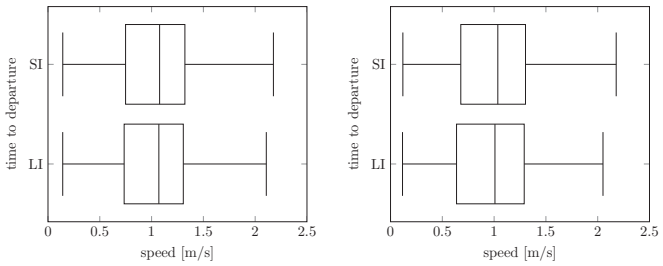


(d) LoS D



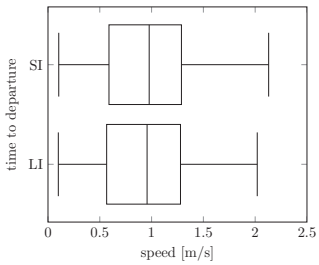
(e) LoS E

Figure C.3: Speed distribution for peak and off-peak periods and different LoS

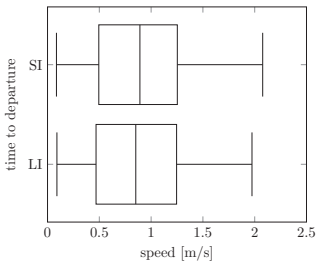


(a) LoS A

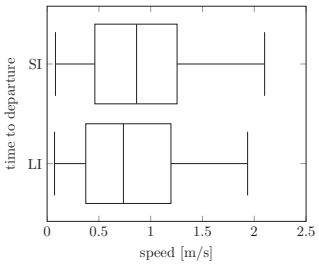
(b) LoS B



(c) LoS C



(d) LoS D



(e) LoS E

Figure C.4: Speed distribution for different time to departure and different LoS

LoS	p-value					
	AP-DP	AP-TP	AP-NP	DP-TP	DP-NP	TP-NP
LoS A	0	0	0	$2.38e^{-213}$	0	$5.16e^{-65}$
LoS B	$3.21e^{-113}$	$9.02e^{-94}$	$3.73e^{-256}$	$1.04e^{-38}$	$6.75e^{-133}$	$1.99e^{-37}$
LoS C	$4.61e^{-188}$	$1.88e^{-106}$	0	$7.71e^{-24}$	$3.97e^{-147}$	$5.80e^{-96}$
LoS D	$3.16e^{-82}$	$1.28e^{-93}$	$4.22e^{-192}$	$1.53e^{-37}$	$1.40e^{-84}$	$2.29e^{-34}$
LoS E	$1.70e^{-24}$	$8.80e^{-29}$	$1.13e^{-51}$	$3.63e^{-09}$	$7.73e^{-19}$	$3.00e^{-11}$

(a) Pedestrian types

LoS	p-value		
	SD-MD	SD-LD	MD-LD
LoS A	$2.55e^{-187}$	0	0
LoS B	$3.25e^{-51}$	$2.14e^{-175}$	$4.08e^{-156}$
LoS C	$9.71e^{-89}$	$6.05e^{-215}$	$8.78e^{-131}$
LoS D	$1.55e^{-50}$	$1.81e^{-106}$	$1.22e^{-70}$
LoS E	$2.57e^{-14}$	$9.45e^{-30}$	$4.50e^{-24}$

(b) OD distances

LoS	p-value
LoS A	$1.89e^{-206}$
LoS B	$9.28e^{-48}$
LoS C	$1.24e^{-79}$
LoS D	$6.76e^{-36}$
LoS E	$1.24e^{-18}$

(c) Peak/off-peak periods

LoS	p-value
LoS A	$2.54e^{-12}$
LoS B	$1.11e^{-04}$
LoS C	$2.83e^{-02}$
LoS D	$4.35e^{-02}$
LoS E	$2.66e^{-06}$

(d) Time to departure

Table C.2: p-value of Kolmogorov-Smirnov statistic

Bibliography

Alahi, A., Bagnato, L., Chanel, D. and Alahi, A. (2013). Technical report for SBB network of sensors, *Technical report*, VisioSafe SA, Switzerland.

Alahi, A., Bierlaire, M. and Vandergheynst, P. (2014). Robust real-time pedestrians detection in urban environments with a network of low resolution cameras, *Transportation Research Part C: Emerging Technologies* **39**: 113–128.

Alahi, A., Jacques, L., Boursier, Y. and Vandergheynst, P. (2011). Sparsity driven people localization with a heterogeneous network of cameras, *Journal of Mathematical Imaging and Vision* **41**(1-2): 39–58.

Algadhi, S. A. and Mahmassani, H. S. (1990). Modelling crowd behavior and movement: application to makkah pilgrimage, *Transportation and traffic theory* **1990**: 59–78.

Antonini, G., Martinez, S. V., Bierlaire, M. and Thiran, J. P. (2006). Behavioral priors for detection and tracking of pedestrians in video sequences, *International Journal of Computer Vision* **69**(2): 159–180.

Bachu, P., Dudala, T. and Kothuri, S. (2001). Prompted recall in global positioning system survey: Proof-of-concept study, *Transportation Research Record: Journal of the Transportation Research Board* (1768): 106–113.

Ball, R., Ghorpade, A., Nawarathne, K., Baltazar, R., Pereira, F. C., Zegras, C. and Ben-Akiva, M. (2014). Battery patterns and forecasting in a large-scale smartphone-based travel survey, *10th International Conference on Transport SurveyMethods, Fairmont Resort, Jamison Valley, BlueMountains National Park, Australia*.

Bauer, D., Brandle, N., Seer, S., Ray, M. and Kitazawa, K. (2009). Measurement of pedestrian movements: A comparative study on various existing systems, *in* H. Timmermans (ed.), *Pedestrian Behavior: Models, Data Collection and Applications*, Emerald Group Publishing Limited, pp. 325–344.

Beck, C. and Roepstorff, G. (1987). Effects of phase space discretization on the long-time behavior of dynamical systems, *Physica D: Nonlinear Phenomena* **25**(1-3): 173–180.

Bierlaire, M. and Robin, T. (2009). Pedestrians choices, *in* H. Timmermans (ed.), *Pedestrian Behavior: Models, Data Collection and Applications*, Emerald Group Publishing Limited, pp. 1–26. ISBN:978-1-84855-750-5.

Blue, V. and Adler, J. (1998). Emergent fundamental pedestrian flows from cellular automata microsimulation, *Transportation Research Record: Journal of the Transportation Research Board* (1644): 29–36.

Blue, V. J. and Adler, J. L. (2001). Cellular automata microsimulation for modeling bi-directional pedestrian walkways, *Transportation Research Part B: Methodological* **35**(3): 293–312.

Bohte, W. and Maat, K. (2009). Deriving and validating trip purposes and travel modes for multi-day gps-based travel surveys: A large-scale application in the netherlands, *Transportation Research Part C: Emerging Technologies* **17**(3): 285–297.

Borgers, A. and Timmermans, H. (1986). A model of pedestrian route choice and demand for retail facilities within inner-city shopping areas, *Geographical analysis* **18**(2): 115–128.

Bovy, P. H. and Stern, E. (2012). *Route Choice: Wayfinding in Transport Networks: Wayfinding in Transport Networks*, Vol. 9, Springer Science & Business Media.

Bowerman, B., Braverman, J., Taylor, J., Todosow, H. and Von Wimmersperg, U. (2000). The vision of a smart city, *2nd International Life Extension Technology Workshop, Paris*, Vol. 28.

Bowman, B. L. and Vecellio, R. L. (1994). Pedestrian walking speeds and conflicts at urban median locations, *Transportation research record* (1438): 67–73.

Campanella, M., Hoogendoorn, S. and Daamen, W. (2014). The nomad model: theory, developments and applications, *Transportation Research Procedia* **2**: 462–467.

Cascetta, E. and Russo, F. (1997). Calibrating aggregate travel demand models with traffic counts: Estimators and statistical performance, *Transportation* **24**(3): 271–293.

Cheah, J. Y. and Smith, J. M. (1994). Generalized M/G/c/c state dependent queueing models and pedestrian traffic flows, *Queueing Systems* **15**(1-4): 365–386.

Chen, M., Mao, S. and Liu, Y. (2014). Big data: a survey, *Mobile Networks and Applications* **19**(2): 171–209.

Chen, T.-H., Chen, T.-Y. and Chen, Z.-X. (2006). An intelligent people-flow counting method for passing through a gate, *2006 IEEE Conference on Robotics, Automation and Mechatronics*, IEEE, pp. 1–6.

Cheung, C. and Lam, W. H. (1998). Pedestrian route choices between escalator and stairway in mtr stations, *Journal of transportation engineering* **124**(3): 277–285.

Colombo, R. M. (2003). Hyperbolic phase transitions in traffic flow, *SIAM Journal on Applied Mathematics* **63**(2): 708–721.

Çöltekin, A., De Sabbata, S., Willi, C., Vontobel, I., Pfister, S., Kuhn, M. and Lacayo, M. (2011). Modifiable temporal unit problem, *ISPRS/ICA workshop Persistent problems in geographic visualization(ICC2011), Paris, France*, Vol. 2.

Cottrill, C., Pereira, F., Zhao, F., Dias, I., Lim, H., Ben-Akiva, M. and Zegras, P. (2013). Future mobility survey: Experience in developing a smartphone-based travel survey in singapore, *Transportation Research Record: Journal of the Transportation Research Board* (2354): 59–67.

Daamen, W. (2004). *Modelling passenger flows in public transport facilities*, PhD thesis, Delft University of Technology, Delft.

Daamen, W. and Hoogendoorn, S. P. (2003). Controlled experiments to derive walking behaviour, *European Journal of Transport and Infrastructure Research* **3**(1): 39–59.

Daamen, W., Hoogendoorn, S. P. and Bovy, P. H. (2005). First-order pedestrian traffic flow theory, *Transportation Research Record: Journal of the Transportation Research Board* **1934**(1): 43–52.

Daganzo, C. and Newell, G. (1995). Methods of analysis for transportation operations, Institute of Transportation Studies, University of California at Berkeley.

Danalet, A. (2015). *Activity choice modeling for pedestrian facilities*, PhD thesis, École Polytechnique Fédérale de Lausanne.

Danalet, A., Farooq, B. and Bierlaire, M. (2014). A bayesian approach to detect pedestrian destination-sequences from wifi signatures, *Transportation Research Part C: Emerging Technologies* **44**: 146–170.

Dashti, S., Palen, L., Heris, M. P., Anderson, K. M., Anderson, S. and Anderson, S. (2014). Supporting disaster reconnaissance with social media data: a design-oriented case study of the 2013 colorado floods, *Proceedings of the 11th International ISCRAM Conference*, pp. 18–21.

Davidson, R. and MacKinnon, J. G. (2004). *Econometric theory and methods*, Vol. 5, Oxford University Press New York.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap methods and their application*, Vol. 1, Cambridge university press.

DiNenno, P. J. (2002). *SFPE handbook of fire protection engineering*, National Fire Protection Association, Quincy, Massachusetts.

Diogenes, M., Greene-Roesel, R., Arnold, L. and Ragland, D. (2007). Pedestrian counting methods at intersections: a comparative study, *Transportation Research Record: Journal of the Transportation Research Board* (2002): 26–30.

Drake, J., Schofer, J. and May, A. (1967). A statistical analysis of speed-density hypotheses, *Highway Research Record* **154**: 53–87.

Duives, D. C., Daamen, W. and Hoogendoorn, S. (2014a). Trajectory analysis of pedestrian crowd movements at a dutch music festival, *Pedestrian and Evacuation Dynamics 2012*, Springer, pp. 151–166.

Duives, D. C., Daamen, W. and Hoogendoorn, S. P. (2013). State-of-the-art crowd motion simulation models, *Transportation Research Part C: Emerging Technologies* **37**: 193–209.

Duives, D. C., Daamen, W. and Hoogendoorn, S. P. (2015). Quantification of the level of crowdedness for pedestrian movements, *Physica A: Statistical Mechanics and its Applications* **427**: 162–180.

Duives, D., Daamen, W. and Hoogendoorn, S. (2014b). Anticipation behavior upstream of a bottleneck, *Transportation Research Procedia* **2**: 43–50.

Edie, L. C. (1963). *Discussion of traffic stream measurements and definitions*, Port of New York Authority, New York, USA.

Flötteröd, G. and Lämmel, G. (2015). Bidirectional pedestrian fundamental diagram, *Transportation research part B: methodological* **71**: 194–212.

Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*, Springer Series in Statistics, Springer Science & Business Media, LLC.

Fruin, J. J. (1971). Designing for pedestrians: A level-of-service concept, number 355, Highway Research Board, Washington, DC, pp. 1–15.

Fuchida, T., Kashima, M., Nakamura, H., Mori, K. and Murashima, S. (2005). Constructing three-dimensional discrete voronoi diagrams by the incremental method and application to self-organizing maps, *Systems and Computers in Japan* **36**(5): 55–67.

Gantz, J. and Reinsel, D. (2011). Extracting value from chaos, *IDC iview* **1142**: 1–12.

Ge, W., Collins, R. T. and Ruback, R. B. (2012). Vision-based analysis of small groups in pedestrian crowds, *IEEE transactions on pattern analysis and machine intelligence* **34**(5): 1003–1016.

Geroliminis, N. and Daganzo, C. F. (2008). Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings, *Transportation Research Part B: Methodological* **42**(9): 759–770.

Gourieroux, C., Monfort, A. and Trognon, A. (1984). Pseudo maximum likelihood methods: Theory, *Econometrica: Journal of the Econometric Society* pp. 681–700.

Greene-Roesel, R., Diogenes, M. C., Ragland, D. R. and Lindau, L. A. (2008). Effectiveness of a commercially available automated pedestrian counting device in urban environments: Comparison with manual counts, *Transportation Research Board 87th Annual Meeting*, pp. 1–16.

Greenshields, B., Bibbins, J., Channing, W. and Miller, H. (1935). A study of traffic capacity, *Proceedings of the Highway Research Board*, Vol. 14, Highway Research Board, Washington, DC.

Hall, P., Horowitz, J. L. and Jing, B.-Y. (1995). On blocking rules for the bootstrap with dependent data, *Biometrika* **82**(3): 561–574.

Hänseler, F., Bierlaire, M., Farooq, B. and Mühlematter, T. (2014). A macroscopic loading model for time-varying pedestrian flows in public walking areas, *Transportation Research Part B: Methodological* **69**: 60 – 80.

Hänseler, F., Lam, W., Bierlaire, M., Lederrey, G. and Nikolić, M. (2017). A dynamic network loading model for anisotropic and congested pedestrian flows, *Transportation Research Part B: Methodological* **95**: 149–168.

Hänseler, F. S. (2016). *Modeling and estimation of pedestrian flows in train stations*, PhD thesis, École Polytechnique Fédérale de Lausanne.

Haq, S. and Luo, Y. (2012). Space syntax in healthcare facilities research: A review, *HERD: Health Environments Research & Design Journal* **5**(4): 98–117.

Helbing, D., Buzna, L., Johansson, A. and Werner, T. (2005). Self-organized pedestrian crowd dynamics: Experiments, simulations, and design solutions, *Transportation science* **39**(1): 1–24.

Helbing, D. and Johansson, A. (2010). Pedestrian, crowd and evacuation dynamics, *Encyclopedia of Complexity and System Science* **16**: 6476–6495.

Helbing, D., Johansson, A. and Al-Abideen, H. Z. (2007). Dynamics of crowd disasters: An empirical study, *Physical review E - Statistical, Nonlinear, and Soft Matter Physics* **75**(4): 1–7.

Helbing, D. and Molnar, P. (1995). Social force model for pedestrian dynamics, *Physical review E* **51**(5): 4282–4286.

Hillier, B. and Tzortzi, K. (2006). Space syntax: the language of museum space, *A companion to museum studies* pp. 282–301.

Hoff III, K. E., Keyser, J., Lin, M., Manocha, D. and Culver, T. (1999). Fast computation of generalized voronoi diagrams using graphics hardware, *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, ACM Press/Addison-Wesley Publishing Co., pp. 277–286.

Hoogendoorn, S. and Bovy, P. (2000). Gas-kinetic modeling and simulation of pedestrian flows, *Transportation Research Record: Journal of the Transportation Research Board* (1710): 28–36.

Hoogendoorn, S. and Daamen, W. (2004). Self-organization in walker experiments, *Traffic and Granular Flow*, Vol. 3, pp. 121–132.

Hoogendoorn, S. and HL Bovy, P. (2003). Simulation of pedestrian flows by optimal control and differential games, *Optimal Control Applications and Methods* **24**(3): 153–172.

Hoogendoorn, S. P. (2001). Normative pedestrian flow behavior theory and applications, *Technical report*, Delft University of Technology, Faculty Civil Engineering and Geosciences.

Hoogendoorn, S. P. and Bovy, P. H. (2001). State-of-the-art of vehicular traffic flow modelling, *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering* **215**(4): 283–303.

Hoogendoorn, S. P. and Bovy, P. H. (2004). Pedestrian route-choice and activity scheduling theory and models, *Transportation Research Part B: Methodological* **38**(2): 169–190.

Hoogendoorn, S. P. and Daamen, W. (2005). Pedestrian behavior at bottlenecks, *Transportation science* **39**(2): 147–159.

Hoogendoorn, S. P. and Daamen, W. (2007). Microscopic calibration and validation of pedestrian models: Cross-comparison of models using experimental data, *Traffic and Granular Flow05*, Springer, pp. 329–340.

Hoogendoorn, S. P., Daamen, W. and Bovy, P. H. (2003). Extracting microscopic pedestrian characteristics from video data, *Transportation Research Board Annual Meeting*, pp. 1–15.

Hoogendoorn, S. P., van Wageningen-Kessels, F. L., Daamen, W. and Duives, D. C. (2014). Continuum modelling of pedestrian flows: From microscopic principles to self-organised macroscopic phenomena, *Physica A: Statistical Mechanics and its Applications* **416**: 684–694.

Hughes, R. L. (2002). A continuum theory for the flow of pedestrians, *Transportation Research Part B: Methodological* **36**(6): 507–535.

Hui, S. K., Bradlow, E. T. and Fader, P. S. (2009). Testing behavioral hypotheses using an integrated model of grocery store shopping path and purchase behavior, *Journal of consumer research* **36**(3): 478–493.

Jabari, S. E., Zheng, J. and Liu, H. X. (2014). A probabilistic stationary speed–density relation based on Newell's simplified car-following model, *Transportation Research Part B: Methodological* **68**: 205–223.

Johansson, A., Helbing, D. and Shukla, P. K. (2007). Specification of the social force pedestrian model by evolutionary adjustment to video tracking data, *Advances in complex systems* **10**(2): 271–288.

Kalakou, S. and Moura, F. (2014). Effects of terminal planning on passenger choices, *14th Swiss Transport Research Conference (STRC), Monte Verità, Ascona, Switzerland*.

Kanda, T., Shiomi, M., Perrin, L., Nomura, T., Ishiguro, H. and Hagita, N. (2007). Analysis of people trajectories with ubiquitous sensors in a science museum, *Proceedings 2007 IEEE International Conference on Robotics and Automation*, IEEE, pp. 4846–4853.

Kazagli, E., Chen, J. and Bierlaire, M. (2014). Individual mobility analysis using smartphone data, *in* S. Rasouli and H. Timmermans (eds), *Mobile Technologies for Activity-Travel Data Collection and Analysis*, pp. 187–208.

Khoshyaran, M. M. and Lebacque, J.-P. (2009). A stochastic macroscopic traffic model devoid of diffusion, *Traffic and Granular Flow07*, Springer, pp. 139–150.

Kim, K. M., Hong, S.-P., Ko, S.-J. and Kim, D. (2015). Does crowding affect the path choice of metro passengers?, *Transportation Research Part A: Policy and Practice* **77**: 292–304.

Kim, T. and Zhang, H. (2008). A stochastic wave propagation model, *Transportation Research Part B: Methodological* **42**(7): 619–634.

Lam, W. H. and Cheung, C.-y. (2000). Pedestrian speed/flow relationships for walking facilities in hong kong, *Journal of transportation engineering* **126**(4): 343–349.

Lam, W. H., Lee, J. Y. and Cheung, C. (2002). A study of the bi-directional pedestrian flow characteristics at hong kong signalized crosswalk facilities, *Transportation* **29**(2): 169–192.

Lam, W. H., Morrall, J. F. and Ho, H. (1995). Pedestrian flow characteristics in Hong Kong, *Transportation Research Record* (1487): 56–62.

Lämmer, S. and Helbing, D. (2008). Self-control of traffic lights and vehicle flows in urban road networks, *Journal of Statistical Mechanics: Theory and Experiment* **2008**(04): 1–36.

Lebacque, J.-P. and Khoshyaran, M. M. (2013). A variational formulation for higher order macroscopic traffic flow models of the gsom family, *Procedia-Social and Behavioral Sciences* **80**: 370–394.

Lee, J.-G., Han, J. and Whang, K.-Y. (2007). Trajectory clustering: a partition-and-group framework, *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, ACM, pp. 593–604.

Lesani, A. and Miranda-Moreno, L. F. (2016). Development and testing of a real-time wifi-bluetooth system for pedestrian network monitoring and data extrapolation, *Transportation Research Board 95th Annual Meeting*, number 16-5665.

Lin, Y.-H. and Chen, C.-F. (2013). Passengers' shopping motivations and commercial activities at airports–the moderating effects of time pressure and impulse buying tendency, *Tourism Management* **36**: 426–434.

Liu, H., Darabi, H., Banerjee, P. and Liu, J. (2007). Survey of wireless indoor positioning techniques and systems, *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* **37**(6): 1067–1080.

Løvås, G. G. (1994). Modeling and simulation of pedestrian traffic flow, *Transportation Research Part B: Methodological* **28**(6): 429–443.

Massey, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit, *Journal of the American statistical Association* **46**(253): 68–78.

McPhail, C. and Wohlstein, R. T. (1982). Using film to analyze pedestrian behavior, *Sociological Methods & Research* **10**(3): 347–375.

Mehner, W., Boltes, M. and Seyfried, A. (2015). Methodology for generating individualized trajectories from experiments, *Traffic and Granular Flow'15,* pp. 3–10.

Naini, F. M., Dousse, O., Thiran, P. and Vetterli, M. (2011). Population size estimation using a few individuals as agents, *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, IEEE, pp. 2499–2503.

Navin, F. and Wheeler, R. (1969). Pedestrian flow characteristics, *Traffic Engineering, Inst Traffic Engr* **39**: 30–36.

Newell, G. F. (1961). Nonlinear effects in the dynamics of car following, *Operations Research* **9**(2): 209–229.

Nikolić, M. and Bierlaire, M. (2014). Pedestrian-oriented flow characterization, *Transportation Research Procedia* **2**: 359–366.

Nikolić, M., Bierlaire, M. and Farooq, B. (2014). Probabilistic speed-density relationship for pedestrians based on data driven space and time representation, *Proceedings of the Swiss Transportation Research Conference*, Ascona, Switzerland.

Ohmori, N., Nakazato, M. and Harata, N. (2005). Gps mobile phone-based activity diary survey, *Proceedings of the Eastern Asia Society for Transportation Studies*, Vol. 5, pp. 1104–1115.

Okabe, A., Boots, B., Sugihara, K. and Chiu, S. N. (2000). *Spatial tessellations: concepts and applications of Voronoi diagrams*, Second edn, Wiley, New York.

Older, S. (1968). Movement of pedestrians on footways in shopping streets, *Traffic engineering and control* **10**: 160–163.

Openshaw, S. (1983). The modifiable areal unit problem, *Concepts and Techniques in Modern Geography*, number 38, Geobooks, Norwich, England.

Park, S. W., Linsen, L., Kreylos, O., Owens, J. D. and Hamann, B. (2006). Discrete sibson interpolation, *IEEE Transactions on Visualization and Computer Graphics* **12**(2): 243–253.

Petrolo, R., Loscrí, V. and Mitton, N. (2014). Towards a smart city based on cloud of things, *Proceedings of the 2014 ACM international workshop on Wireless and mobile technologies for smart cities*, ACM, pp. 61–66.

Plaue, M., Bärwolff, G. and Schwandt, H. (2014). On measuring pedestrian density and flow fields in dense as well as sparse crowds, *Pedestrian and Evacuation Dynamics 2012*, Springer, pp. 411–424.

Rastogi, R., Ilango, T. and Chandra, S. (2013). Pedestrian flow characteristics for different pedestrian facilities and situations, *European Transport* **53**: 1–21.

Rigaux, P., Scholl, M. and Voisard, A. (2001). *Spatial databases: with application to GIS*, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Robin, T., Antonini, G., Bierlaire, M. and Cruz, J. (2009). Specification, estimation and validation of a pedestrian walking behavior model, *Transportation Research Part B: Methodological* **43**(1): 36–56.

Rong, G. and Tan, T.-S. (2007). Variants of jump flooding algorithm for computing discrete voronoi diagrams, *Voronoi Diagrams in Science and Engineering, 2007. ISVD'07. 4th International Symposium on*, IEEE, pp. 176–181.

Ross, S. (2013a). Chapter 11 - Statistical Validation Techniques, *Simulation (Fifth Edition)*, Academic Press, pp. 247 – 270.

Ross, S. (2013b). Chapter 5 - Generating Continuous Random Variables, *Simulation (Fifth Edition)*, Academic Press, pp. 69 – 96.

Saberi, M. and Mahmassani, H. (2014). Exploring areawide dynamics of pedestrian crowds: Three-dimensional approach, *Transportation Research Record: Journal of the Transportation Research Board* **2421**(1): 31–40.

Schueller, A. (2007). A nearest neighbor sweep circle algorithm for computing discrete voronoi tessellations, *Journal of Mathematical Analysis and Applications* **336**(2): 1018–1025.

Seneviratne, P. and Morrall, J. (1985). Analysis of factors affecting the choice of route of pedestrians, *Transportation Planning and Technology* **10**(2): 147–159.

Sexton, G., Zhang, X., Redpath, G. and Greaves, D. (1995). Advances in automated pedestrian counting, *Security and Detection, 1995., European Convention on*, IET, pp. 106–110.

Seyfried, A., Boltes, M., Kähler, J., Klingsch, W., Portz, A., Rupprecht, T., Schadschneider, A., Steffen, B. and Winkens, A. (2010). Enhanced empirical data for the fundamental diagram and the flow through bottlenecks, *Pedestrian and Evacuation Dynamics 2008*, Springer.

Skabardonis, A. and Geroliminis, N. (2008). Real-time monitoring and control on signalized arterials, *Journal of Intelligent Transportation Systems* **12**(2): 64–74.

Steffen, B. and Seyfried, A. (2010). Methods for measuring pedestrian density, flow, speed and direction with minimal scatter, *Physica A: Statistical mechanics and its applications* **389**(9): 1902–1910.

Stubenschrott, M., Kogler, C., Matyus, T. and Seer, S. (2014). A dynamic pedestrian route choice model validated in a high density subway station, *Transportation Research Procedia* **2**: 376–384.

Talebpour, A. and Mahmassani, H. S. (2016). Influence of connected and autonomous vehicles on traffic flow stability and throughput, *Transportation Research Part C: Emerging Technologies* **71**: 143 – 163.
URL: *http://www.sciencedirect.com/science/article/pii/S0968090X16301140*

Tan, B., Zhang, J. and Wang, L. (2011). Semi-supervised elastic net for pedestrian counting, *Pattern Recognition* **44**(10): 2297–2304.

Tanaboriboon, Y., Hwa, S. S. and Chor, C. H. (1986). Pedestrian characteristics study in singapore, *Journal of Transportation Engineering* **112**(3): 229–235.

Tregenza, P. (1976). *The design of interior circulation*, Van Nostrand Reinhold, New York, USA.

Treiber, M. and Helbing, D. (2003). Memory effects in microscopic traffic models and wide scattering in flow-density data, *Physical Review E* **68**(4): 046119–1.

Treiber, M. and Kesting, A. (2013). *Traffic flow dynamics*, Springer, Heidelberg, New York, Dordrecht, London.

Trognon, A. (1987). Les méthodes du pseudo-maximum de vraisemblance, *Annales d'Économie et de Statistique* pp. 117–134.

Underwood, R. T. (1961). Speed, volume, and density relationships: Quality and theory of traffic flow, *Yale Bureau of Highway Traffic, New Haven*, pp. 141–188.

Van den Heuvel, J. and Hoogenraad, J. (2014). Monitoring the performance of the pedestrian transfer function of train stations using automatic fare collection data, *Transportation Research Procedia* **2**: 642–650.

Van der Putte, T. (2009). *Using the discrete 3D Voronoi diagram for the modelling of 3D continuous information in geosciences*, PhD thesis, TU Delft, Delft University of Technology.

van Wageningen-Kessels, F. (2013). *Multi-class continuum traffic flow models: analysis and simulation methods*, PhD thesis, Delft University of Technology/TRAIL Research school, Delft.

van Wageningen-Kessels, F., Hoogendoorn, S. P. and Daamen, W. (2014). Extension of Edie's definitions for pedestrian dynamics, *Transportation Research Procedia* **2**: 507–512.

Verlander, N. Q. and Heydecker, B. G. (1997). Pedestrian route choice: An empirical study, *Proceedings of Seminar F of the PTRC European Transport Forum, P415*, pp. 39–49.

Versichele, M., Neutens, T., Delafontaine, M. and Van de Weghe, N. (2012). The use of bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the ghent festivities, *Applied Geography* **32**(2): 208–220.

Walker, J. L. and Li, J. (2007). Latent lifestyle preferences and household location decisions, *Journal of Geographical Systems* **9**(1): 77–101.

Wang, H., Ni, D., Chen, Q.-Y. and Li, J. (2013). Stochastic modeling of the equilibrium speed–density relationship, *Journal of Advanced Transportation* **47**(1): 126–150.

Wasserman, L. (2000). Bayesian model selection and model averaging, *Journal of mathematical psychology* **44**(1): 92–107.

Weidmann, U. (1993). Transporttechnik der fussgänger, *Technical Report Schriftenreihe des IVT Nr. 90*, Institut für Verkehrsplanung,Transporttechnik,Strassen- und Eisenbahnbau, ETH Zürich. (In German).

Whitt, M. C., DuBose, K. D., Ainsworth, B. E. and Tudor-Locke, C. (2004). Walking patterns in a sample of african american, native american, and caucasian women: the cross-cultural activity participation study, *Health Education & Behavior* **31**(4): 45S–56S.

Wong, S., Leung, W., Chan, S., Lam, W. H., Yung, N. H., Liu, C. and Zhang, P. (2010). Bidirectional pedestrian stream model with oblique intersecting angle, *Journal of transportation Engineering* **136**(3): 234–242.

Wood, S. A., Guerry, A. D., Silver, J. M. and Lacayo, M. (2013). Using social media to quantify nature-based tourism and recreation, *Scientific reports* **3**: 2976.

Wooldridge, J. M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity, *Journal of applied econometrics* **20**(1): 39–54.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*, MIT Press.
URL: *http://www.jstor.org/stable/j.ctt5hhcfr*

Xie, S. and Wong, S. (2015). A bayesian inference approach to the development of a multidirectional pedestrian stream model, *Transportmetrica A: Transport Science* **11**(1): 61–73.

Yaeli, A., Bak, P., Feigenblat, G., Nadler, S., Roitman, H., Saadoun, G., Ship, H. J., Cohen, D., Fuchs, O., Ofek-Koifman, S. et al. (2014). Understanding customer behavior using indoor location analysis and visualization, *IBM Journal of Research and Development* **58**(5/6): 3:1–3:12.

Yoshimura, Y., Sobolevsky, S., Ratti, C., Girardin, F., Carrascal, J. P., Blat, J. and Sinatra, R. (2014). An analysis of visitors' behavior in the louvre museum: A study using bluetooth data, *Environment and Planning B: Planning and Design* **41**(6): 1113–1131.

Zhang, J. (2012). *Pedestrian fundamental diagrams: Comparative analysis of experiments in different geometries*, PhD thesis, Forschungszentrum Jülich.

Zhang, X. and Sexton, G. (1997). Automatic human head location for pedestrian counting, *Image Processing and Its Applications, 1997., Sixth International Conference on*, Vol. 2, IET, pp. 535–540.

Zhao, F., Pereira, F. C., Ball, R., Kim, Y., Han, Y., Zegras, C. and Ben-Akiva, M. (2015). Exploratory analysis of a smartphone-based travel survey in singapore, *Transportation Research Record: Journal of the Transportation Research Board* **2**(2494): 45–56.

Marija Nikolić
EPFL ENAC TRANSP-OR
Station 18
CH-1015 Lausanne
Switzerland

Phone : +41 21 6939327
Website: http://transp-or.epfl.ch/
http://people.epfl.ch/marija.nikolic
E-mail : marija.nikolic@epfl.ch

Education

2012–present	**PhD in Electrical Engineering** École Polytechnique Fédérale de Lausanne (EPFL), Switzerland PhD thesis title: Data-driven fundamental models for pedestrian movements Thesis Supervisor: Prof. Michel Bierlaire
2009–2010	**MSc in Electrical and Computer Engineering (GPA: 10 / 10)** Faculty of Technical Sciences, University of Novi Sad, Serbia MSc thesis title: Filtering data collected from various sources in SCADA and UMS systems Thesis Supervisor: Prof. Aleksandar Erdeljan
2004–2009	**BSc in Electrical and Computer Engineering (GPA: 9.86 / 10)** Faculty of Technical Sciences, University of Novi Sad, Serbia

Professional experience

2012–present	**Research and Teching Assistant** EPFL, Transport and Mobility Laboratory, Switzerland
2010–2012	**Software Engineer** Schneider Electric DMS NS, Novi Sad, Serbia
2008–2010	**Part-time Software Engineer** Schneider Electric DMS NS, Novi Sad, Serbia

Research projects

2012–2015	**Pedestrian dynamics: flows and behavior** Swiss National Science Foundation grant This project aims at developing mathematical models of pedestrian dynamics, both at aggregate and disaggregate levels.

Publications

Papers in international journal

Nikolić, M., Bierlaire, M., Farooq, B., and de Lapparent, M. (2016). Probabilistic speed-density relationship for pedestrian traffic, *Transportation Research Part B: Methodological* **89**:58-81. doi:10.1016/j.trb.2016.04.002

Nikolić, M., and Bierlaire, M. (2014). Pedestrian-oriented Flow Characterization, *Transportation Research Procedia* **2**:359-366. doi:10.1016/j.trpro.2014.09.032

Hänseler, F.S., Lam, W.H.K., Bierlaire, M., Lederrey, G., and Nikolić, M. (2017). A dynamic network loading model for anisotropic and congested pedestrian flows, *Transportation Research Part B: Methodological* **95**:149-168. doi:10.1016/j.trb.2016.10.017

Papers in conference proceedings

Nikolić, M., and Bierlaire, M. (to appear). Data-driven spatio-temporal discretization for pedestrian flow characterization. Proceedings of the 22nd International Symposium on Transportation and Traffic Theory (ISTTT), Northwestern University, 24-26 July, 2017.

Nikolić, M., and Bierlaire, M. (2015). Pedestrian flow characterization based on spatio-temporal Voronoi tessellations. Proceedings of the 15th Swiss Transportation Research Conference (STRC) 15-17 April, 2015.

Nikolić, M., Bierlaire, M., and Farooq, B. (2014). Probabilistic speed-density relationship for pedestrians based on data driven space and time representation. Proceedings of the Swiss Transportation Research Conference (STRC) 14-16 May, 2014.

Nikolić, M., Farooq, B., and Bierlaire, M. (2013). Exploratory analysis of pedestrian flow characteristics in mobility hubs using trajectory data. Proceedings of the Swiss Transportation Research Conference (STRC) 24-26 April, 2013.

Technical reports (selection)

Nikolić, M., Bierlaire, M., de Lapparent, M., and Scarinci, R. (2017). Multi-class speed-density relationship for pedestrian traffic. Technical report TRANSP-OR 170115.

Nikolić, M., and Bierlaire, M. (2016). Data-driven characterization of pedestrian flows. Technical report TRANSP-OR 160815.

Nikolić, M., Bierlaire, M., and Farooq, B. (2015). Probabilistic speed-density relationship for pedestrian traffic: a data-driven approach. Technical report TRANSP-OR 150411.

Seminars (selection)

Nikolić, M., and Bierlaire, M., Data-driven characterization of pedestrian flows. hEART 2016, 5th Symposium of the European Association for Research in Transportation, Delft University of Technology, September 14, 2016, Delft, the Netherlands

Nikolić, M., Bierlaire, M., de Lapparent, M., and Scarinci, R., Multi-class speed-density relationship for pedestrian traffic. hEART 2016, 5th Symposium of the European Association for Research in Transportation, Delft University of Technology, September 14, 2016, Delft, the Netherlands

Nikolić, M., and Bierlaire, M., Data-driven characterization of pedestrian traffic. Ninth Triennial Symposium on Transportation Analysis (TRISTAN IX), June 14, 2016, Oranjestad, Aruba

Nikolić, M., Bierlaire, M., de Lapparent, M., and Scarinci, R., Multi-class speed-density relationship for pedestrian traffic. Ninth Triennial Symposium on Transportation Analysis (TRISTAN IX), June 14, 2016, Oranjestad, Aruba

Nikolić, M., and Bierlaire, M., Data-driven characterization of pedestrian flows. 16th Swiss Transportation Research Conference, May 18, 2016, Ascona, Switzerland

Nikolić, M., Bierlaire, M., and Hänseler, F.S., Data-driven characterization of multidirectional pedestrian traffic. TGF '15, TU Delft, October 28, 2015, Delft, The Netherlands

Nikolić, M., Bierlaire, M., and Hänseler, F.S., Characterization of multidirectional pedestrian flows based on three-dimensional Voronoi tessellations. 4th Symposium of the European Association for Research in Transportation, Technical University of Denmark, September 09, 2015, Copenhagen, Denmark

Nikolić, M., and Bierlaire, M., Pedestrian flow characterization based on spatio-temporal Voronoi tessellations. 15th Swiss Transportation Research Conference, April 15, 2015, Ascona, Switzerland

Nikolić, M., and Bierlaire, M., Pedestrian-oriented flow characterization. Pedestrian and Evacuation Dynamics conference (PED 2014), Delft University of Technology, October 24, 2014, Delft, the Netherlands

Nikolić, M., Bierlaire, M., and Farooq, B., Probabilistic speed-density relationship for heterogeneous pedestrian traffic. 3rd Symposium of the European Association for Research in Transportation, University of Leeds, September 11, 2014, Leeds, United Kingdom

Nikolić, M., Bierlaire, M., and Farooq, B., Probabilistic speed-density relationship for pedestrians based on data driven space and time representation. Swiss Transport Research Conference, May 15, 2014, Ascona, Switzerland

Nikolić, M., Bierlaire, M., and Farooq, B., Probabilistic speed-density relationship for pedestrians based on data driven space and time representation. WORKSHOP ON PEDESTRIAN MODELS 2014, EPFL, April 10, 2014, Lausanne, Switzerland

Nikolić, M., Bierlaire, M., and Farooq, B., Spatial tessellations of pedestrian dynamics. 2nd Symposium of the European Association for Research in Transportation, September 05, 2013, Stockholm, Sweden

Nikolić, M., Pedestrian flow characteristics based on individual trajectories. DATA SIM Summer School 2013, Mobility Modeling and Big Data Sources, Hasselt University, July 15, 2013, Hasselt, Belgium

Nikolić, M., Farooq, B., and Bierlaire, M., Exploratory analysis of pedestrian flow characteristics in mobility hubs using trajectory data. Swiss Transportation Research Conference (STRC), April 25, 2013, Ascona, Switzerland

Teaching

Teaching assistant for courses

- Optimization and simulation, PhD course, EPFL (2014–2017)

Supervisor for Master theses

- Information system for passenger discretionary activities in transport terminals, Master thesis, Federico Orsini (Section of Civil Engineering), 2016/2017
- A multi-class framework for a pedestrian cell transmission model accounting for population heterogeneity, Master thesis, Guy Cooper (Section of Mathematics), 2014

- Mode Choice Analysis Using Smartphone Data, Master thesis, Michael Friederich (Section of Civil Engineering), 2014

Supervisor for semester projects

- Planning of feeding station installment for a full electric large capacity urban bus system, Adrien Ruault (Section of Mathematics), 2017.
- Models for pedestrian movements based on integrated and sequential clustering, Konde Romain Olivier Bondo (Section of Civil Engineering), 2017.
- Accounting for dynamics in pedestrian multi-class speed-density relationship, Marc-Edouard Schultheiss (Section of Civil Engineering), 2016.
- Specification testing of fundamental diagrams for an anisotropic pedestrian network loading model, Joel Mateus Fonseca (Section of Communication Systems), 2016.
- Development of a novel pedestrian walking model applicable to congested flows, Gael Lederrey (Section of Mathematics), 2015.
- Pedestrian movement in train stations: modeling speed-density relationship for different classes of passengers, Laure Emma Rosine (Section of Communication Systems), 2015.
- Exploring pedestrian mobility using video tracking data in Lausanne train station, Babel Hugo Louis (Section of Mathematics), 2014.
- Mode choice analysis from a large smartphone dataset, Mikael Nicolas Xavier Friederich (Section of Civil Engineering), 2014.
- Exploration of smartphone users trip data to investigate travel behavior, Mikael Nicolas Xavier Friederich (Section of Civil Engineering), 2014.
- Movement patterns of pedestrians on platforms prior to/after train departures/arrivals in Gare de Lausanne: Exploitation of pedestrian tracking data, Isabel Tovar (Section of Civil Engineering), 2013.
- Pedestrian flow simulation in Lausanne train station, Nicolas de Lamberterie (Section of Civil Engineering), 2013.

Reviewing

- Transportation Research Part B: Methodological
- Transportation Research Part C: Emerging Technologies
- Transportation Research Part F: Traffic Psychology and Behaviour
- Journal of Advanced Transportation

- IET Intelligent Transport Systems
- Transportation Letters: the International Journal of Transportation Research

Awards and distinctions

- Best Student Award at the promotion of the Faculty of Technical Sciences (2010)
- Scholarship of the Republic of Serbia for the best one thousand talented students (2009-2010)
- Scholarship of the Schneider Electric DMS NS gifted students (2008-2010)
- Extraordinary Award for Excellent Achievement at the University of Novi Sad (2006-2009)
- 2nd Prize as a team at International Competition of Electrical Engineering in Automatics (2008)