TECHNISCHE UNIVERSITÄT MÜNCHEN

Ingenieurfakultät Bau Geo Umwelt,
Lehrstuhl für Verkehrstechnik

# Development of tactical and operational behaviour models for bicyclists based on automated video data analysis

Heather A. Twaddle, M.Sc.

Vollständiger Abdruck der von der Ingenieurfakultät Bau Geo Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines Doktor-Ingenieurs genehmigten Dissertation.

Vorsitzender: Prof. Dr.-Ing. Gebhard Wulfhorst

Prüfer der Dissertation:

1. Prof. Dr.-Ing. Fritz Busch
2. Prof. Dr. Nicolas Saunier
   Associate Professor, École Polytechnique de Montréal, Kanada
3. Prof. Dr. phil. Klaus Bengler
   (schriftliche Beurteilung)

Die Dissertation wurde am 05.07.2017 bei der Technischen Universität München eingereicht und durch die Ingenieurfakultät Bau Geo Umwelt am 15.09.2017 angenommen.

## Acknowledgements

The work presented in this dissertation was carried out while working as a researcher at the Chair of Traffic Engineering and Control, mainly within the framework of the research project UR:BAN (Urban Space: User oriented assistance systems and network management), which was supported by the German Federal Ministry of Economic Affairs and Energy. I would like to thank Fritz Busch for giving me the opportunity to work on such an interesting, relevant and challenging topic. I very much appreciate his trust in my research and German skills as well as his valuable input to this dissertation.

This work would not have been possible without the efforts of Nicolas Saunier. His development of the open source software Traffic Intelligence (and his support using the software) made the data analysis carried out in this dissertation possible. Not only that, but my stay with him and his team at the École Polytechnique de Montréal gave me the opportunity to take big leaps forward in my research. Finally, I am very thankful to Nicolas for taking over the role of second supervisor and for the detailed feedback he provided.

Thank you very much to Michel Bierlaire for his valuable input and warm welcome during my stay at the TRANSP-OR Lab at École Polytechnique Fédérale de Lausanne. Thank you to all the wonderful researchers at the lab – what a great team. Special thanks to Marija Nikolic and Riccardo Scarinci for their excellent input and suggestions concerning the operational behaviour model.

Thank you to all my colleagues, current and former, at the Chair of Traffic Engineering and Control for making the last six years so enjoyable and memorable. Thank you especially to Silja Hoffmann, my group leader, for her unwavering support and cheerful mood. Special thanks to Tobias Schendzielorz for giving me such a good start into the project UR:BAN and working in the German language.

A special thank you goes to all the building owners who allowed me to install my data collection system on their buildings for couple of days. Thank you also to the City of Munich (Klaus Krämer in particular) for providing me with signal timing data from the intersections. It was a pleasure to meet so many helpful people in the course of this research.

Thank you to my family and friends, both here in Germany and back home in Canada. Your love and support mean so much to me. Special thanks to my dad for editing an early version of this dissertation.

Last but certainly not least, I would like to thank my brilliant husband-to-be Jakob Kaths for all his support, input, commiseration and excellent humour during this whole process. I'll never know how I got so lucky.

## Executive Summary

The bicycle offers an environmentally friendly transportation option that is space efficient, inexpensive and supportive of personal health through physical activity. Moreover, bicycling is often the fastest mode of transportation for short trips, particularly in urban areas. Despite the numerous advantages of bicycling, difficulties persist in integrating bicycle traffic into the road environment while protecting bicyclists' safety and ensuring efficient traffic flow for all road users. A commonly used method for developing and evaluating measures to improve traffic safety and efficiency is microscopic traffic simulation. However, modified driver or pedestrian models are typically used to simulate bicycle traffic in these tools. These approaches do not adequately reflect the flexibility and adaptability of bicyclists.

In this dissertation, real data from bicyclists at signalised intersections are used to specify, calibrate and validate behavioural models for application in microscopic traffic simulation. Video data is collected at four intersections, differing from one another in terms of geometry and traffic volume. Automated video analysis is used to extract trajectories, which quantify the spatial progression of road users, in a subset of the video data. Automated methods for classifying road users as bicyclists, pedestrians or motor vehicles, correcting distortion in the trajectory data resulting from a wide angle lens and identifying the manoeuvre (right turn, left turn or travelling straight across the intersection) of each bicyclist are implemented.

Two methods for clustering the pathways used by bicyclists to carry out their desired manoeuvres are introduced. The first of which clusters the pathways of bicyclists arriving on a given approach of a given intersection. This information is used to specify the desired pathways in the forthcoming operational behaviour modelling approach. The second method generically clusters the types of pathways used by bicyclists to carry out a given manoeuvre (e.g. left turn) independent of the approach or intersection. The clustering results from the generic method are used in the subsequent modelling of the tactical decisions of bicyclists.

Models are developed based on the clustered trajectories to better represent the behaviour of bicyclists in microscopic simulation on two levels; tactical and operational behaviour. Tactical behaviour encompasses conscious decisions made on a time scale of seconds to minutes that allow a road user to fulfil strategic plans while coping with the current situation. Behaviour at this level is modelled using logistic regression models with a choice set defined based on the results of the generic clustering method. Operational behaviour includes subconscious action patterns realised on a timescale of milliseconds to seconds in order to respond to the immediate situation. A modified social force model is specified here that enables the straightforward restriction of movement in the lateral and longitudinal directions. Models at both the tactical and operational level are calibrated and evaluated individually using K-fold cross validation.

The resulting behavioural models are integrated with the microscopic traffic simulation software SUMO to evaluate the overall ability of the approach to realistically simulate bicycle traffic. The integrated models are evaluated by comparing the positions and speeds of simulated bicyclists (SUMO default approach and proposed modelling approach) and bicyclists observed in reality. The shape of simulated queues as well as the ability to realistically simulate left turn manoeuvres are evaluated. Results indicate that the proposed integrated modelling approach is capable of realistically simulating the flexible behaviour of bicyclists at signalised intersections. The dispersed positioning of bicyclists across the area of the intersection can be simulated without manually specifying a multitude of links and intersection points. This enables the realistic simulation of manoeuvres that are tedious to simulate in currently available simulation software. Particularly the simulation of left turn manoeuvres, including indirect left turns with and against the mandatory direction of travel, is simplified using the proposed approach. Finally, the simulation of bicycle queues that are irregular and differ between phases is possible.

## Zusammenfassung

Das Fahrrad ist ein umweltfreundliches, platzsparendes und kostengünstiges Verkehrsmittel, das die persönliche Gesundheit des Nutzers durch körperliche Bewegung fördert. Des Weiteren stellt das Fahrradfahren auf kürzeren Entfernungen oft die schnellste Verkehrsmitteloption dar, insbesondere im urbanen Raum. Trotz der zahlreichen Vorteile der Fahrradnutzung treten weiterhin Schwierigkeiten bei der Integration des Fahrradverkehrs in den Straßenraum auf. So müssen sowohl Sicherheit als auch Effizienz unterschiedlicher Verkehrsteilnehmergruppen wie dem motorisierten Individualverkehr, Fußgängern und Radfahrern berücksichtigt werden, wodurch abweichende oder sogar gegensätzliche Ziele auftreten können. Eine häufig verwendete Methode zum Entwurf und zur Bewertung von verkehrlichen Maßnahmen, die auf die Verkehrssicherheit und -effizienz ausgerichtet sind, ist die mikroskopische Verkehrssimulation. Die angepassten Fahrzeug- oder Fußgängermodelle, die in der mikroskopischen Verkehrssimulation üblicherweise zur Abbildung des Radfahrerverhaltens verwendet werden, reichen jedoch nicht aus, um die gewünschte Genauigkeit und Realitätsnähe bei der Abbildung des Radverkehrs zu erreichen.

In dieser Dissertation werden Realdaten von Radfahrern an signalisierten Knotenpunkten verwendet, um Verhaltensmodelle für den Einsatz in der mikroskopischen Verkehrssimulation zu spezifizieren, zu kalibrieren und zu validieren. Es werden Videodaten an vier städtischen Knotenpunkten erhoben, die sich jeweils in ihrer Geometrie und dem auftretenden Verkehrsaufkommen unterscheiden. Bei der Überquerung eines Knotenpunktes durch einen Verkehrsteilnehmer wird sein Bewegungsablauf räumlich und zeitlich quantifiziert. Diese Bewegungsabläufe werden mittels einer automatisierten Videodatenanalyse extrahiert. Zur Klassifizierung der Verkehrsteilnehmer als Radfahrer, motorisierte Fahrzeuge oder Fußgänger, zur Korrektur der Verzerrung in den Bewegungsabläufen, die durch das Weitwinkelobjektiv verursacht wurden, sowie zur Identifizierung der Fahrmanöver (Rechtsabbiegen, Linksabbiegen, Geradeausfahren) werden automatisierte Werkzeuge entwickelt und bewertet.

Zur Gruppierung der Pfade, welche von den Radfahrern zur Umsetzung ihres gewünschten Fahrmanövers gewählt werden, werden zwei unterschiedliche Ansätze eingeführt. Der erste Ansatz gruppiert die Pfade der Radfahrer, die den betrachteten Knotenpunkt an einer bestimmten Zufahrt erreichen. Die aus der Gruppierung resultierenden repräsentativen Pfade dienen als Vorgabe für das entwickelte Modell zur Abbildung des operativen Verhaltens. Mithilfe des zweiten Ansatzes werden die Pfade gruppiert, die von Radfahrern zur Durchführung eines bestimmten Fahrmanövers (z.B. Linksabbiegen) genutzt werden. Im Falle des zweiten Ansatzes geschieht dies jedoch unabhängig vom Knotenpunkt oder der betrachteten Zufahrt. Diese Gruppierungsergebnisse fließen in die Modellierung der taktischen Entscheidungen der Radfahrer ein.

Basierend auf den Ergebnissen der Gruppierungsanalyse werden Modelle entwickelt, um das Verhalten von Radfahrern in der mikroskopischen Verkehrssimulation auf zwei Ebenen besser abzubilden: der taktischen und der operativen Verhaltensebene. Taktisches Verhalten umfasst bewusste Entscheidungen, die in einem zeitlichen Horizont von Sekunden bis Minuten getroffen werden, um strategische Pläne zu erfüllen, während die aktuelle Verkehrssituation bewältigt wird. Das Verhalten auf dieser Ebene wird anhand von logistischen Regressionsmodellen prädiziert. Dabei werden die Alternativen für die Modelle basierend auf den Ergebnissen der Gruppierungsanalyse definiert. Das operative Verhalten beinhaltet hingegen unterbewusste Bewegungsmuster zur Reaktion auf die aktuelle Situation, die in einem zeitlichen Horizont von Millisekunden bis Sekunden realisiert werden. Bekannte Social Force Modelle werden um die Einschränkung der Bewegung in lateraler und longitudinaler Richtung erweitert. Modelle beider Verhaltensebenen werden mit den realen Daten kalibriert und mithilfe von k-facher Kreuzvalidierung bewertet.

Die resultierenden Verhaltensmodelle werden zur Evaluierung der Gesamtleistungsfähigkeit der Ansätze in das mikroskopische Verkehrssimulationsprogramm SUMO integriert. Mit Hilfe von Heatmaps, welche Positionen und Geschwindigkeiten des Radfahrerkollektivs anzeigen, werden die entwickelten Modelle mit dem SUMO Standardmodell sowie den Realdaten der beobachteten Radfahrer verglichen. Des Weiteren werden die Form der simulierten Warteschlangen sowie die Fähigkeit des entwickelten Ansatzes, realistische Linksabbiegemanöver abzubilden, bewertet. Die Ergebnisse zeigen, dass der vorgestellte Modellierungsansatz das flexible und anpassungsfähige Verhalten der Radfahrer an signalisierten Knotenpunkten realistisch abbilden kann. Die im Knotenpunktbereich stark verteilten Positionen der Radfahrer können implizit simuliert werden, ohne eine Vielzahl von Pfaden und Konfliktpunkten manuell vorgeben zu müssen. Dies ermöglicht eine realitätsnahe Simulation verschiedener Fahrmanöver, die derzeit nur mit erheblichem Aufwand mit Hilfe herkömmlicher Simulationsprogramme abgebildet werden können. Insbesondere die Simulation von Linkabbiegemanövern, einschließlich dem indirekten Linkabbiegen mit und entgegen der vorgegebenen Fahrtrichtung, wird mit den entwickelten Ansätzen vereinfacht. Darüber hinaus ist eine realitätsnahe Abbildung von Warteschlangen möglich, die deren zeitlichen und räumlichen Unregelmäßigkeiten Rechnung trägt.

# Table of Contents

“Whatever you do, don't congratulate yourself too much, or berate yourself either. Your choices are half chance, so are everybody else's.”

Mary Theresa Schmich - Wear Sunscreen

# 1. Introduction

In this section, the motivation and framework for this research work are established. Starting with a historical review of bicycling as a mode of transport, the developments and current problems with safety and traffic efficiency relating to bicycling in urban areas are discussed. The scope and research questions pursued in this work are then defined are based on this contextual review and the methodology used to investigate the research questions is outlined. Finally, the contributions to research in the field of trajectory data processing and modelling the behaviour of bicyclists are presented.

## 1.1 Context

Since the introduction of the first two-wheeled vehicle, the *Laufmaschine* (dandy horse), by Baron Karl von Drais in Germany in 1817, the bicycle has played a pivotal role in providing utilitarian mobility and a leisure activity. Although first viewed as a daredevil machine for young, sporty men, design innovations quickly made the bicycle attractive and accessible to all. In the late 19$^{th}$ century and early 20$^{th}$ century, the bicycle enjoyed immense popularity in both Europe and North America. The growing bourgeoisie class rode for leisure in city parks and toured the countryside on bicycles. The practical applications of the bicycle were soon realised as well and the bicycle became popular for commuting and carrying out daily tasks [OLDENZIEL & ALBERT DE LA BRUHÈZE, 2011]. By the year 1930, there were over 40 million bicycles in Europe, more than seven bicycles for each car [EBERT, 2004].

It was not until after World War I that the focus of urban developers and traffic engineers shifted to the needs of motorised vehicles. Bicyclists and pedestrians were held largely responsible for traffic congestion and accidents and a movement was made to relocate these modes to the side of the roadway on newly built sidewalks and bicycle lanes. Although the stated motivation for this shift was the safety and comfort of vulnerable road users, the underlying incentive was likely to clear the roadway for motor vehicles. This marginalisation, as well as the cultural trend of associating motor vehicles with the future of mobility and the bicycle with the poorer working class, resulted in a massive decrease in bicycling rates from the end of World War I until about 1970 [PUCHER & BUEHLER, 2008].

In response to growing problems with traffic safety, environmental pollution and road congestion caused by automobiles, transport and land use policies in many European countries shifted back towards favouring bicycling during the 1970s. Taxation and policies that restricted motor vehicle use and increased driving costs were put in place, while bicycle infrastructure was expanded and bicycle use was politically supported. This resulted in an increase in bicycling in European countries such as Germany, the Netherlands and Denmark from the mid-1970s until today. Current cycling rates in Europe, measured as the percentage

of all trips, range from about 2% in the United Kingdom to about 30% in the Netherlands [PUCHER & BUEHLER, 2008].

No such shift in policy took place during the 1970s in North America. Sprawling cities built during the motor vehicle centric years of the mid-20th century are less conducive to bicycle transportation and the lack of bicycling history made it difficult to reignite the cultural flame for utilitarian bicycling. Today, the United States and Canada have two of the lowest bicycling rates in the world with only 1% and 2% of all trips made by bicycle, respectively [PUCHER & BUEHLER, 2008]. Nevertheless, a movement to increase the attractiveness of bicycling has taken hold there as well during the last decades [DILL & CARR, 2003]. In response to many of the same motivating factors experienced in Europe in the 1970s, North American transport planners and urban policy makers today are seeking to (re-)build infrastructure and enact policies that reduce motor vehicle use and encourage bicycling.

However, the safety of bicyclists must be protected if this mode is to be supported as a daily transport option. Due to the expansion and improvement of bicycle infrastructure, the reduction of speed limits in urban areas and numerous other factors, the number of bicyclist fatalities has decreased by about 70% in Germany, the Netherlands and Denmark since 1970 [PUCHER & BUEHLER, 2008]. Although there is a general trend towards a safer road environment, the sheer number of serious injuries and fatalities per unit of exposure is much higher for bicyclists than motorised road users. In 2015, 19.7% of all injured and killed road users in Germany were bicycling [STATISTISCHES BUNDESAMT, 2016] despite an overall modal share of bicycling of 9.8% [RADKE, 2015]. This disproportion is even more pronounced in urban areas, where 27.9% of severe and minor injuries were inflicted on bicyclists. At urban intersections, GERSTENBERGER [2015] found that 39.1% of all collisions involve a bicyclist.

The development of road fatalities and serious injuries for motorists (drivers and passengers) and bicyclists in Germany, the UK and the Netherlands per billion person-kilometres travelled are shown in Fig 1.1. Serious injuries and fatalities are defined and listed by the respective national traffic data analysts in the three countries (Germany: STATISTISCHES BUNDESAMT [2014; 2016], The Netherlands,: SWOV [2016], UK: DEPARTMENT FOR TRANSPORT [2017]). The exposure rate used is billion person-km and these data are also collected from respective national traffic data analysts (Germany: RADKE [2015], The Netherlands: KENNISINSTITUUT VOOR MOBILITEITSBELEID (KIM) [2013], UK: DEPARTMENT FOR TRANSPORT [2017]).

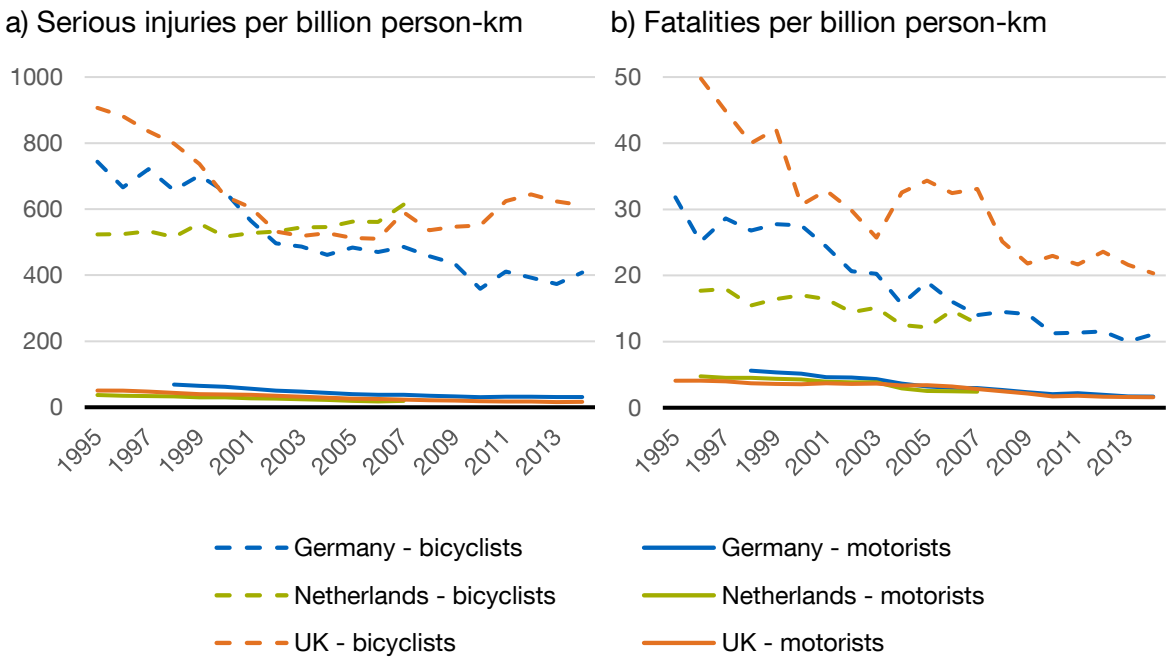



**Fig 1.1** Serious injuries (a) and fatalities per billion person-km in Germany, the Netherlands and the UK between 1995 and 2014

In addition to concerns about traffic safety, bicycle traffic has a significant impact on overall traffic flow in cities with at least a moderate share of bicycling. If bicyclists share the road with motor vehicles, slower moving bicyclists can impede the movement of motor vehicles and slow the overall driving speed on road segments. Bicyclists too are impacted by congestion caused by motor vehicle traffic. In general, the influence of both modes on one another plays an important role in the overall efficiency and capacity of shared roadways. If bicycle traffic is separated from motor vehicle traffic on bicycle specific facilities, the segregated traffic streams must nonetheless interact at intersections. The most obvious impact of bicycle traffic on the capacity at intersections occurs due to the delay of left and right turning vehicles that share a signal phase with bicyclists travelling straight across the intersection. According to the traffic laws in many countries, these left and right turning vehicles must yield to the right of way of bicycles and pedestrians that are crossing the intersection in the same phase. At non-signalised intersections, the interactions and cooperation between bicyclists, pedestrians and motor vehicles strongly influence the safety and efficiency of all road users.

As the demand for mobility continues to rise, with increasing motor vehicle, bicycle and pedestrian flows in urban centres, the need for innovative solutions to satisfy this demand using limited road space rises. Experts in many fields, including traffic engineering, urban planning and automotive engineering are creating solutions to improve bicyclist safety and to create the most efficient traffic flow for all road users. Advances in technologies for sensing and interpreting the road environment and then reacting safely to the perceived situation form the basis of advanced driver assistance systems (ADAS). Within the project UR:BAN (Urban Space: User oriented assistance systems and network management), innovative systems

were developed by partners in the automotive industry for the automated detection and reaction to vulnerable road users [UR:BAN PROJECT PARTNERS, 2016]. In the near future, automated motor vehicles must be able to identify bicyclists, predict their behaviour and respond safely in order to move independently through the urban environment. Innovations in the realm of intelligent transportation systems (ITS), including green light speed advice for bicyclists and adaptive traffic control capable of prioritising bicycle traffic, are also being developed and have the potential to further improve mobility for bicyclists.

In addition to innovations in the realm of ADAS and ITS, the bicycle in itself is changing. Electrification is enabling bicycle use for longer trips and by a more diverse group of people. Cargo bicycles and trailers are becoming more popular for transporting goods and children in urban areas. Expected outcomes of these trends include wide-ranging dynamic characteristics (e.g. speed, acceleration and turning radii) and diverse tactical decisions (e.g. stopping at red lights or carrying out various left turn manoeuvre types) of bicyclists. The application of various ITS and ADAS measures to target the safety and efficiency of different types of bicycles and bicyclists will be an important aspect for future research. However, in order to develop these new measures, it is imperative to understand the behaviour of different types of bicyclists. Not only is an understanding of the behaviour necessary, but mathematical models for describing and predicting behaviour are essential for creating useful solutions and evaluating their effects.

A tool used commonly for development and evaluation ADAS and ITS measures is microscopic traffic simulation. Simulation tools are built from a collection of models that desribe the behaviour of road users and the road environment. According to BARCELÓ [2010], the cornerstone of all types of traffic simulation is that traffic is a system comprised of many interconnected, complex and functionally related components, each of which can be modelled independently. The behaviour of the entire traffic system cannot be reduced to the sum of the individual components. In other words, interactions and interdependencies between the components lead to a traffic system that is more than the merely the sum of its components. This can be illustrated by a number of road users moving independently through the road environment. The independent motion of each road user can be predicted relatively easily. However, when road users interact with one another, it becomes more difficult to predict the actions of each individual road user. This is an example of an emerging complex system. The realistic simulation of the complete traffic system, therefore, requires accurate modelling of all the components of the system as well as realistic representations of the interactions and interdependencies between the components.

This holistic view of traffic systems as well as the ensuing analysis of traffic using traffic simulation is particularly attractive for urban environments, where many actors and situational factors come into play simultaneously. A microscopic traffic simulation calibrated and validated using data from an existing scenario can be used to infer or predict the effects of

measures that are not yet in place or the large-scale deployment of existing measures. Microscopic traffic simulation offers an attractive solution for assessing the effect of ADAS and ITS. Due to the scale of microscopic traffic simulation, efficiency parameters such as the average and maximum travel time, the number of stops, delay and the speed profiles of individual road users can be examined. Although simulations are typically accident free, the interactions between road users and the resulting surrogate measures of safety, such as time to collision (TTC) and post-encroachment time (PET), can be investigated using microscopic traffic simulation, once a thorough calibration and validation have been carried out.

Unfortunately, models currently used to include bicycle traffic in microscopic traffic simulations do not adequately reflect the unique and flexible behaviour of bicyclists in a realistic and detailed manner. There are many ways in which bicyclists take advantage of their small size, high manoeuvrability and wide range of travelling speeds to tactically move through the road environment. For example, bicyclists are able to choose between riding on the roadway, sidewalk or a bicycle facility. The direction of travel, as well as the type of turning manoeuvres carried out at intersections can be tactically selected by the bicyclist depending on their preferences, strategic goals and willingness to disobey traffic regulations. In a study of the accident risk and rule acceptance of bicyclists in six German cities, ALRUTZ ET AL. [2009] found that about 500 of the 1400 bicyclists (36%) violated a traffic regulation at least once during their trip. This flexible and occasionally non-compliant behaviour of bicyclists makes it a unique challenge to realistically model and simulate bicycle traffic.

Despite these differences, bicycle traffic is typically simulated using the same models that are used to model the behaviour of motorised vehicles, with model extensions allowing simulated bicycles more freedom of movement in the lateral direction in some cases. It is nonetheless very difficult to simulate the high flexibility and the wide range of trajectories observed in reality using current simulation approaches. Particularly the complex tactical behaviour of bicyclists is difficult to simulate using commercially or publically available microscopic simulation software [BARCELÓ, 2010; KRAJZEWICZ ET AL., 2014; TWADDLE, SCHENDZIELORZ & FAKLER, 2014]. As a result of the unrealistic modelling of the bicycle traffic, the subsequent microscopic simulation of the entire traffic system is flawed due to the interactions between models in the simulation.

Based on these motivating factors, the goal of this dissertation is to provide an approach for modelling the operational and tactical behaviour of bicyclists for application in microscopic traffic simulation. This method will enable more accurate investigations of the effects of ADAS, ITS and infrastructure design on the safety of bicyclists and the efficiency of traffic flow using microscopic traffic simulation.

## 1.2 Scope

The three aspects explained below are used to define the focus area and boundaries of this dissertation.

**Location:** The entire road network is relevant for the analysis of road safety and efficiency using microscopic traffic simulation. The scope of this dissertation, however, is confined to signalised intersections. This focus is justified by the disproportionately high safety risk to bicyclists at intersections in comparison to road segments [GERSTENBERGER, 2015; STATISTISCHES BUNDESAMT, 2016]. In addition, many measures for improving the safety and efficiency of bicycle traffic are deployed at signalised intersections. Examples of such measures include signal optimisation for bicyclists and ADAS systems that warn drivers of bicyclists approaching from behind a vehicle. Furthermore, the accurate modelling of interactions between motorised road users and bicyclists at intersections, which occur even when bicycle traffic is separated on a bicycle path, is paramount in the realistic simulation of urban networks.

**Behavioural level:** According to MICHON [1985], the behaviour of road users can be classified into three categories depending on the time scale and consciousness of the decisions. At the strategic level, decisions about general mobility (e.g. housing and workplace location, recreational activities) and particular trips (e.g. destination, time of departure, mode of travel, route) are made. Once on a given trip, actions of a road user must be adjusted and controlled on a time horizon of minutes or seconds to cope with the immediate traffic situation. These behaviours, which include pathway planning and the adherence to traffic regulations, are included at the tactical level. At the bottom level of the categorization framework, the operational level, road users make subconscious decisions to realise the behaviours determined at the tactical level. These automatic action patterns include acceleration, deceleration and positioning. Although all levels are important for the realistic simulation of bicycle traffic, the scope of this dissertation is restricted the tactical choices and operational behaviour of bicyclists. This selection is made in consideration of the important role of the behaviour of bicyclists at these two levels in traffic safety and efficiency at signalised intersections.

**Simulation resolution:** Three types of traffic simulation can be identified with regard to the level of aggregation. According to BARCELÓ [2010], macroscopic traffic simulation is based on an analogy to hydrodynamics; traffic streams with defined volumes, densities and speeds flow through the road network in the same way that a fluid flows through a network of pipes. Microscopic traffic simulation is based on a disaggregated approach in which the behaviour of each road user and the interactions between road users are modelled. The overall traffic flow results from the aggregation of the individual movements. Mesoscopic traffic simulation is an intermediate approach that makes use of simplified models. Microscopic traffic

simulation is most suitable for investigating the safety and traffic efficiency on specific road segments or intersections and is thus the focus of the dissertation research.

The following research question guides the work done in this dissertation:

> *How can the operational and tactical behaviour of bicyclists at signalised intersections be modelled to evaluate bicyclist safety and overall traffic efficiency using microscopic traffic simulation?*

In this dissertation, all bicyclists, including all types of people (e.g. gender, age, experience) and all forms of bicycles (e.g. pedelecs, e-bikes, cargo bikes), are included in one population. Although the dynamic characteristics and tactical behaviour of different types of bicyclists are assumedly very different, the aim here is to develop a generic model for all bicyclists that can be calibrated for different types of bicyclists in future applications.

## 1.3 Methodology

The methodology followed in this dissertation is summarised in Fig 1.2. In a first step, previous research investigating the behaviour of bicyclists is thoroughly reviewed (Section 2). Mathematical models proposed by researchers to replicate the operational and tactical behaviour of bicyclists are examined. Particular focus is placed on models that are currently implemented in commercially and publically available microscopic traffic simulation tools. The product of the literature review is a needs assessment derived from the critical comparison of the capabilities of current modelling approaches and the behaviour of bicyclists in reality.

In Section 3, the experimental design is elaborated including data specification, the sampling approach and the selection of research intersections. In addition to the behavioural data from bicyclists, necessary data describing interacting road users and the situation in which the bicyclists find themselves are specified. Once the overarching procedure is determined, the specific methods for data collection and processing are defined. The medium for data collection and the techniques used to process the collected data are selected and necessary software extensions are identified. Finally, data collection and processing including video data collection, the automated extraction of road user trajectories, the post processing of the trajectory data and the manual collection of additional variables are explained.

The trajectory data is clustered in Section 4 to gain an understanding of the pathways used by bicyclists to cross an intersection while performing different manoeuvres, including travelling straight across the intersection, turning right and turning left. An approach specific and generic clustering approach are presented.

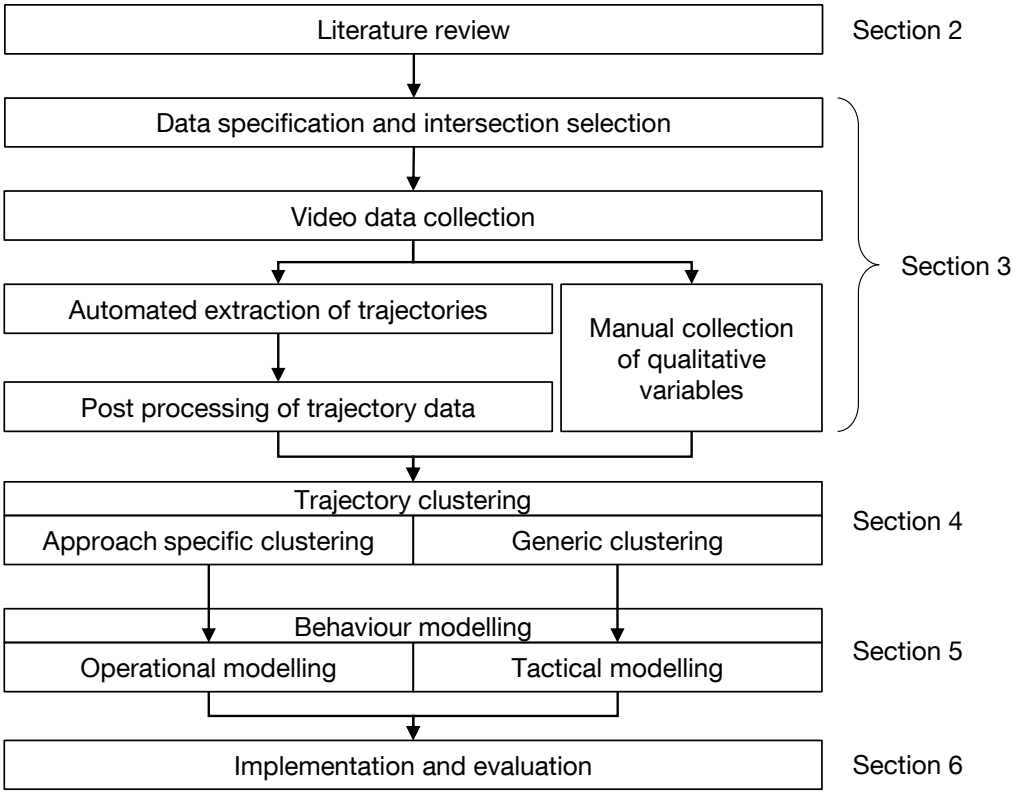



**Fig 1.2** Research workflow

Models for the operational and tactical behaviour are specified in Section 5 based on knowledge gained through the literature review and behavioural hypotheses developed by observing the behaviour of bicyclists at the research intersections. The operational behaviour models are calibrated and validated using trajectory data from three of the four research intersections. This is performed using an iterative process in which a given model is specified, calibrated, validated, improved and then re-calibrated and validated until a final version emerges that accurately replicates the behaviour of bicyclists. Tactical models are developed using data from bicyclists observed carrying out selected behaviours at each of the four research intersection.

The feasibility of the final models is evaluated in Section 6 using the microscopic simulation tool SUMO [KRAJZEWICZ ET AL., 2006; KRAJZEWICZ & HERTKORN, 2002]. The fourth research intersection, which was not used for the development and calibration of the operational behaviour models, is simulated. The ability of the models to replicate the behaviour of bicyclists at this intersection is assessed using measures that are important in the simulation of driver assistance and traffic control systems.

## 1.4 Contributions

The overarching goal of this dissertation is to model the operational and tactical behaviour of bicyclists. In the process of accomplishing this objective, a number of additional research contributions are achieved. These contributions include:

1 A thorough review of international literature concerning the operational and tactical behaviour of bicyclists as well as the methods for modelling and simulating these behaviours in microscopic traffic simulation.

2 The extension of an existing software for the automated extraction of trajectory data and the subsequent processing of this data. In particular, a method for classifying road users as pedestrians, motor vehicles or bicyclists based on their spatial progression through the road geometry and their dynamic characteristics is developed and implemented. In addition, a method for post processing trajectory data to rectify distortion caused by collecting video data with a wide-angle lens by applying available image processing tools is proposed.

3 The development and evaluation of two trajectory clustering approaches:

- A method for generic pathway clustering, which is used to identify the general types of pathways used by bicyclists to carry out their desired manoeuvres (left turn, right turn, straight).
- A method for approach specific pathway clustering, which identifies representative pathways used at a given approach of a given intersection.

4 The specification, calibration and validation of models describing the tactical and operational behaviour of bicyclists, including:

- A model of the operational behaviour of bicyclists based on the NOMAD social force model for pedestrian movement. Adaptations and extensions enable the explicit restriction in the change in direction and speed at each time step. A new parameter, the velocity direction factor, which accounts for the relative velocity of interacting road users, is added and is found to improve the simulation of situations in which bicyclists follow one another.
- Logistic regression models to predict four tactical behaviours of bicyclists; infrastructure selection (bicycle facility, roadway or sidewalk), direction of travel (with or against the mandatory direction), compliance with traffic signals and the type of left turn (direct, indirect or indirect against the given direction of travel).

5 The integration of the developed operational and tactical models with the microscopic traffic simulation software SUMO using the interface *TraCI*. This integration allows for the evaluation of the external models within a complete traffic environment.

# 2. Literature review

Modelling the behaviour of bicyclists at signalised intersections requires an extensive understanding of two main domains; the behaviour of bicyclists and the modelling of human behaviour within the context of road traffic. In his critical review of driver behaviour models, MICHON [1985] outlined three levels of driver control that are useful for classifying road user behaviour based on the time horizon of the decision and the type of action (Fig 2.1). At each behavioural level, road users are assumed to be well-informed, rational decision makers who collect and process situational input, predict the risks and benefits associated with each possible action and select the action with the highest expected utility.

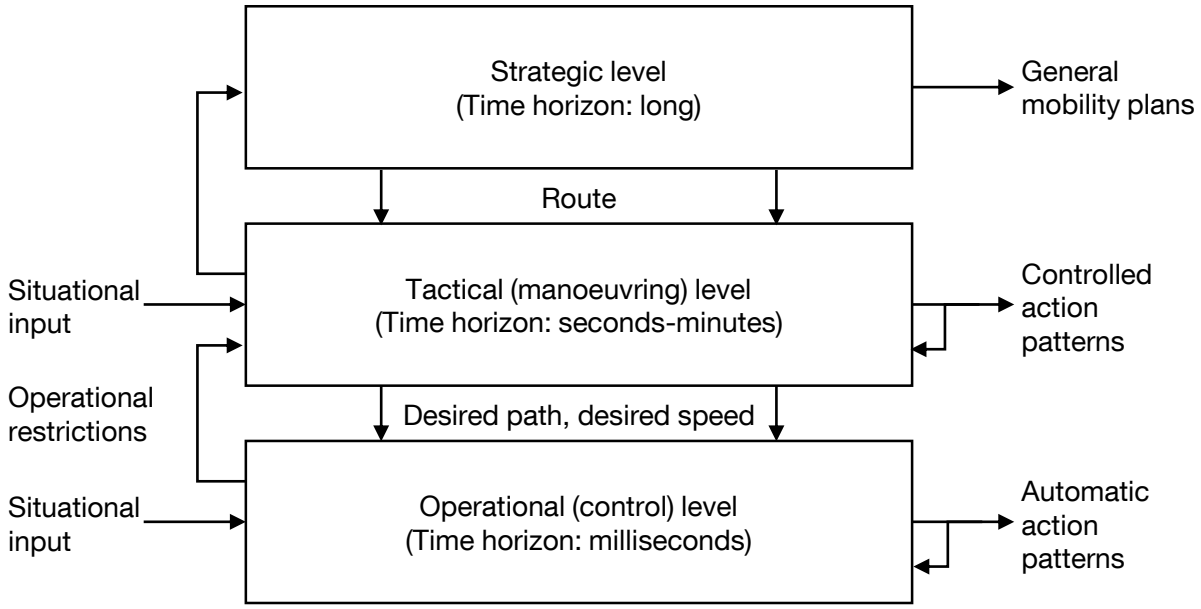



**Fig 2.1** Categorization framework for road user behaviour (adapted from MICHON [1985])

Feedback mechanisms between the levels of the framework reflect the criteria and restraints resulting from decisions made at neighbouring levels. For example, the decision to stop at a red light (tactical level) is restrained by general mobility choices and route requirements established at the strategic level. It is also influenced by the current speed of the bicyclist and necessity to avoid other road users, both of which are determined at the operational level. Once the choice to stop at a red light is made, the outcome will govern action patterns at the operational level and influence future strategic decisions. Generally, task levels higher up in the hierarchy have both an activity initiating and a supervising and correcting function to lower task levels [VAN DER MOLEN & BÖTTICHER, 1988]. Lower task levels influence the choices made at task levels higher up in the hierarchy. The framework suggested by MICHON [1985] and extended by VAN DER MOLEN & BÖTTICHER [1988] is used to organise the review of the literature concerning bicyclist behaviour and modelling approaches. Later in this dissertation, the same framework is used to guide model development.

As defined in Section 1.2, the scope of this dissertation is limited to the operational and tactical behaviour of bicyclists. The framework shown in Fig 2.2 is used to facilitate a comprehensive and concerted literature review with regard to the research question. Specific aspects of the tactical and operational behaviour, which are shown as bullet points in Fig 2.2, are identified for review based on a preliminary investigation of the literature in the field. For each of the aspects, general findings, influential factors, safety effects and efficiency effects are explored. A review of existing approaches for modelling and simulating bicycle behaviour is carried out for tactical and operational behaviour. Three important model types for each behavioural level were identified in a preliminary review and are examined here in detail.

| **2.1 Behavioural review** | |
|---|---|
| 2.1.1 Tactical behaviour | 2.1.2 Operational behaviour |
| • Reaction to red signal<br>• Infrastructure selection<br>• Riding direction<br>• Path selection | • Speed<br>• Acceleration<br>• Spacing, density and flow<br>• Overtaking and meeting<br>• Gap acceptance<br>• Position |

| **2.2 Modelling review** | |
|---|---|
| 2.2.1 Tactical models | 2.2.2 Operational models |
| • Probabilistic approaches<br>• Discrete choice models<br>• Pathfinding approaches | • Cellular automata<br>• Car following models<br>• Social force models |

**Fig 2.2** Literature review framework

Finally, the findings from both the behavioural review and the modelling review as well as the research motivation and scope defined in Section 1 are examined together to determine which behavioural analyses and model developments are necessary to enable the accurate microscopic simulation of bicycle traffic. This workflow is depicted in Fig 2.3.

In Section 2.1, the findings of studies investigating the operational and tactical behaviour of bicyclists are presented according to the framework shown in Fig 2.2. The most important findings are summarised at the end of Section 2.1.1 and Section 2.1.2. Literature pertaining to modelling human behaviour in road traffic is thoroughly examined and summarised in

Section 2.2. Finally, in Section 2.3, a comparison of the findings is given along with the identified areas for further analysis and model development.

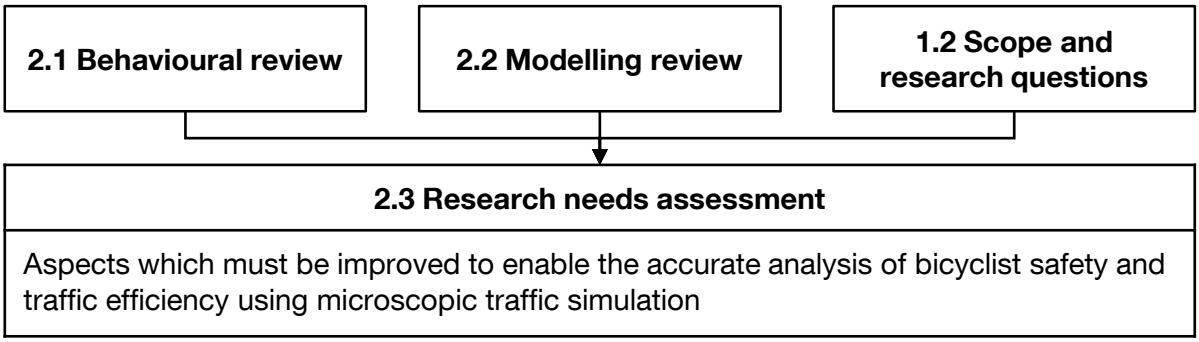



**Fig 2.3** Workflow for the research needs assessment

## 2.1 Behavioural review

Road user behaviour describes the decisions made and action patterns realised by motorists, bicyclists and pedestrians to fulfil goals. In Section 2.1.1, the findings from literature pertaining to the tactical behaviour of bicyclists are presented. Literature concerning the operational behaviour of bicyclists is reviewed in Section 2.1.2. In both sections, international literature describing bicyclist behaviour is summarised. Although the behaviour of bicyclists likely differs from country to country (and even between cities in one country), the relatively small number of studies that have examined the operational and tactical behaviour of bicyclists necessitates a broad-based review. Furthermore, the small number of studies makes it difficult to identify empirically quantified regional differences in bicyclist behaviour. For this reason, regional differences are not further investigated or specified in this review. However, this could be an important point to address in the future when more studies in this field have been published.

### 2.1.1 Tactical behaviour of bicyclists

The tactical behaviour of bicyclists, which includes conscious decisions made on a time horizon of seconds to minutes to cope with the immediate traffic situation, is an important and complex subject. Due to their size, manoeuvrability and ability to adapt to riding with motor vehicles on the roadway or with pedestrians on the sidewalk, a bicyclist is faced with a number of tactical decisions that are not faced by other types of road users. It is imperative to include these tactical behaviours in microscopic traffic simulations in consideration of the significant consequences these behaviours have on overall traffic safety and efficiency. Four particularly relevant tactical behaviours are investigated in this review; the reaction to a red traffic signal, the choice between using a bicycling facility, the roadway or the sidewalk, the direction of travel and the selection of a pathway across an intersection. In this section, the findings of research investigating these behaviours are summarised. Although there are other

tactical behaviours that may be interesting to investigate, these four aspects were found to be most relevant for traffic safety and efficiency at signalised intersections in a preliminary review. The review is separated into four sections corresponding to the four tactical decisions. Relationships identified in previous studies between each behaviour and traffic safety and efficiency are presented.

**Reaction to red traffic signal**

The propensity of bicyclists to stop at a red light has received a significant amount of attention from researchers. The resulting estimated compliance rates identified in these studies ranges greatly. In their review of 16 international studies pertaining to red light violations of bicyclists, RICHARDSON & CAULFIELD [2015] found percentages of red light violation (or self-reported violators for surveys) to range from 6.9% to 87.5%.

Researchers collected revealed choice data at intersections and developed choice models that use the personal characteristics of the bicyclists and the situational parameters to predict red light violation [JOHNSON ET AL., 2011; PAI & JOU, 2014; RICHARDSON & CAULFIELD, 2015]. To supplement the observed data, JOHNSON ET AL. [2013] carried out a survey with the goal of understanding why bicyclists run red lights. The factors found to influence the decision to stop at a red light in these studies are summarised below:

- **Personal characteristics of the bicyclist:** All four studies found that male bicyclists are more likely than female bicyclists to violate red lights. RICHARDSON & CAULFIELD [2015] found gender to be the second most significant determinant of red light infringements (after type of infrastructure). The age of the bicyclist was also found to have a significant effect on the likelihood of red light violations with younger bicyclists disobeying red lights more often than older bicyclists [JOHNSON ET AL., 2013]. Other characteristics of the bicyclist, such as the type of bicycle or clothing, were not found to be related to red light violations [JOHNSON ET AL., 2011]. JOHNSON ET AL. [2013] found that single or never married bicyclists are most likely to violate red lights and students are more likely than employed bicyclists to violate a red light. No significant correlation was found between the level of education or the income of the bicyclist and the propensity to violate a red light. Personality attributes, including aggressiveness and the tendency to violate traffic rules and regulations overall, are likely to be related to red light behaviour specifically. Bicyclists who have been charged with violating a red light while driving a car were found to be more likely to do so while bicycling than those who had not [JOHNSON ET AL., 2013]. PAI & JOU [2014] found wearing a helmet to be negatively correlated with disobeying at a red light, however, JOHNSON ET AL. [2011] did not find a significant relationship between these variables. Bicyclists who have experienced an accident while bicycling are more likely to violate a red light than those who have not [JOHNSON ET AL., 2013].

- **Type of infrastructure:** Bicyclists using a 'cycle track' ("segregated cycling facility with bicycle lights that give cyclists right of way when they interact with other traffic" [RICHARDSON & CAULFIELD, 2015, P.67]) are much more likely to infringe on a red light than those using other types of bicycle lanes. At intersections, bicyclists were found to violate a red light slightly more often if an advanced stop line for bicyclists is provided [ALLEN ET AL., 2005]. Similarly, JOHNSON ET AL. [2011] found that bicyclists on a 'centre lane' (mixed traffic road with a bicycle stop box ahead of the stopped motor vehicle traffic) are more likely than bicyclists using other types of facilities to infringe on red lights. However, ZANGENEHPOUR ET AL. [2013] found that bicycle boxes significantly reduce the number of red light violations.

- **Traffic characteristics:** The traffic flow at the intersection plays an important role in the red light compliance of bicyclists. Low traffic volumes on the crossing road lead to red light infringing behaviour [PAI & JOU, 2014; JOHNSON ET AL., 2013]. The presence of other road users on the same approach, either cars or other bicyclists, was found to have a deterring effect on red light violations [JOHNSON ET AL., 2011].

- **Manoeuvre:** The desired manoeuvre of the bicyclist (straight, right turn or left turn) was found to be related to red light violations. Bicyclists turning left in Australia (akin to turning right in countries with right-hand traffic) violate red lights more often than those carrying out other manoeuvres [JOHNSON ET AL., 2013; JOHNSON ET AL., 2011].

- **Signal control:** The length of the signal phase was found to be positively correlated with opportune infringing of a red light [PAI & JOU, 2014].

- **Time of day and weather:** PAI & JOU [2014] found that bicyclists are more likely to violate red lights in off-peak hours and in good weather conditions.

According to the literature, the safety risk associated with bicyclists violating red lights is relatively low. BACCHIERI ET AL. [2010] studied a group of bicyclists over one year and did not find a significant correlation between reported red signal violation and accidents. LAWSON [1991] found that red light violation on the part of a bicyclist caused 1.8% of collisions. GERSTENBERGER [2015] did not report red light violation on the part of a bicyclist to be a considerable cause of accidents in Germany. In an analysis of Australian crash data, researchers found that traffic light violation on the part of a bicyclist caused 6.5% of accidents caused by bicyclists, which is 2.3% of all accidents involving at least one bicyclist [SCHRAMM ET AL., 2008].

No studies were found that measure the effect of red light violations by bicyclists on the traffic flow at intersections. It is hypothesised that red light violations carried out opportunistically by bicyclists in situations with little or no traffic on the crossing roadway improve the overall efficiency at intersections by reducing waiting times for interacting traffic streams.

**Infrastructure selection**

Unlike almost any other type of road user, bicyclists are able to make tactical decisions about where to travel. Although there are traffic regulations in some regions that mandate the use of available bicycle infrastructure, bicyclists still tend to choose between using a bicycle facility, riding with motor vehicles on the roadway or with pedestrians on the sidewalk. This choice has a considerable impact on the overall efficiency of traffic flow as well as on the safety of bicyclists and other road users. Unfortunately, little research has been done to examine infrastructure selection and the motivation of bicyclists to select the roadway or sidewalk when a bicycle facility is available. A number of studies have investigated the role of infrastructure within the framework of route choice. These studies, however, offer little insight as to which part of the infrastructure is actually used by the bicyclist once they are on-route.

KULLER ET AL. [1986] carried out an observational study to determine how often bicyclists use the roadway or sidewalk when a bicycle facility is available. Interviews were carried out immediately after the behaviours were observed to uncover motivating factors. They found that bicyclists use the roadway when a bicycle facility is available for three main reasons; the bicycle facility is blocked by an obstacle or road user, the bicyclist is preparing for an anticipated manoeuvre, such as turning left, or the bicyclist is not satisfied with the quality of the bicycle facility. Bicyclists who use the sidewalk instead of a bicycle facility also cited obstacle avoidance, the anticipation of upcoming manoeuvre and the quality of the bicycle facility. In addition, they also referenced other rule-breaking behaviours, such as riding two abreast and riding against the given direction or travel, the desire to ride slowly and a feeling of safety as reasons for riding on the sidewalk.

ALRUTZ ET AL. [2009] carried out an in-depth study of the accident risk and rule acceptance of bicyclists in Germany and found that 90% of bicyclists use a provided bicycle facility regardless of type. In cases where bicycle facility use is mandatory, 92% use the facility, while 87% do so if bicycle facility use is not mandatory.

Three studies were found that investigate factors motivating bicyclists' choice between using different parts of the infrastructure [KULLER ET AL., 1986; ALRUTZ ET AL., 2009; GUO ET AL., 2013]. The factors identified by these studies are summarised below.

- **Personal characteristics of the bicyclist:** Male bicyclists and younger bicyclists are more likely to ride on the roadway than female and older bicyclists [ALRUTZ ET AL., 2009]. KULLER ET AL. [1986] noted that women and people under 18 years of age are more likely to use the sidewalk than other bicyclists.

- **Infrastructure characteristics:** The type of bicycle infrastructure available (e.g. physically separated bicycle paths, bicycle lanes on the roadway or marked roadways) has a strong influence on whether bicyclists will choose to use this facility. GUO ET AL.

[2013] studied the tendency of bicyclists to temporarily cross from a bicycle lane into the roadway and found that the width of the bicycle lane is negatively correlated with the probability of moving out of the bicycle lane. ALRUTZ ET AL. [2009] found that narrow bicycle lanes (<1 m wide) and on-road bicycle lanes have a lower rate of acceptance than other types of bicycle infrastructure.

- **Speed:** The average speed of the bicyclist was found by GUO ET AL. [2013] to have a small, significant correlation with bicyclists moving from a bicycle lane into the roadway.

- **Parking:** The presence of kerbside parking along an on-road bicycle lane has a positive correlation with bicyclists moving from a bicycle lane to the roadway [ALRUTZ ET AL., 2009].

- **Traffic characteristics:** Neither the volume of car traffic in the roadway nor the presence of a moving car in the adjacent driving lane are significant factors in the decision of a bicyclist to veer out of a bicycle lane into the roadway [GUO ET AL., 2013]. ALRUTZ ET AL. [2009] also found traffic volume does not play a role in the decision of a bicyclist to use a bicycle lane or not. The presence of another bicyclist riding against the given direction of travel, however, was found to have a strong positive correlation with moving from the bicycle lane into the roadway [GUO ET AL., 2013]. KULLER ET AL. [1986] found that bicyclists who rode on the sidewalk are likely to cite high traffic volume and speed as a reason for riding on the sidewalk rather than the roadway when no bicycle facility is available.

Infrastructure selection appears to play a crucial role in bicyclists' safety, although few studies have investigated this link. In his analysis of accidents in Germany, GERSTENBERGER [2015] found that "bicyclist used the wrong infrastructure" was the stated accident cause in nearly half of the cases in which a bicyclist was responsible for an accident. Similarly, WACHTEL & LEWISTON [1994] found that the average bicyclist faces a collision risk on the sidewalk (including separated bicycle paths) 1.8 times as great as on the roadway (including bicycle lanes on the roadway). A significant weakness of this study, however, is the coarse clustering of sidewalks and roadways with and without a bicycle facility.

A number of studies have investigated the type of infrastructure available for bicyclists and the risk of injury [TESCHKE ET AL., 2012; RODGERS, 1995; RIVARA ET AL., 1997; REYNOLDS ET AL., 2009; MORITZ, 1998; LUSK ET AL., 2011; AULTMAN-HALL & HALL, 1998]. In a review of 23 papers that studied the role played by the type of infrastructure in bicyclist safety, REYNOLDS ET AL. [2009, P.1] concluded that "sidewalks and multi-use trails pose the highest risk, major roads are more hazardous than minor roads, and the presence of bicycle facilities (e.g. on-road bike routes, on-road marked bike lanes, and off-road bike paths) was associated with the lowest risk". Unfortunately, these studies are only marginally useful in determining the influence of tactical infrastructure selection because bicyclists may not use the intended facility. For

example, on busy roads, bicyclists may use the sidewalk instead of the roadway to avoid situations that make them feel unsafe [KULLER ET AL., 1986].

The effect of bicyclists' tactical infrastructure selection on the traffic efficiency was not investigated directly in any studies found in this literature review. Due to their slower travelling speed, bicyclists who ride on the roadway can hinder the movement of motor vehicles driving in the same lane. The extent of this effect is presumably linked to the width and number of driving lanes, the volume of both bicycle and motor vehicle traffic in both directions and the portion of bicyclists who choose to use the roadway. Research has shown that bicyclists tend to move out of the bicycle facility and onto the sidewalk or roadway when passing a slower moving bicyclist in the bicycle facility [FALKENBERG ET AL., 2003]. While potentially degrading the traffic flow of motor vehicles in the adjacent road lane or the movement of pedestrians on the sidewalk, this behaviour likely improves the overall flow of bicycle traffic as faster moving bicyclists are able to pass slower individuals. To the knowledge of the author, neither of these premises have been quantified using observed data.

**Riding direction**

Most roads are either unidirectional or bidirectional and are comprised of a number of driving lanes, each characterised by a mandatory direction of travel. In regions with right-hand traffic, road users proceed along the right-hand side of the roadway and a dividing line on bidirectional roads indicates the width of the directional right of way. Similarly, bicycle facilities can be classified into two categories with regard to the direction of travel; unidirectional bicycle facilities, which are normally positioned on the respective side of the roadway, and bidirectional bicycle facilities, which are positioned on one side of the road and are used by bicyclists riding in both directions.

Bicyclists, more so than motor vehicles, can tactically decide to ride with or against the mandatory direction of travel. Despite the fact that this behaviour can lead to dangerous situations in which the bicyclist approaches an intersection from an unexpected direction, KULLER ET AL. [1986] found that 11.1% of bicyclists ride against the mandatory direction on a bicycle facility, 1.0% do so on the roadway and 29.0% ride the wrong way on the sidewalk. In the United States, WACHTEL & LEWISTON [1994] found that 14.2% of bicyclists ride against the given direction of travel (5.3% on the roadway and 32.4% on the sidewalk).

Two main reasons for riding against the given direction of travel were identified by KULLER ET AL. [1986] through interviews with bicyclists. First, bicyclists are more likely to violate traffic rules upon approaching their final destination or an intermediate goal. Consequently, many bicyclists ride the last few meters of the trip against the mandatory direction of travel. The second reason for riding against the mandatory direction of travel is route simplification.

Although the direction of travel plays a key role in the likelihood of a bicycle-motor vehicle collision occurring, only two research papers were found that investigate the factors associated with this behaviour [ALRUTZ ET AL., 2009; KULLER ET AL., 1986]. The findings of these studies are summarised below:

- **Personal characteristics of the bicyclist:** ALRUTZ ET AL. [2009] found that men are more likely to ride against the mandatory direction of travel than women. They also found that children ride against the given direction of travel more often than adults. KULLER ET AL. [1986], however, found riding against the mandatory direction of travel to occur with equal frequency for all bicyclists, regardless of their personal characteristics.

- **Infrastructure characteristics:** No information was found concerning the influence of the infrastructure design on the likelihood of bicyclists riding against the mandatory direction of travel. Findings from ALRUTZ ET AL. [2009] suggest that if a bicyclist rides against the mandatory direction of travel, there is a 66% likelihood he or she will use a bicycle facility and a 33% chance of using the sidewalk. If the bicycle facility is situated in the roadway, the portion of wrong way bicyclists using the sidewalk increases to 80%.

The direction of travel is one of the key factors influencing the safety of bicyclists, particularly at intersections. A situation that dominates bicycle crash statistics in many countries occurs at non-signalised intersections when a vehicle turning right collides with a bicycle approaching from the right side who is travelling against the given direction of travel [HERSLUND & JØRGENSEN, 2003; GERSTENBERGER, 2015; RÄSÄNEN & SUMMALA, 1998; SUMMALA ET AL., 1996]. In this case, drivers expect conflicting vehicles or bicycles to approach from the left side, and as a consequence, scan the area to their left more thoroughly than the area to the right [SUMMALA ET AL., 1996]. Even if drivers do look to the right, they are more susceptible to a “look-but-failed-to-see-error” because the position of the bicycle does not belong to the driver’s fixed search strategy [HERSLUND & JØRGENSEN, 2003]. GERSTENBERGER [2015] found that more than half of the bicycle collisions recorded in the GIDAS (German In-Depth Accident Study) database occurred when a bicyclist approached from an unexpected direction. The accident situations with bicyclists riding against the expected direction of travel identified by GERSTENBERGER [2015] are shown in Fig 2.4. The bicycle location in the diagram does not indicate the type of bicycle facility used (on road or separated).

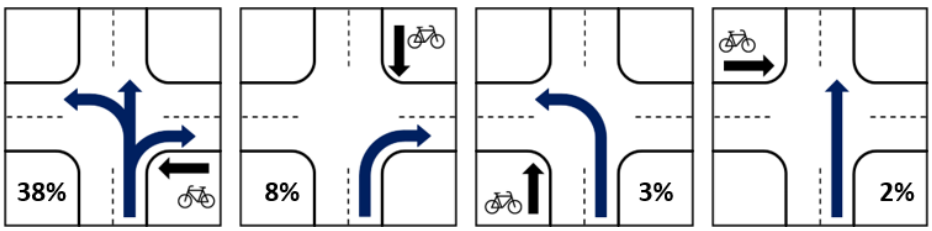



**Fig 2.4** Accident situations in which bicyclists approach from an unexpected direction and the percentage of bicycle-vehicle accidents (adapted from GERSTENBERGER [2015])

In general, bicyclists riding against the mandatory direction of travel were found by WACHTEL & LEWISTON [1994] to face an increased risk of a collision with a motor vehicle of 3.6 times compared to those riding in the expected direction of travel. This is particularly true for bicyclists riding on the sidewalk, in which case a risk increase of 4.5 times was estimated.

The effect of riding against the mandatory direction of travel on traffic flow was not analysed in any studies uncovered in the course of this literature review. Due to the infrequency of bicyclists riding the wrong way on the roadway [KULLER ET AL., 1986], the behaviour does not likely have a significant influence on the flow of motor vehicle traffic. The flow of bicycle traffic on bicycle facilities, however, will be affected by a bicyclist riding against the direction of travel. According to the theory formulated by BOTMA [1995] and implemented in the American Highway Capacity Manual [NATIONAL RESEARCH COUNCIL, 2000; NATIONAL RESEARCH COUNCIL, 2010] and the German design guidelines "*Handbuch für die Bemessung von Straßenverkehrsanlagen*" [FORSCHUNGSGESELLSCHAFT FÜR STRAßEN- UND VERKEHRSWESEN, 2015], the level of service on bicycle facilities is influenced by hindrance caused by meeting and passing events. When a bicyclist rides against the mandatory direction of travel, he or she creates meeting events for all other bicyclists using the facility in the expected direction.

**Path Selection**

The path of a bicyclist is understood here as movement through space with the purpose of realising a given manoeuvre. In the context of this work, path selection is limited to signalised intersections. The corresponding manoeuvres include turning right, turning left and riding straight across the intersection. An interesting example of this tactical behaviour is the path selection of bicyclists turning left. There are three main possibilities for bicyclists to carry out this manoeuvre, depicted in Fig 2.5.

1. Direct left turn – bicyclists turn with the motor vehicle traffic in one signal phase.
2. Indirect left turn – bicyclists turn over two phases using a pedestrian style turn.
3. Indirect left turn (wrong way) – similar to the indirect left turn but the bicyclist rides against the mandatory direction of travel during both of the turning phases.

The possible number of pathway options for bicyclists turning left explodes when combinations of infrastructure use (bicycle facility, roadway or sidewalk), the direction of travel (with or against the mandatory direction) and the left turn type are considered. Although less pronounced than in the case of bicyclists turning left, bicyclists turning right and those riding straight across the intersections also face choices concerning where and how to carry out their manoeuvre.

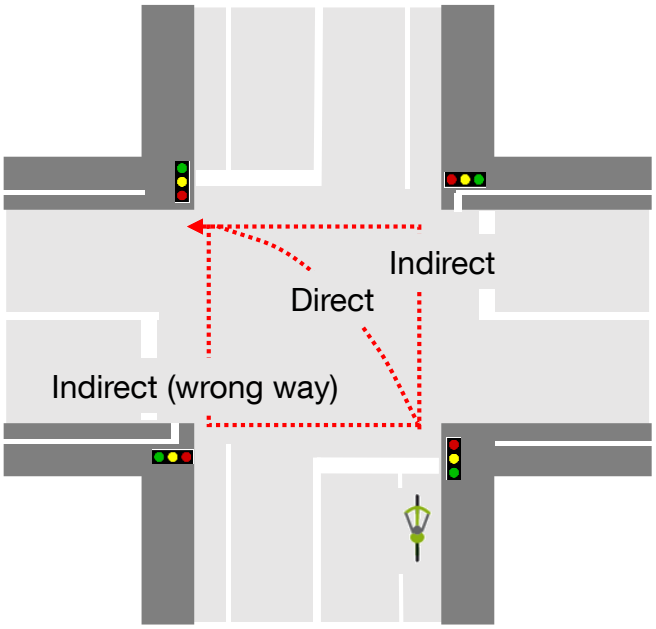



**Fig 2.5** Depiction of the three main types of left-hand turns for bicyclists

Although the pathfinding behaviour of bicyclists at intersections has important consequences concerning the overall safety and efficiency at intersections, very little research was found that examines this behaviour. AMINI, TWADDLE & LEONHARDT [2016] used revealed preference data collected at three intersections to examine the tactical path selection of bicyclists turning left. LING & WU [2009] developed a model for predicting the paths of bicyclists at signalised intersections. Unfortunately, the data used to validate the model and the model predictions were only briefly explained in the paper and no conclusions could be extracted for this review. The factors that influence the path selection of bicyclists turning left identified by AMINI, TWADDLE & LEONHARDT [2016] are summarised below.

- **Signal control:** Bicyclists using a bicycle facility become less likely to make an indirect left turn against the mandatory direction of travel as the time since the red phase began increases. The green phase of the adjacent pedestrian signal increases the likelihood for an indirect left turn against the direction of travel.

- **Infrastructure selection:** Path selection is strongly dependent on the infrastructure used upon arrival. Two discrete choice models were developed to predict the behaviour of bicyclists, one for those arriving on the roadway and another for those arriving on the bicycle facility or sidewalk, both during the red phase. Bicyclists using a bicycle facility or sidewalk either make an indirect left turn with or against the given direction of travel and not a direct left turn. Bicyclists using the roadway, on the other hand, carry out either a direct left turn or an indirect left turn against the direction of travel.

No studies were found that examined the effect of path selection across an intersection and the overall traffic safety or efficiency. However, an effect concerning safety is expected due

to differing patterns of interaction between bicyclists and other road users. In addition, similar safety consequences as described in the riding direction section are expected for bicyclists who make indirect left turns against the direction of travel.

With regard to traffic efficiency, the rate of indirect left turns against the direction of travel is expected to influence the flow of traffic at intersections. Similar to traffic signal violation, queue building at red signals is reduced by bicyclists who turn indirectly against the direction of travel. If bicycle traffic shares the road with motor vehicles, this can also reduce the effect of bicycle queues hindering vehicles at intersections.

**Summary**

A short overview of the tactical behaviours examined in Section 2.1.1 is presented in Tab. 2.1. Each of the four reviewed aspects is assessed with regard to the current state of knowledge and the influence of the behaviour on the safety of bicyclists and the overall traffic efficiency. The current state of knowledge is evaluated using a three-category ordinal scale (poor, fair, good). The influence of the behaviour on the safety of bicyclists and the overall traffic efficiency is rated using a second three-category ordinal scale (small, moderate, large).

| Aspect | Assessment |
|---|---|
| **Reaction to red signal** | **General knowledge:** More than 16 studies were found that examine the reaction to red traffic signals and identify the motivating factors for violation. **Good**<br>**Safety effects:** The results of studies investigating the relationship between red light violations and accident rates indicate a surprisingly small relationship. **Small**<br>**Efficiency effects:** Although no studies were found that investigated the relationship between red light violations and efficiency, it is hypothesised that bicyclists who violate red lights increase the capacity of signalised intersections, as they do not hinder other road users when the signal turns green. **Moderate** |
| **Infrastructure selection** | **General knowledge:** Although a number of studies used route choice as a surrogate measure of infrastructure preference, these studies provide limited insight into infrastructure selection while using a given road segment. Only three studies were found that examined tactical selection of infrastructure. **Poor**<br>**Safety effects:** Although few studies have explored the link between infrastructure selection and bicyclist safety, the relationship appears significant. **Large**<br>**Efficiency effects:** Overall traffic efficiency is highly dependent on where bicyclists choose to ride. Hindrance to motor vehicles and pedestrians can occur when bicyclists select the sidewalk or roadway instead of a bicycle facility. **Large** |
| **Riding direction** | **General knowledge:** Few studies investigated the probability and motivation for riding against the mandatory direction of travel. **Poor**<br>**Safety effects:** Bicyclist approaching from an unexpected direction is arguably the most important factor leading to collisions involving bicyclists. The direction of travel is crucial to consider in evaluating bicyclist safety. **Large**<br>**Efficiency effects:** Due to the infrequency of bicyclists riding against the mandatory direction of travel on the road, this is unimportant for vehicular traffic. Pedestrian and bicyclist level of service (LOS) and flow are affected. **Moderate** |
| **Path selection** | **General knowledge:** Only two studies were identified regarding the path selection of bicyclists at intersections. There is great potential to build on the preliminary results offered by these papers. **Poor**<br>**Safety effects:** Although the role of this aspect in bicyclist safety was not investigated in any studies uncovered in the literature review, it is hypothesised that the effect is similar to riding against the mandatory direction of travel. **Large**<br>**Efficiency effects:** The path selection across the intersection is expected to influence the overall flow of traffic by influencing the position of bicyclists and the queuing of bicyclists at red signals. **Moderate** |

**Tab. 2.1** Assessment of the current state of knowledge and traffic safety and efficiency effects of the analysed aspects of tactical behaviour

### 2.1.2 Operational behaviour of bicyclists

The operational behaviour of road users is defined as subconscious action patterns that take place on a time horizon of milliseconds to seconds. Actions at this level are carried out to respond to the immediate situation while executing plans made at the tactical level and have important consequences concerning traffic flow and safety. The main differences between the operational behaviour of bicyclists and that of motor vehicles are the roles of physical and legal restraints. Whereas the speed travelled by motor vehicles is constrained by the speed limit (and the willingness to follow this limit), the speed travelled by bicyclists is controlled by many additional parameters, such as the preferences, physical fitness and skill of the bicyclist, the type and quality of the infrastructure, the road gradient and the weather. Similarly, the lateral spacing of motor vehicles is determined mainly by the width of the lanes. Bicyclists, on the other hand, are free to select their position within the infrastructure and their spacing to other road users. Six aspects of the operational behaviour of bicyclists at signalised intersections are selected for review: speed, acceleration, spacing between bicyclists and the resulting density and flow of bicycle traffic, overtaking and meeting events, gap acceptance and lateral position.

#### Speed

Most of the research that has been done to date concerning bicyclist speed has been motivated by the need to consider bicycle traffic in signal control plans [PEIN, 1997; RUBINS & HANDY, 2005; SHLADOVER ET AL., 2011; TAYLOR & DAVIS, 1999]. PEIN [1997] measured the speed and crossing time of bicyclists who started from a complete stop at a red light. The average crossing speed was found to be 4.3 m/s. RUBINS & HANDY [2005] measured crossing times at 10 intersections in the USA and calculated the average speed of bicyclists starting from a standing, rolling or quasi-rolling state. They found an average crossing speed of 4.1 m/s for all bicyclists, 3.4 m/s for bicyclists starting from a standing state, 4.9 m/s for bicyclists starting from a rolling state and 3.6 m/s for bicyclists starting from a quasi-rolling state. The researchers noted a very large range in the crossing times and calculated crossing speeds, even among relatively homogeneous populations of bicyclists. SHLADOVER ET AL. [2011] analysed the trajectories of bicyclists crossing two signalised intersections in the USA to derive the average crossing speed. They found an average of 5.0 m/s with a large difference between the average speeds found at each of the two intersections. WU, LING & ZHAO [2004] measured speeds at signalised intersections in China and found an average of 3.2 m/s.

Other studies focus on the level of service for bicycle traffic. KHAN & RAKSUNTORN [2001] analysed overtaking events in the USA to quantify the effects of hindrance events on speed. Hindrance occurs when a bicyclist passes another bicyclist travelling in the same direction, meets a bicyclist travelling in the opposite direction or a combination of both events. They found an average speed of 7.5 m/s for overtaking bicyclists and 4.9 m/s for bicyclists being overtaken. FALKENBERG ET AL. [2003] used video data collected at 21 road segments in

Germany to analyse the free flow speeds and the speed and spacing during overtaking and other types of hindrance events. They found an average free flow riding speed of 4.7 m/s, an average speed of 5.5 m/s while overtaking and an average speed of 3.6 m/s during hindrance events.

The results of several speed studies were used to calibrate microscopic traffic simulations of bicycle traffic. RAKSUNTORN [2002] analysed the behaviour of bicyclists at five signalised intersections in the USA and found an average speed of 4.6 m/s. DE ZHAO ET AL. [2013] measured the operational behaviour of conventional bicycles and e-bikes in order to develop a calibrated cellular automaton simulation and found an average speed of 4.4 m/s.

The free flow speed travelled by bicyclists is dependent on several factors, the role of which has been investigated by a number of researchers. The findings of these studies are briefly described below.

- **Type of bicycle:** DE ZHAO ET AL. [2013] found an average speed of e-bicyclists of 6.2 m/s and an average speed of conventional bicyclists of 4.4 m/s. SCHLEINITZ ET AL. [2016] equipped the bicycles of 85 participants (conventional bicycles, pedelecs and e-bikes) with sensors to measure speed and distance as well as two cameras to record the environment and the face of the bicyclist. They found an average speed of 4.3 m/s, 4.8 m/s and 6.8 m/s for conventional bicycles, pedelecs and e-bikes, respectively[1]. No studies were found that compared the speed of bicyclists using different types of conventional bicycles.

- **Personal characteristics of the bicyclist:** There is little agreement in the findings of studies that investigate the effects of the age of the bicyclist on speed. While PARKIN & ROTHERAM [2010] found no significant difference in average bicycling speed based on the age of the bicyclist, SCHLEINITZ ET AL. [2016] performed a pairwise comparison of three age groups (≤ 40 years, 41-64 years and ≥65 years) and found that the younger group rides significantly faster than the older group in all cases. THOMPSON ET AL. [1997] analysed the speed of 152 recreational bicyclists divided into two groups (<14 years and ≥14 years). They found a small, statistically insignificant difference between the average speeds of the two groups (4.0 m/s and 4.3 m/s, respectively). FALKENBERG ET AL. [2003] created a linear regression model to predict speed which included age as an ordinal variable with four categories (<18 years, 18-30 years, 31-60 years and >60 years) and found a decrease in the average speed of 0.4 m/s for each subsequent age category.

  This disagreement is even more pronounced concerning the role of gender. LING & WU [2004] found that women ride slightly faster than men, although this difference was found

[1] Pedelecs offer electric motor support up to 25 km/h while e-bikes offer support up to 45 km/h.

to be small (0.1 m/s) and insignificant. PARKIN & ROTHERAM [2010] found a small, significant difference in the speed of male and female bicyclists, though fewer data were collected from female bicyclists. They reported average speeds of 5.7 m/s and 6.1 m/s for female and male bicyclists, respectively. However, they found that including gender in the regression model for speed prediction had no significant influence. THOMPSON ET AL. [1997] found the average speed of male and female bicyclists to be 4.3 m/s and 4.0 m/s, respectively but did not find this difference to be statistically significant. WHEELER, CONRAD & FIGLIOZZI [2010] found a small, significant difference between the average crossing speeds of male and female bicyclists at two intersections in the USA, one flat and one at a gradient, with men riding significantly faster than women. The linear regression model developed by FALKENBERG ET AL. [2003] indicates that women ride on average 0.6 m/s slower than men.

In the only study found that investigated physical fitness and bicycling speed, PARKIN & ROTHERAM [2010] found no significant relationship between the Body Mass Index (BMI), which was used as an indicator of physical fitness, and the average speed of a bicyclist.

- **Gradient of the roadway:** PARKIN & ROTHERAM [2010] found an average speed on flat segments of 6.0 m/s and an increase in the average speed of 0.24 m/s for each 1% decrease in gradient (downhill) and a decrease in the average speed of 0.40 m/s for each 1% increase in gradient (uphill). FALKENBERG ET AL. [2003] included gradient as an ordinal variable with three categories (downhill, flat and uphill) in their regression model and found an increase and decrease of 0.8 m/s for bicyclists riding downhill and uphill, respectively.

- **Type of infrastructure:** There is a high level of agreement between the few studies that examined the effect of infrastructure type on the speed of bicyclists. Researchers in the USA collected video data at 15 intersections to develop a multimodal intersection simulation [OPIELA ET AL., 1980]. They examined the average approach speed of bicyclists using four different types of infrastructure and found a mean speed on bicycle paths of 5.6 m/s, on bicycle lanes of 6.9 m/s, on sidewalks of 5.1 m/s and on roadways of 5.3 m/s. In a naturalistic study of bicycle speeds, SCHLEINITZ ET AL. [2016] found the highest speeds on bicycle lanes and roadways (4.6 m/s), followed by the sidewalk (3.7 m/s). These speeds are quite a bit lower than those measured by OPIELA ET AL. [1980] but show the same tendency concerning the role of infrastructure type. FALKENBERG ET AL. [2003] found that bicyclists ride the fastest on bicycle lanes on the roadway, followed by separated bicycle lanes and bicycle paths (0.9 m/s decrease between subsequent infrastructure types). They did not measure the speed of bicyclists riding on the sidewalk or on the roadway in mixed traffic conditions. ALRUTZ ET AL. [2009] also found the bicyclists travel the fastest on bicycle lanes on the roadway than on any other type of

road infrastructure. The average speed for each type of infrastructure is included in the meta-analysis presented in Fig 2.6.

- **Manoeuvre:** LING & WU [2004] used video data to measure the average speed of 331 bicyclists in China while turning left and travelling straight across an intersection and found average values of 3.2 m/s and 3.7 m/s, respectively. RAKSUNTORN [2002] found an average crossing speed of 4.6 m/s for 997 bicyclists at five intersections in the USA. Bicyclists turning right were found to have an average speed of 3.1 m/s to 3.6 m/s when the traffic signal is red or green, respectively (N=43).

- **Miscellaneous:** Other factors, such as the speed and direction of the wind and other weather-related aspects, likely influence the speed of bicyclists [TAYLOR & DAVIS, 1999].

A meta-analysis is carried out in which all the reported means from the reviewed studies are aggregated. The average speed result from each study is included once in the dataset (studies are not weighted depending on the number of observations). The boxplots shown in Fig 2.6 portray the results of the meta-analysis.

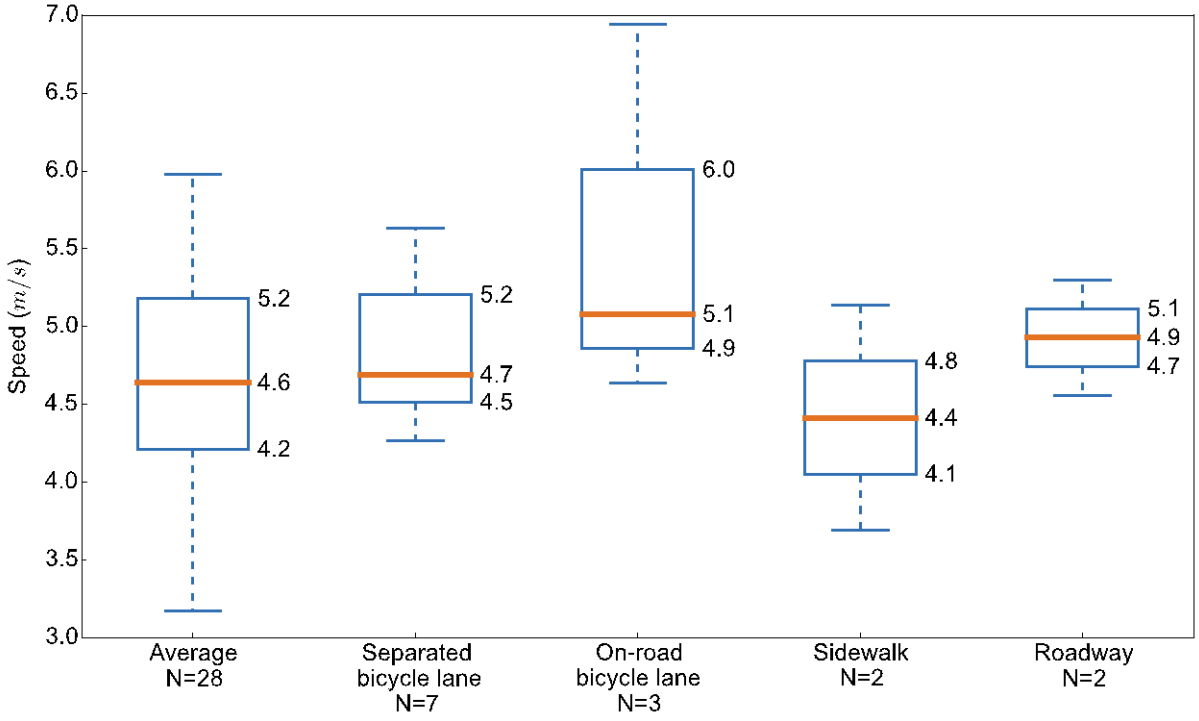



**Fig 2.6** Meta-analysis of the reported average speeds of bicyclists

Speed is an important variable in traffic safety. RIVARA ET AL. [1997] studied the severity of injury of bicyclists involved in an accident and found the speed of the bicyclist, expressed by the dichotomous variable (< 6.7 m/s and ≥ 6.7 m/s), to be associated with a 1.4 times increase in the risk for severe injury, 1.5 times increased risk for hospital admission and 2.6 times increased risk for fatal injury (when speed ≥ 6.7 m/s) while controlling for the speed of the

other road user. Furthermore, the speed variation in a population of bicyclists likely plays a role in difficulties experienced by drivers in estimating gap sizes in conflicting streams of bicyclist traffic [HERSLUND & JØRGENSEN, 2003].

Bicyclist speed is a determinant in overall traffic efficiency at intersections and on-road segments. Intergreen times, the length of which has a strong influence on the capacity of signalised intersections, must be allotted to ensure that all road users can cross the intersection before conflicting streams are allowed to enter the intersection. The maximum clearance time is determined from the distribution of speed and acceleration rates among bicyclists. This aspect has been thoroughly analysed by researchers with widely ranging results. The methodology used to estimate LOS in the Highway Capacity Manual assumes a normally distributed average speed of 5 m/s with a standard deviation of 0.8 m/s [NATIONAL RESEARCH COUNCIL, 2000; GOULD & KARNER, 2010]. No information is given concerning how different speed distributions can be implemented. Speed distribution, however, could be an important parameter to consider in analysing bicyclist traffic, as large speed disparities can cause problems with traffic safety and efficiency. No research was found that has investigated this link.

In mixed traffic, particularly on single lane roads where motor vehicles cannot always overtake slower moving bicyclists, the speed travelled by bicyclists sets the overall travel speed on the roadway until a passing manoeuvre can be carried out. Although a number of papers were found that use microsimulation to analyse this situation [YAO ET AL., 2009; CHENG ET AL., 2008; GUO ET AL., 2013], none measured this effect in the field.

A final application for average speed in traffic engineering is the design of coordinated traffic signals. The progression speed of signal coordination for bicyclists is determined using characteristics of the average speed distribution (mean and standard deviation) of the bicycling population. Even slight errors in the speed estimation can degrade the effectiveness of the coordination. No research was found investigating the effect of different progression speeds for signal coordination and the efficiency of bicycle and overall traffic.

**Acceleration and deceleration**

In relation to the number of studies that investigated the speed, relatively few were found that examine the acceleration and deceleration of bicyclists. It is important here to distinguish between instantaneous acceleration, which is an acceleration value at a particular point in time, an acceleration profile, which is a sequence of acceleration values over time, and mean acceleration, which is the average acceleration of an acceleration profile. LUO & MA [2016] carried out a detailed analysis of the acceleration and deceleration profiles of bicyclists using GPS data from 11 bicyclists in Stockholm, Sweden. They found acceleration rates to fall most often between 0.5-1.0 m/s$^2$. The majority of bicyclists were found to complete an acceleration process from a stopped position within 10 s. The deceleration process was not quantified

precisely in the paper. However, for both acceleration and deceleration, the maximum acceleration or deceleration occurs near the mid-point of the process. PARKIN & ROTHERAM [2010] collected GPS data from 16 bicyclists in the United Kingdom for one week and examined the speed and acceleration characteristics. They found a mean and maximum acceleration of 0.25 m/s$^2$ and 0.71 m/s$^2$, respectively. A linear regression model was estimated to predict the acceleration given the gradient of the roadway. A model constant of 0.23 m/s$^2$ was estimated, which represents the acceleration without any gradient. This estimate leads to an acceleration duration of 26 s from a stop to an ultimate speed of 6.0 m/s, which is significantly longer than the acceleration duration estimated by LUO & MA [2016]. PEIN [1997] measured the acceleration of 442 bicyclists on a bicycle path in the USA and found mean acceleration rates to range between 0.82 m/s$^2$ and 1.07 m/s$^2$. The mean acceleration was estimated from the crossing time and the average a mean cruising speed estimation. In another study in the USA, TAYLOR [1993], measured the mean acceleration of 18 subjects and found the mean acceleration to range between 0.43 m/s$^2$ and 1.15 m/s$^2$. The limited number of factors identified in the literature review to influence the acceleration of bicyclists are summarised below.

- **Personal characteristics of the bicyclist:** FIGLIOZZI, WHEELER, & MONSERE [2013] found no significant difference in the mean acceleration of male and female bicyclists, either at flat or graded intersections. They did note, however, that male bicyclists accelerate for a longer period of time than female bicyclists, reaching a higher speed by the end of the acceleration process. LUO & MA [2016] could not discern any significant difference in acceleration based on gender, although they suggest this is due to the small sample size in the experiment. PARKIN & ROTHERAM [2010] found a small but significant difference in the average acceleration of male and female bicyclists (0.26 m/s$^2$ and 0.22 m/s$^2$, respectively). However, including gender was not a significant predictor in the resulting regression model.

- **Gradient:** FIGLIOZZI, WHEELER, & MONSERE [2013] studied the acceleration profiles of bicyclists at two intersections in the USA, one flat and one at a gradient, and found the largest portion of measured acceleration rates to fall between 1.5-1.8 m/s$^2$ and 0.9-1.2 m/s$^2$, respectively. The mean, median and variation were not quantified in more detail in the paper. Using a linear regression model, PARKIN & ROTHERAM [2010] estimated that each degree of downhill gradient increases the mean acceleration by 0.02 m/s$^2$ and each degree of uphill gradient decreases the mean acceleration by 0.01 m/s$^2$.

No research was found that thoroughly examined the deceleration (negative acceleration) of bicyclists.

Acceleration and deceleration have a significant influence on traffic safety and efficiency. Acceleration from the stop line at intersections influences the overall crossing time and, therefore, the necessary intergreen times. Maximum deceleration rates are important

parameters in evaluating safety critical situations. No research was found, however, that investigates or quantifies either of these relationships. In general, the acceleration and deceleration characteristics of bicyclists share many of the influences on efficiency and safety with speed. These effects are explained in the previous section.

**Spacing, density and flow**

Motorised road traffic predominantly moves in single file along road lanes with intermittent lane changes. Spacing and density can therefore be reduced to one-dimensional parameters; the longitudinal following distance between two vehicles in the same lane and the resulting density given in vehicles/km. Although motor vehicles do not always drive in the centre of the lane and the lateral position is variable, the lateral component when calculating density is typically fixed by the width of the lane. Bicyclists, on the other hand, have significantly more freedom of lateral movement and can often pass one another in the same facility (road lane or bicycle facility), making it necessary to measure the spacing and density in two dimensions, longitudinally and laterally.

KHAN & RAKSUNTORN [2001] examined 29 passing manoeuvres and found an average lateral spacing of 1.78 m with a range between 1.35 m and 2.36 m over the entire observation period. The average maximum lateral spacing, which is an average of the maximum lateral spacing observed for each bicycle during the observation period, was found to be 1.88 m with little variation. In their analysis of 135 passing manoeuvres in various German cities, FALKENBERG ET AL. [2003] found a much smaller average lateral spacing of approximately 0.60 m, with a minimum value of 0.20 m and a maximum of over a meter. This discrepancy could be related to the facility width, as KHAN & RAKSUNTORN [2001] studied a single 3 m wide bicycle path while FALKENBERG ET AL. [2003] investigated 21 facilities ranging in width from 1.2 m to 4.3 m. FALKENBERG ET AL. [2003] also measured the volume and speed of bicycle traffic on the observed road segments and used this information to calculate the density of bicycle traffic. They found values ranging between 1.6 bicyclists/km/lane and 17.9 bicyclists/km/lane. The following factors were found in the literature to have an influence on the spacing, density and flow of bicyclist traffic.

- **Speed:** LING & WU [2004] measured the average density of 49 groups of bicyclists at normal speeds through a signalised intersection in China and found a value of 0.31 bicyclists/$m^2$. They also analysed the behaviour of 46 groups of bicyclists that were moving at a slower speed due to interference with motor vehicle traffic and found an average density of 0.42 bicyclists/$m^2$. KHAN & RAKSUNTORN [2001], however, did not find a significant relationship between the speed of the bicyclists and the maximum lateral spacing. This could be due to the small sample size of 29 events.

- **Paired riding:** A phenomenon unique to bicycle traffic is two or more bicyclists moving together or riding two abreast in order to chat. KHAN & RAKSUNTORN [2001] measured the

spacing between two bicyclists riding together over a distance of at least 91 m (300 ft.) and found an average lateral spacing of 1.05 m and an average longitudinal spacing of 0.60 m. The lateral spacing is much lower than that found in overtaking or meeting events (1.78 m and 1.95 m, respectively).

The spacing, density and flow of bicycle traffic likely influence bicyclist safety. However, to the knowledge of the author, the direct relationships between safety and bicyclists' spacing, density and flow are yet to be studied.

The spacing, density and flow of bicycle traffic have, per definition, an important impact on traffic efficiency, both on road segments and at intersections. The applicability and determination of the fundamental relationship between density, flow and speed for bicycle traffic, analogous to the relationship for vehicle traffic, have been pursued by a number of researchers. GOULD & KARNER [2010] provide a helpful review of research investigating the bicycle fundamental relationship. Estimates for maximum capacity for bicycle facilities range from 770 bicyclists/h/m [SMITH, 1972] to 4600 bicyclists/h/m [BOTMA & PAPENDRECHT, 1991]. The relatively large body of research in this area is not summarised here, however, as the focus of this dissertation is microscopic traffic analysis. Interested readers are referred to the papers by ALLEN ET AL. [1998], GOULD & KARNER [2010] and TAYLOR & DAVIS [1999].

**Lateral position**

The position of the bicyclist is defined here as the physical location within the road space that a bicyclist selects for riding or stopping. The lateral position of bicyclists measured from the kerb or edge of the infrastructure is examined in a number of studies [DUTHIE ET AL., 2011; HARKEY & STEWART, 1997; HUNTER ET AL., 1999; HUNTER & STEWART, 1999; KROLL & RAMEY, 1977; MCHENRY & WALLACE, 1985]. DUTHIE ET AL. [2011] observed the lateral position of 96 bicyclists while they rode loops on a predefined circuit with multiple video data recording sites. They used the manually processed data describing the lateral position of the bicyclists to estimate a linear regression model. The base case of the model is a road with a wide kerb lane (no bicycle facility) during an overtaking event (car overtaking bicyclist). In this situation, the average lateral distance from the kerb face was found to be 0.5 m. During normal riding (no overtaking event), bicyclists are predicted to ride 0.7 m from the kerb face. KHAN & RAKSUNTORN [2001] analysed the lateral position of bicyclists riding on a 3 m wide separated bicycle path and found an average distance from the right side of the bicycle path of 0.9 m during normal riding and 0.6 m for a passed bicycle during an overtaking event. The factors described below were found to influence the position of bicyclists.

- **Infrastructure:** The linear regression model developed by DUTHIE ET AL. [2011] indicates that bicyclists ride further from the kerb face if there is a bicycle lane. The width of the bicycle lane also plays a role with each additional 0.3 m in width leading to an increase of 0.1 m in distance from the kerb face. HARKEY & STEWART [1997] used video data of

1583 bicycle-motor vehicle interactions to estimate the lateral position of bicyclists and motor vehicles during overtaking events on roadways with different types of bicycle infrastructure. They found a lateral distance from the road edge of 0.4 m on roadways with wide kerb lanes and 0.8 m on roadways with a bicycle lane or paved shoulder. VAN HOUTEN & SEIDERMAN [2005] found bicycle lanes marginally increase the average distance from the kerb and decrease the distribution in observed distances.

- **Parking:** DUTHIE ET AL. [2011] found the presence of parking to increase the distance from the kerb by 0.15 m and by 0.3 m if the parking facility is continuous.

- **Speed:** KHAN & RAKSUNTORN [2001] found no significant relationship between the lateral positioning of the bicyclist and the speed.

The position of bicyclists while riding along road segments and across intersections has an impact on the overall flow of traffic as well as bicyclist safety. On roadways without bicycle lanes, a bicyclist who chooses to ride further from the kerb can hinder motor vehicle traffic driving in the same lane to a larger extent than a bicyclist riding closer to the kerb. However, bicyclists riding close to cars parked along the side of the road put themselves in danger of being hit by a car door that is opened unexpectedly. No studies were identified in which these relationships were quantitatively investigated.

**Overtaking and meeting**

An overtaking event occurs when a faster bicyclist approaches a slower bicyclist travelling in the same direction and adjusts his or her speed and position to pass the slower bicyclist. A meeting event is defined as a situation in which two bicyclists travelling in opposite directions approach each other head on and must adjust their lateral position to move past one another without colliding.

An overtaking event is characterised by a number of parameters, including the length of the event in time and distance, the speeds of the two bicyclists throughout the event and the spacing between the bicyclists. KHAN & RAKSUNTORN [2001] carried out a study of overtaking and meeting manoeuvres on a 3 m wide separated bicycle path in the USA. The analysis of a relatively small sample of 29 overtaking events yielded an average speed difference between the overtaking and passed bicyclists of 2.6 m/s, which was found to remain relatively constant throughout the manoeuvre. A minimum speed difference of 1.5 m/s was noted. If the difference dropped below this threshold, the overtaking bicyclist was found to increase his or her speed. The authors derived a linear equation to express the speed of the overtaking bicyclist as a function of the speed of the passed bicyclist. In contrast, BOTMA & PAPENDRECHT [1991] found overtaking bicyclists maintain a constant speed while carrying out an overtaking manoeuvre. FALKENBERG ET AL. [2003] observed 135 overtaking manoeuvres on 21 roadway sections in Germany and found that the overtaking bicyclist usually does not have to reduce

his or her riding speed in reaction to the bicyclist who is to be passed. KHAN & RAKSUNTORN [2001] found an average length of an overtaking manoeuvre to be 91.4 m (3 m wide bicycle path), while BOTMA & PAPENDRECHT [1991] found overtaking manoeuvre lengths of 57 m (11.0 s) and 24 m (4.5 s) for 2.4 m and 1.8 m wide separated bicycle paths. As well as the type and width of the facility, other conditions, such as whether it was in an urban or rural region, may have had an influence on the results.

Meeting events can be reduced to the lateral distance between two bicyclists as they move past one another. KHAN & RAKSUNTORN [2001] measured an average lateral spacing of 1.95 m on a 3 m wide separated bicycle path. Unlike overtaking events, the length and duration of the event are irrelevant. Although the speed of both bicyclists during the meeting event is likely important, this factor was not investigated in the study. No information was found regarding the factors that influence overtaking or meeting events.

The number of overtaking and meeting events influences the LOS for bicyclists. BOTMA [1995] suggested using the number of hindrance events, which are overtaking, meeting or combined overtaking and meeting events, as an indicator of the LOS for bicycles on separated facilities [BOTMA, 1995]. The main idea of this concept is that each hindrance event causes bicyclists to adjust their speed or path, which in turn decreases the quality of the bicycle trip. This idea has been adopted in a modified form in both the American Highway Capacity Manual [NATIONAL RESEARCH COUNCIL, 2000; NATIONAL RESEARCH COUNCIL, 2010] and the German "*Handbuch für die Bemessung von Straßenverkehrsanlagen*" [FORSCHUNGSGESELLSCHAFT FÜR STRAßEN- UND VERKEHRSWESEN, 2015]. Although the number of overtaking and meeting events is widely used in determining LOS on bicycle facilities, GOULD & KARNER [2010] noted that they are relatively difficult to observe and measure in the field.

Although the number and characteristics of the overtaking and meeting events likely impact the safety of bicyclists, research assessing this effect was not found in the literature review.

**Gap acceptance**

Gap acceptance plays an extremely important role in the interaction of conflicting traffic streams at intersections. At signalised intersections, conflicting streams that are serviced in the same signal group (partially conflicting streams) are regulated by priority rules. The most common examples here are left turning vehicles or bicyclists serviced in the same phase as vehicles or bicyclists moving straight across the intersection in the opposite direction. Another example is road users turning left or right that must pass through a stream of pedestrians or bicyclists crossing adjacently in the same phase. In all mentioned cases, the turning vehicle or bicyclist must wait for a large enough gap in the prioritised stream. The only factor found to affect the gap acceptance of bicyclists is the type of stop. The findings are explained below:

- **Type of Stop:** FERRARA [1975] investigated the gap acceptance behaviour of bicyclists crossing two lanes of motor vehicle traffic. He separated the observed bicyclists into two groups, those who came to a complete stop and those who came to a rolling stop. However, he had difficulty categorising the observed bicyclists and in any case found a critical gap of between 3.5 s and 4 s for both groups. He rejected the idea of differing behaviour between the groups and proposed a single linear model for gap acceptance. OPIELA ET AL. [1980] studied the gap acceptance of 260 bicyclists as they crossed two lanes of one-way motor vehicle traffic. They found gap acceptance to be affected by the type of stop, with bicyclists who came to a rolling stop accepting much shorter gaps compared to those who came to a complete stop. The observed gap acceptance data was found to follow a logarithmic distribution. The critical gap, which represents the intersection between the gap acceptance and gap rejection, was found to be 3.2 s.

According to FERRARA [1975], SAITZ [1968] studied gap acceptance of bicyclists in East Germany and found critical gaps of 8.3 s and 8.0 s for bicyclists crossing two-way and one-way streams of motor vehicle traffic, respectively (original publication could not be located). The reported critical gaps found by SAITZ [1968] are much larger than those found by other researchers.

Overall traffic efficiency at intersections is significantly impacted by gap acceptance of road users. One factor affecting traffic efficiency is the delay of left and right turning vehicles or bicyclists due to interactions with partially conflicting streams, which are conflicting traffic streams that are included in the same signal phase. ALLEN ET AL. [1998] studied the relationship between the bicycle volume on a given intersection approach and the percentage of the green phase in which the conflict area for left and right turning vehicles is blocked by bicyclists. They concluded that there is very little impact on traffic flow for bicycle volumes less than 60 bicycle/h. A linear equation was developed to predict the proportion of green time during which the conflict zone is occupied based on the volume of bicycle traffic. Extrapolation was used to predict that a full blockage of the conflict zone occurs at 2646 bicyclists/hour green.

Gap acceptance plays a central role in the safety of bicyclists at intersections. A major safety concern at signalised intersections involves vehicles turning right in the same signal phase as bicyclists travelling straight across the intersection, which are often positioned to the right of the turning vehicle traffic. This leads to situations in which drivers do not see bicyclists (look-but-failed-to-see-error) or accept gaps in bicycle traffic that are not large enough. This, however, is related to the gap acceptance behaviour of motor vehicle drivers and not bicyclists. The relationship between gap acceptance behaviour of bicyclists, the cause of different gap acceptance behaviours and bicyclist safety was not examined in the literature reviewed.

**Summary**

A short overview of the operational behaviours examined in Section 2.1.2 is presented in Tab. 2.2. Each of the six reviewed aspects is assessed with regard to the current state of knowledge and the influence of the behaviour on the safety of bicyclists and the overall traffic efficiency. The current state of knowledge is evaluated using a three-category ordinal scale (poor, fair, good). The influence of the behaviour on the safety of bicyclists and the overall traffic efficiency are rated using a second three-category ordinal scale (small, moderate, large).

| Aspect | Assessment |
|---|---|
| **Speed** | **General knowledge:** A relatively large number of studies were found that examine the speed of bicyclists crossing intersections (28). The link between tactical behaviour and speed was not investigated in previous studies. **Good**<br>**Safety effects:** Bicyclist speed is positively correlated with the severity of accidents involving bicyclists. Speed variation also plays an important role in drivers misjudging gap size. **Large**<br>**Efficiency effects:** Speed is crucial in the design and evaluation of traffic signals. At signalised intersections, the determination of intergreen times and the offset of coordinated signals depend largely on speed distribution. **Large** |
| **Acceleration and deceleration** | **General knowledge:** In comparison with the large body of research investigating bicycle speed, only four papers have investigated acceleration. **Poor**<br>**Safety effects:** Acceleration and deceleration are important parameters in the development of safety-critical situations. **Moderate**<br>**Efficiency effects:** Acceleration controls speed and as such has a similar, but arguably, less pronounced effect on efficiency (explained above). **Moderate** |
| **Spacing, density and flow** | **General knowledge:** Spacing, density and flow have received a significant amount of attention in previous studies. However, relatively little information exists regarding the environmental factors that influence these parameters. **Good**<br>**Safety effects:** Spacing, density and flow of bicycle traffic have not been shown in the literature to be significantly related to traffic safety. **Small**<br>**Efficiency effects:** By definition, the efficiency of bicycle traffic is dependent on spacing, density and flow. The efficiency of vehicular and pedestrian traffic can also be impacted by these aspects of bicyclist behaviour. **Large** |

**Tab. 2.2** Assessment of the current state of knowledge and traffic safety and efficiency effects of the analysed aspects of operational behaviour

| Aspect | Assessment |
|---|---|
| Lateral position | **General knowledge:** Although only a limited number of studies have investigated bicyclist position while riding, the most important parameters describing this aspect have been quantified. **Good**<br>**Safety effects:** The main influence of riding position on bicyclist safety arise from the danger of the doors of parked cars being opened or motor vehicles overtaking with a small safety gap. **Moderate**<br>**Efficiency effects:** The position of bicyclists mainly influences the efficiency of vehicular traffic when bicyclists ride on the roadway in narrow vehicle lanes. Otherwise, the effect is negligible. **Small** |
| Overtaking and meeting | **General knowledge:** Although only a limited number of studies have investigated overtaking and meeting, the most important parameters describing these manoeuvres have been quantified. **Fair**<br>**Safety effects:** No studies were found that investigated the safety impact of overtaking and meeting events. Due to the degree of interaction required, it is hypothesised that this relationship is important. **Moderate**<br>**Efficiency effects:** The number of hindrance events (overtaking, meeting or both) is used to derive bicycle LOS in many countries. **Large** |
| Gap acceptance | **General knowledge:** Three relatively dated studies were found that examine the gap acceptance behaviour of bicyclists. A systematic investigation of gap acceptance in a wide variety of situations would be useful. **Fair**<br>**Safety effects:** While the gap acceptance of vehicle drivers (particularly truck drivers turning right) has a strong, well-understood influence on bicyclist safety, the effect of bicyclists' gap acceptance is not well studied. **Moderate**<br>**Efficiency effects:** Gap acceptance on the part of bicyclists plays a role in bicycle and overall traffic efficiency depending on the manoeuvre and type of infrastructure used by bicyclists. **Moderate** |

**Tab. 2.2** Assessment of the current state of knowledge and traffic safety and efficiency effects of the analysed aspects of operational behaviour (cont.)

## 2.2 Modelling review

Modelling human behaviour is an incredibly complex task. As drivers, humans are able to perceive an enormous amount of input from the environment, filter and process this information, and react to the situation, all within between 0.7 – 1.5 s [GREEN, 2007]. It is likely impossible to develop a comprehensive collection of models to exhaustively describe the situation, road user perception, filtering and processing of information and reaction mechanisms. Even if it were possible to completely model these aspects, execution in real

time would be challenging. Consequently, all simulations are a simplification of reality that can be useful in understanding some situations or predicting some future developments.

In the 1970s and 1980s, road safety researchers established a collection of theories and models describing the balance maintained by a driver between the subjective risk or difficulty of a task and the expected benefit realized by carrying out the task [MICHON, 1985; NÄÄTÄNEN & SUMMALA, 1974; NÄÄTÄNEN & SUMMALA, 1975; TAYLOR, 1964; VAN DER MOLEN & BÖTTICHER, 1988; WILDE, 1982; WILDE, 1988]. From a macroscopic viewpoint, these theories are used to explain why measures introduced to increase road safety, such as creating forgiving road infrastructure and improving vehicle design, do not always have the expected positive impact on accident rates. In conditions that are objectively safer or easier to navigate, drivers increase the riskiness of their behaviour to arrive at the same accepted level of subjective risk or task difficulty. From a microscopic perspective, the models can be used to predict the actions carried out by a given road user in a certain situation. An example given in the literature involves a car overtaking a slow-moving truck on a two-way road with one lane in each direction [VAN DER MOLEN & BÖTTICHER, 1988]. The psychological models can be used to predict whether or not the expected benefit of overtaking the truck outweigh the inherent risk of overtaking using the opposing lane. These theories developed by road safety researchers have acted as a backbone for subsequent model development for microscopic traffic simulation. Many simulation tools are built upon a common concept: modelled road users progress through the simulated environments by balancing the drive to move forward with the desired speed along the desired path with the necessity of avoiding conflicts with obstacles and other road users.

Microscopic traffic simulation is used as a platform to interconnect many models that represent the behaviour of drivers, bicyclists and pedestrians, interactions between road users, the road environment and the traffic control. The resulting simulated traffic environment is a complex system in which the specific models interact to form an overall system that is more than the sum of all the models acting independently [BARCELÓ, 2010]. Once the traffic simulation is created, calibrated and validated using data from an existing situation, target aspects of the system can be manipulated by researchers to investigate the effects of these changes on the whole system. For example, engineers can use microscopic simulation to predict the effect of a given traffic control measure (e.g. coordinated traffic signals) on traffic efficiency before this measure is implemented. Surrogate indicators for traffic safety that describe the interactions between road users, such as Time-To-Collision (TTC), can be used to investigate probable changes in road safety. Because the local characteristics of the site are included in the simulation, the engineer does not solely rely on partially relevant experience at other sites in the planning and evaluation of traffic measures. Another motivation for using microscopic traffic simulation is to project the effects of a given measure or system when deployed on a large scale. For example, vehicle to vehicle (V2V) and vehicle to infrastructure (V2I) communication as well as automated driving technologies are still in the

development phase and can only be tested in reality on a relatively small scale. Microscopic traffic simulation tools are useful for predicting the effects of these technologies on road capacity, traffic flow and road safety with varying rates of equipment.

The methods used for microscopic simulation can be classified based on the continuity of the simulation space and time. The vast majority of simulation tools run in discretised time steps of a given length. The position and attributes of each road user and the road environment are updated in each time step. Time step size can range from quite small (e.g. 0.04 s per time step) to relatively large (e.g. >1 s per time step) depending on the level of detail required from the simulation. The road space can be modelled either continually or in discrete sections. Cellular automata approaches, based on the idea first proposed by NAGEL & SCHRECKENBERG [1992], discretize the road space into an array of cells. Vehicles within the simulation follow predefined routes through the cell array using four driving regimes, acceleration, deceleration, randomization and update/move. Other simulation approaches define space continually in the longitudinal direction and discretely in the lateral direction. Car-following models are used to position road users at any point in the longitudinal direction based on the individual driving characteristics and the reaction to leading vehicles. Lateral movement is restricted to lane changes, which are modelled using rule-based or discrete choice models. A final method uses a continuous representation of space in both the lateral and longitudinal direction. The most common example of complete space continuity is found in pedestrian modelling. Models such as the social force model, first proposed by HELBING & MOLNAR [1995], allow pedestrians to move to any point on a continuous two-dimensional plane.

A wide variety of commercial and open source tools are available for the microscopic simulation of traffic. The methods used for building the road network, including road segments, sidewalks, bicycle facilities and intersections, as well as the models and methods used to emulate road user behaviour vary widely between tools. The various tools are not reviewed in this dissertation due to the wealth of literature available in this field. If the reader is interested in learning more about available simulation tools, the book *Fundamentals of Traffic Simulation* [BARCELÓ, 2010] provides a good introduction to traffic simulation in general and more detailed information concerning selected simulation tools. In the German book *Simulation des Straßenverkehrs in der Großstadt: das Mit-und Gegeneinander verschiedener Verkehrsteilnehmertypen*, DALLMEYER [2014] summarises and evaluates a number of simulation tools.

In the next sections, the state of the art in modelling the tactical and operational behaviour of bicyclists is presented. Only a handful of the reviewed approaches are implemented in currently available simulation tools. The remainder includes theoretical or independently implemented models that could be used to improve the simulation of bicycle traffic.

In Section 2.2.1, existing models for recreating the tactical behaviour of bicyclists at signalised intersections are summarised. This overview is followed by a review of approaches for

modelling the operational behaviour of bicyclists in Section 2.2.2. An assessment of the capability of existing modelling approaches to capture the complex behaviour of bicyclists is given at the end of each of the sections[2].

### 2.2.1 Tactical models

Modelling and simulation approaches reviewed in this section are used to recreate the tactical behaviour of bicyclists, which is limited here to the four decisions defined in Section 2.1.1, infrastructure selection (bicycle facility, roadway or sidewalk), reaction to a red signal, travelling with or against the mandatory direction of travel and path selection across an intersection. Relatively few approaches for modelling the tactical behaviour of bicyclists were found in the literature review. Even fewer approaches were found that have been implemented in currently available microscopic simulation tools. In this section, three types of models are investigated, probabilistic approaches, discrete choice models and continuous pathfinding approaches. The first two types are based on the conceptualization of tactical behaviour as a set of discrete decisions that are made at a defined position or point in time. The third type formulates tactical behaviour as a continuous decision process. For each of the three approach types, examples are given and the application in currently available simulation tools is noted.

#### Probabilistic approaches

The most straightforward method of implementing tactical choice models in microscopic simulations is to use observed frequencies of certain behaviours. A number of previous studies have investigated the rate of various tactical behaviours occurring (see Section 2.1.1). These measured values can be used to specify the proportion of simulated road users carrying out a given behaviour. Some simulation tools offer users the option to specify frequencies of tactical behaviour directly. For example, using the microscopic simulation tool PTV Vissim [FELLENDORF & VORTISCH, 2010; PTV PLANUNG TRANSPORT VERKEHR AG, 2015], it is possible to regulate the percentage of road users who obey red traffic signals by adjusting the given parameter. The outcome of the tactical decision, in this case red light violation, is decided randomly based on the specified probability. Other tactical behaviours, such as infrastructure selection, are more difficult or impossible to simulate using the basic functionalities of the software and must be controlled using an Application Programming

[2] Parts of the text in this section were originally published in the paper:

TWADDLE, H.; SCHENDZIELORZ, T. & FAKLER, O., Bicycles in Urban Areas: Review of Existing Methods for Modeling Behavior. Transportation Research Record: Journal of the Transportation Research Board, 2434, pp.140–146. [TWADDLE, SCHENDZIELORZ & FAKLER, 2014]

Interface (API). Although the portion of bicyclists carrying out given behaviours can be calibrated to agree with reality, the main drawback of this approach is that the causation of the tactical behaviours is not taken into account. The simulated road users are therefore insensitive to the environment and cannot respond to different situations by realising different tactical behaviours.

**Discrete choice models**

The vast majority of the theoretic or independently implemented models for replicating or predicting the tactical behaviour of bicyclists are logistic regression models. In such approaches, observed choice outcomes are used to quantify relationships between independent variables describing the bicyclist and the environment and a nominal dependent variable describing a discrete tactical behaviour. These resulting models are used to predict the outcome of future choice situations. For example, JOHNSON ET AL. [2011] developed a binomial logistic regression model that relates the type of facility, gender of the bicyclist, direction of travel, presence of other road users, and average traffic volume to a red light violation. The resulting model predicts if a bicyclist in a given situation will violate a red signal.

An overview of the regression models developed for predicting the selected tactical behaviours for bicyclists is given in Section 2.1.1. This type of model can be implemented in a number of traffic simulation tools with an API that allows for the direct manipulation of the simulated road users, such as the COM interface for PTV Vissim and *TraCI* for SUMO. These models can be implemented in the same way as probabilistic approaches. When the simulated bicyclist reaches a predefined decision point, the independent variables are extracted from the simulation using the API and the probabilities of the various choice outcomes are calculated using the logistic regression or discrete choice model. The option with the highest probability is selected and returned to the simulation. If for example, the model proposed by JOHNSON ET AL [2011] were implemented in a microscopic simulation tool, the simulated bicyclists would be instructed to stop or to violate a red traffic signal depending on the input parameters of the model. The advantage of this approach over probabilistic methods is that the causation of the tactical behaviour, and not only the proportion of behaviour observed to carry out a certain choice, is reflected in the simulation. No examples of the direct implementation of a discrete choice model for tactical behaviour in a currently available simulation tool were found.

**Pathfinding approaches**

Approaches in this category model the tactical behaviour of bicyclists continuously rather than discretely. Two approaches were identified in this review that model the pathfinding behaviour of bicyclists, one at non-signalised intersections and the other in shared space. SCHÖNAUER & SCHROM-FEIERTAG [2010] and SCHÖNAUER ET AL. [2012] added a tactical force vector to their social force model in order to simulate longer term tactical path selection in

order to avoid conflicts for bicyclists, pedestrians and vehicles in shared space environments. The movement of the road users is the result of the sum of an infrastructure guiding force, an adapted social force model and a tactical force model. The tactical force model is based on the game concept proposed by VON STACKELBERG [1934], which is a non-symmetric hierarchical game model with follower and leader players. In the application from SCHÖNAUER ET AL. [2012], potential conflicts between road users are identified based on the planned trajectories of the uninfluenced road users. Various possibilities for avoiding the conflict are considered by both road users. The range of options is created by first determining the strategy of the leading road user and then the reactive strategy of the following road users. The total utility of the strategy is the sum of the partial payoffs for both road users. The approach was validated using trajectory data from conflicts between vehicles and pedestrians and therefore the validity of the approach for bicycle traffic cannot be verified.

LING & WU [2009] proposed a model that makes use of fuzzy logic rules to determine the path selection of bicyclists at uncontrolled intersections. The model consists of three sub-models:

1. Situation detection – The speed, direction and position of other road users within a given area are detected. Conflict points with the detected road users are calculated and fuzzy logic rules are used to estimate the relative danger associated with each.

2. Path sketching – Information collected with the situation detection model is used to determine possible trajectories. The directness, comfort and efficiency are estimated for each of the possible trajectories and fuzzy logic rules are used to evaluate each.

3. Reactive path generation – The path choice is carried out and the information is sent to the situation detection model.

The model was tested using empirical data and the results indicated that the modelled trajectories reflect those observed at the test site.

Researchers in robotics are very active in the field of motion prediction and path planning. Robots, like simulated road users, must navigate around obstacles in their environment to reach a certain objective or location. A number of approaches use observed trajectories in given situations and pattern recognition algorithms to learn pathfinding behaviour from human actors. For example, KUDERER ET AL. [2013] and KITANI ET AL. [2012] use observed human trajectories through a room with furniture and a parking lot, respectively, to create pathway hypotheses based on the characteristics of the physical environment and the presence of interacting objects. A detailed review of methods for motion prediction, both in the realm of robotics and traffic analytics is given by MOHAMED & SAUNIER [2013].

None of the existing pathfinding approaches from robotics or traffic modelling are implemented in currently available microscopic simulation tools. The path selection across an intersection can be partially implemented by increasing the complexity of the network. Users

can define an exhaustive number of potential routes for bicyclists approaching and crossing an intersection. Routing methods used in the simulation tool can be used to define the path selection of the bicyclists. Although this is a viable solution for very small networks, the complexity associated with defining conflict points for many routes makes this option daunting for large networks. Furthermore, this option still represents a discrete choice for pathfinding and does not reflect the actual pathfinding choices of bicyclists made in a continuous two-dimensional space.

No models were identified in this review that consider the choice between riding on a bicycle facility, the roadway or sidewalk. A method similar to that for path selection can be used to implement infrastructure selection in microscopic simulation tools. However, a critical issue in using this approach to simulate infrastructure selection is the current incompatibility between pedestrian simulation and road traffic simulation. In general, pedestrians in microscopic simulation tools are modelled using approaches that allow for continuous movement on a plane. Cars, bicycles and other road users, on the other hand, are confined to unidirectional facilities on the road. Interaction between the models is not currently possible, meaning that pedestrians cannot enter the roadway and interact with cars or bicyclists and cars and bicyclists cannot interact with pedestrians on the sidewalk. A workaround for this issue is to create a virtual pedestrian when a bicyclist moves on the sidewalk, which makes it possible for the other simulated pedestrians to react to the bicyclist. This approach, however, is cumbersome and decreases simulation speed dramatically.

No models were identified that predict the direction of travel of bicyclists on segments with a mandatory direction of travel. The possibility of simulating the tactical choice concerning the direction of travel (riding with or against the mandatory direction of travel) differs depending on the simulation environment. In simulations with continuous lateral and longitudinal space, the direction of travel can be altered quite easily. Indeed, the challenge in these models is to restrain the direction of movement to realistically reflect the dynamics of riding a bicycle [SCHÖNAUER ET AL., 2012]. However, in simulations based on unidirectional links, such as the road segments in PTV Vissim and SUMO, movement is only possible in the predefined direction of travel. It is currently impossible to simulate road users travelling in the opposite direction. This restraint, however, has been recently relaxed in the simulation tool SUMO, with pedestrians moving on bidirectional pedestrian links. This approach could be extended for simulating bicycle traffic.

**Summary**

The current potential to model and simulate each of the aspects of tactical behaviour identified in Section 2.1.1 is assessed in Tab. 2.3. A three-category ordinal scale (poor, fair, good) is used to qualitatively evaluate the current modelling and simulating capabilities.

| Aspect | Current model capability |
|---|---|
| **Reaction to red signal** | It is currently possible to simulate this aspect of behaviour directly in microscopic traffic simulation tools by defining the proportion of road users that obey a red traffic signal using simulation parameters. It is also possible to use APIs to connect external discrete choice models to control the decisions of simulated bicyclists during predefined events or at given positions. Logistic regression models predicting this behaviour have been developed by previous researchers. However, many of these models use personal attributes of the bicyclist as predictors, which only exist if different types of simulated road users are created in current microscopic traffic simulations. **Good** |
| **Infrastructure selection** | No models were found in this review that predict the infrastructure selection of bicyclists. This aspect of behaviour is difficult to simulate using currently available simulation tools due to the segregation of model types. **Poor** |
| **Riding direction** | The ability to simulate riding against the given direction of travel varies greatly depending on how the road environment is simulated. If space is modelled continuously in two dimensions without a given direction of travel, bicyclists can theoretically move in all directions. However, no approaches were found in which bicyclists are simulated riding against the given direction using facilities that are characterised by a mandatory direction of travel. **Poor** |
| **Path selection** | Two models were found in this review that examine the continuous path selection of bicyclists, one at non-signalised intersections and one for shared space. These approaches are most suitable for application in simulation tools with continuous two-dimensional space. With regard to simulation tools with directional infrastructure, complex path selection can be simulated by creating many possible pathways across an intersection and predefining routing probabilities or discrete models for route selection. However, the variety in pathfinding behaviours is limited by the number of predefined routes. **Poor** |

**Tab. 2.3** Current model capability for the examined aspects of tactical behaviour

### 2.2.2 Operational models

In this section, approaches for modelling and simulating the operational behaviour of bicyclists are reviewed. The six aspects of this behaviour that were introduced and examined in Section 2.1.2, speed, acceleration and deceleration, overtaking and meeting, gap acceptance, position and spacing and density, are used to guide the review in this section. Three approaches for modelling the operational behaviour of all road users are selected based on the level of discretization of the road space. Cellular automata use complete discretization

of road space into an array of cells, car-following models generally use continuous longitudinal and discrete lateral space and social force models function in continuous two-dimensional space. These modelling approaches are introduced and applications for modelling bicycle traffic are examined.

**Cellular Automata Models**

Cellular automata (CA) are time and space discrete models. In the first model by NAGEL & SCHRECKENBERG [1992], two-dimensional space is modelled as a one-dimensional array of cells, each having a length of 7.5 m (roughly the length of one car). This was later extended to model two-lane highway traffic by RICKERT ET AL. [1996]. In both cases, the cells in the array do not overlap each other and it is not possible for more than one road user to occupy a cell during one time step. For this reason, CA provide a simple and fast method for modelling homogeneous traffic flows that follow lane discipline. The possibilities for modelling mixed traffic streams and the interaction between different modes of transportation are limited. However, a number of extensions of the original CA have been suggested to make it possible to include bicycle traffic.

One option for including many types of road users in CA is to create an array of cells sized in accordance with the dimensions of the smallest road user (width and length) and allow road users to occupy more than one cell per time step. This method was used by YAO ET AL. [2009] to model situations in China where more than one lane of bicycle traffic runs along a street with one or more lanes of car traffic. Each cell in the model represents 1m x 1m in reality. A bicycle occupies 3x1 cells and a car occupies 5x3 cells. Interactions between bicycles and cars are classified into two types: friction and blockage. The driving resistance for car drivers is determined based on the presence of bicycles in the next lane and by bicycles that move from the bicycle lane into car lanes. This model has not yet been empirically validated or calibrated.

MALLIKARJUNA & RAMACHANDRA RAO [2008] used a similar approach to model mixed traffic streams in India. In this model, the cell lengths are based on the acceleration and deceleration properties of the road users and the cell widths are based on the width and observed lateral spacing maintained between different groups of road users. In order to consider the different lateral behaviour of the road user types, a two-lane road is divided into five sub-lanes and five types of lateral movements are defined. The lateral position of the road users is updated in a first step. In a second step, the longitudinal position is updated depending on the acceleration and deceleration characteristics of the road user type and the available space to move forward.

VASIĆ & RUSKIN [2011A; 2011B; 2012A; 2012B] developed a CA to depict car traffic and single file bicycle traffic. In this model, a cell array with appropriately sized cells is created for each type of road user (in this case bicyclists and cars). When the pathways of multiple traffic

streams intersect, the cells in the CA model overlap. The movement of the vehicles is then determined by the impingement of the leading cells. A cell is impinged if it is occupied or any of the overlapping cells are occupied. The lateral interaction between bicycles and cars travelling in the same direction on the same roadway is only considered on 'narrow roads' where the velocity of the cars is limited based on the longitudinal distance to the next leading bicycle. The advantage of the model proposed by VASIĆ & RUSKIN is that the geometry of the intersection can be directly translated into a fitting array of cells. Complex interactions are extracted from this array.

GOULD & KARNER [2010] used a CA to derive the macroscopic properties of bicycles travelling on a one-way bicycle facility. The simulated lane was divided into two hypothetical lanes and bicyclists were divided into two groups based on their speed (slow and fast). The model was operated using the extended rules proposed by RICKERT ET AL. [1996] and was validated and calibrated using empirical data. The results from the simulation were used to estimate the relationship between speed, density and flow of bicycle traffic. However, no observations were made in reality where the density of bicycles reached or passed a critical level and began to negatively affect the bicycle flow.

A final option proposed for adapting the original CA for simulating bicycle traffic is to allow more than one road user to occupy a cell in one-time step [JIA ET AL., 2007; JIANG ET AL., 2004]. This provides a macroscopic approach to estimate the capacity of bicycle infrastructure and will not be discussed further due to the microscopic focus of this dissertation.

**Car-following models**

The majority of traffic simulation tools employ two components to independently model the longitudinal and lateral motion of road users (e.g. PTV Vissim, Aimsun, SUMO). The longitudinal motion models are space continuous and time discrete and typically use one of the three types of car-following models; Gazis-Herman-Rothery models, safety distance models, and psychophysical models [BARCELÓ, 2010]. All of these models use the speed and position of a leading road user to determine the behaviour of the following road user. If there are no other road users present, or a leading road user is far enough ahead not to influence the behaviour, road users strive to maintain a predefined speed. The lateral movement of road users is modelled using a rule-based or discrete choice model, where the position and speed of other road users and the desired route of the individual road user are taken into account in the lane choosing process. Lateral movement within a lane is not possible in simulations based on discrete lane selection. If mixed traffic streams are simulated using such approaches, single-file bicycle traffic quickly causes a moving bottleneck on the road, which is very rarely the case in reality as faster vehicles overtake bicycles at the earliest opportunity.

The main strategy employed to improve the realism of bicycle simulation is to enable more freedom of movement in the lateral direction. The most straightforward method of

implementation is to divide the lane into a number of narrower sub-lanes. An extension of this idea is to create a continuous lateral axis rather than a discrete number of sub-lanes. This concept has been implemented by at least two microscopic simulation tools, PTV Vissim and SUMO. Users of both tools are able to adjust the global (SUMO) or vehicle specific parameters (PTV Vissim) such that all or specific road users can move freely in the lateral direction and can overtake one another in the same lane.

The desired lateral position of the road users can be specified using a simulation parameter or road user characteristic. This position is favoured in situations in which the movement of the bicyclist is not impeded by the presence or actions of other road users and is adapted if the desired movement of the bicyclist is hindered. In PTV Vissim, the lateral position is selected with the aim of minimising the TTC to other road users. This model was proposed and calibrated by FALKENBERG ET AL. [2003] and is described by FELLENDORF & VORTISCH [2010]. In SUMO, on the other hand, bicyclists use two criteria to select their lateral position, speed maximisation and minimization of the distance to the desired lateral position. This feature was first made publically available in the core code of SUMO in 2016 and no documentation was found at the point of this review describing the exact implementation of the sub-lane concept.

**Social Force Models**

The first social force model was proposed by HELBING & MOLNAR [1995] to model pedestrian dynamics. The basic operating principle of this model is that pedestrians move in reaction to the sum of a number of attractive and repulsive forces acting upon them:

- attraction towards a destination
- repulsion from obstacles
- repulsion from other pedestrians or road users

The movement within social force models is not bound to the longitudinal and lateral axis, but modelled road users move freely on a two-dimensional plane. However, bicycle dynamics differ from pedestrians, mainly because bicyclists move primarily along the longitudinal axis and do not make abrupt turns. A small number of bicycle models based on an adapted form of the original social force model have been developed and are summarised below.

SCHÖNAUER ET AL. [2012] and SCHÖNAUER & SCHROM-FEIERTAG [2010] used an adapted social force approach to simulate shared space. Motor vehicles, pedestrians and bicyclists select their path based on the geometry of the infrastructure and the behaviour of other road users. This is done using three models, an infrastructure model, an operational model and a tactical model. The infrastructure model builds a force field that uses repulsive forces from the infrastructure edges to push the road users into their intended path. The operational model,

which is used to move road users through the road space, is an adapted form of the model proposed by HELBING & MOLNAR [1995]. In order to consider the reduced degrees of freedom associated with motor vehicles and bicycles, the modified single track model for car dynamics proposed by KRAMER [2008] is implemented. The tactical model is discussed in Section 2.2.1. The approach was validated using trajectory data from vehicles and pedestrians, and as such, the validity of the approach for bicycle traffic is not verified.

LI ET AL. [2011] developed a social force model that considers mixed bicycle-automobile traffic. The forces acting on bicyclists are a forward driving force that relates the current speed with the desired speed, repulsive forces from other bicyclists and a repulsive force from the edges of the infrastructure that keeps bicyclists within the bicycle lane. Bicyclists exit the bicycle lane onto the roadway if the repulsive forces from the other bicyclists (high density in the bicycle lane) overtake the repulsive force from the lane edge. The repulsive force enacted on bicyclists by the motor vehicles in the model is much larger than the force enacted by other bicyclists. This is intended to reflect the fact that cars have a greater influence within the road space than bicyclists. The proposed model was not yet validated or calibrated using empirical data.

LIANG ET AL. [2012] developed a social force model that utilises two regimes, free flow traffic and congested traffic, to model bicycle traffic. In free flow conditions, the model uses two of the forces proposed by LI ET AL. [2011], a driving force and a repulsive force from other road users. Bicyclists are depicted within the model as ellipses. If the ellipses of two bicyclists overlap, a physical model takes over that prevents a collision from occurring. The two forces acting, in this case, are the contact force, which counteracts compression of the ellipse and the sliding friction force, which restricts relative tangential motion. The simulation results are compared with observed densities and velocities from bicyclist riding on a closed circuit.

**Summary**

The current capability to model and simulate each of the aspects of operational behaviour identified in Section 2.1.2 is assessed in Tab. 2.4. A three-category ordinal scale (poor, fair, good) is used to qualitatively evaluate the current modelling and simulating capabilities.

| Aspect | Current model capability |
|---|---|
| **Speed** | The capability of models to simulate speed accurately depends on the discretization of the simulated space. Simulation tools based on car-following approaches enable the user to define a distribution of desired speeds for simulated bicyclists. It is, however, difficult and work-intensive to simulate speed adaptations in response to features of the environment (e.g. type and quality of the facility) and the manoeuvre (e.g. turning). The resulting speed profiles of simulated bicyclists are more uniform than those observed in reality. **Good** |
| **Acceleration and deceleration** | The most commonly used approach for this behaviour is the constant acceleration model, although this does not accurately reflect observed profiles. Accuracy increases when acceleration values can be specified for various speeds. If a direct possibility for controlling acceleration is not provided, this aspect is relatively simple to control using simulation interfaces. Like speed, the accuracy and realism of the acceleration and deceleration in simulation are highly dependent on the level of discretization. **Fair** |
| **Spacing, density and flow** | The desired lateral spacing can be adjusted in most simulations by specifying values at predefined speeds. It is not possible to specify desired spacing distributions in any simulation tool, although this would increase the realism of simulated bicycle flow. It is currently impossible to consider changes in the desired spacing due to the environment, manoeuvre or interactions with other road users. Simulated bicycle flows are more uniform and consistent than in reality. **Fair** |
| **Overtaking and meeting** | The capability of current approaches to simulate overtaking events is similar to that described in the spacing, density and flow section above. Meeting events, however, are currently difficult to simulate using models based on unidirectional links. This behaviour can theoretically be simulated using social force models, but no examples of this were found in the literature. **Poor** |
| **Gap acceptance** | Bicyclists tend to avoid stopping and use other tactics such as altering their path or reducing their speed over a longer period of time to avoid coming to a complete stop. This is particularly evident when streams of bicycle traffic intersect with streams of other road users. The simulation of this scenario does not reflect reality and no approaches were found that address this issue. **Poor** |
| **Position** | CA are limited in their capacity to simulate bicyclist position accurately as road users proceed through the centre of the cells. The methods with continuous lateral space, however, offer a much more realistic depiction of positioning. In these approaches, bicyclists are assigned a desired lateral position and strive to maintain this position. Inter and intra bicyclist variations are difficult or work-intensive to simulate. **Fair** |

**Tab. 2.4** Current model capability for the examined aspects of operational behaviour

## 2.3 Research needs assessment[3]

The findings from the literature review are consolidated in this section. The scope and research question defined in Section 1 are reflected and the focus of the subsequent model development is defined. Two main questions are used to assess the need for further research:

1. Which aspects of the operational and tactical behaviour are difficult to model and simulate realistically using currently available approaches?

2. Which aspects of the behaviour are important to model realistically for the accurate evaluation of ADAS and ITS using microscopic traffic simulation?

A summary of the qualitative evaluation of the capability of existing models to simulate the investigated behaviour aspects is shown in Fig 2.7 realistically, addressing the first point above. The impacts of the behaviour aspects on bicyclist safety and overall traffic efficiency are also given in Fig 2.7. The importance of the aspects for the improvement of microscopic simulation models is greatest in the upper left-hand corner of the figure and least in the lower right-hand corner. The second question in the list above can be answered using these qualitative assessments.

The findings summarised in Fig 2.7 indicate that it is difficult to simulate three tactical behaviours using current models; the use of various facets of the infrastructure (bicycle facility, sidewalk or roadway), riding against the mandatory direction of travel and path selection. Path selection at signalised intersections is particularly in need of improvement as current modelling approaches focus on non-signalised intersections and shared space. These problems stem from the conceptualisation of road infrastructure as a network of links, which are unidirectional facilities in which the simulated road users travel in one way, and nodes where intersecting streams are managed using signal control and conflict rules.

An important motivation for this work is the simulation of bicycle behaviour that allows for the development and evaluation of ADAS and automated vehicles. It is therefore important to consider which facets of bicyclist behaviour are most critical from this viewpoint. Generally, automated driving algorithms will require a prediction concerning the future trajectories,

[3] The research needs identified here were also used as a basis for the research done within the project UR:BAN. Results from this project work are published here:

TWADDLE, H. [2017]: Analysis and Modelling of the Operational and Tactical Behaviour of Bicyclists. In K. Bengler, S. Hoffmann, D. Manstetten, A. Neukum, & J. Drüke, eds. UR:BAN Human Factors in Traffic. Springer. [TWADDLE, 2017]

TWADDLE, H. & HOFFMANN, S. [2016]: Forschungsprojekt UR:BAN - Mensch im Verkehr Schlussbericht, Germany. [TWADDLE & HOFFMANN, 2016]

including the position and speed in each time step, of bicyclists in the vicinity of the automated (or ADAS supported) vehicle. In reference to the tactical and operational behaviour framework used in this review, the path and infrastructure selection, as well as the riding direction, are very important. The speed and acceleration, as well as the position within the infrastructure, will also be important elements to consider in the development of useful models for ADAS and automated driving development.

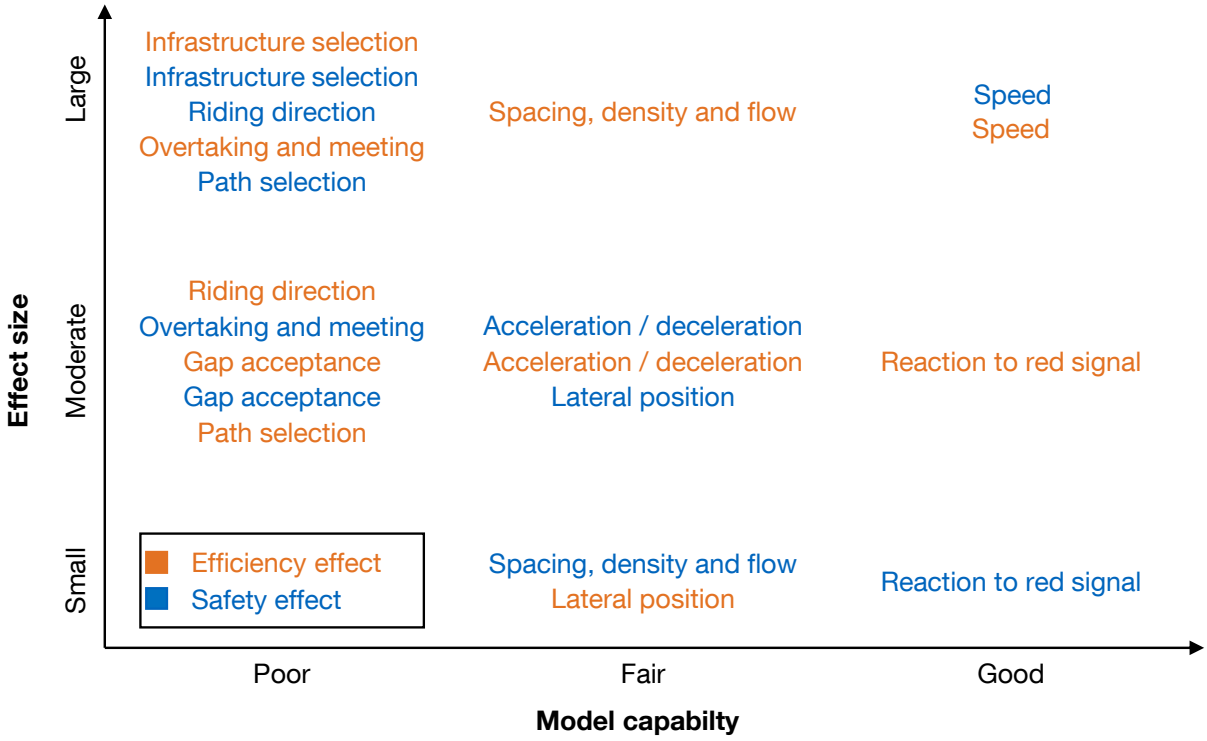



**Fig 2.7** Qualitative evaluation of the model capability and safety and efficiency effects of the investigated operational and tactical behaviour aspects

Modelling approaches should be developed that reflect the flexible nature of bicycle traffic and are compatible with unidirectional link based simulation frameworks. An approach that makes it possible to simulate bicyclists using and switching between different parts of infrastructure (bicycle facility, roadway and sidewalk), riding with and against the mandatory direction of travel and selecting unique pathways across the intersection will greatly increase the realism of bicycle simulation. The compatibility with link based simulation environments will ensure usability with common simulation software. The development of such an approach will not only address the poor ability to model the tactical behaviours mentioned above, but also the simulation of meeting events.

# 3. Experimental design, data collection and processing

In order to solve the research problem defined in Section 1.2 and address the focus points identified in Section 2.3, it is necessary to collect and process data describing the identified aspects of the tactical and operational behaviour of bicyclists at signalised intersections. These data are subsequently used in Section 4 to analyse the pathfinding behaviour of bicyclists as well as in Section 5 to specify, calibrate and evaluate behavioural models to be applied in microscopic traffic simulation. The workflow described in this section is shown in the following figure.

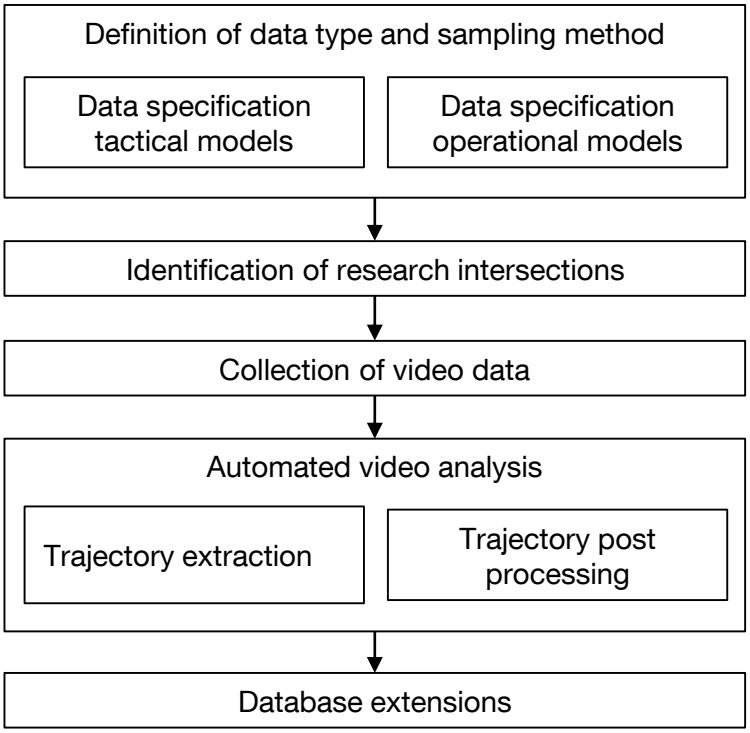



**Fig 3.1** Flowchart for experimental design, data collection and processing

The data specification and sampling approach, variable specification and selection of research intersections are presented in Section 3.1 and Section 3.2. The data collection system is explained in Section 3.3 followed by a description of the methods used for processing the data with an extended version of the open source software *Traffic Intelligence* [JACKSON ET AL., 2013] in Section 3.4. The post processing of the data as well as the extension of the database to include signal timing information and qualitative variables are described in Section 3.5 and Section 3.6, respectively.

## 3.1 Data specification and sampling approach

Model specification and calibration require data from a sample of bicyclists, interacting road users and the encompassing situation. Two types of data can be used for this purpose:

- **Stationary data** collected at a limited number of locations for a large sample of subjects.
- **Floating vehicle (bicycle) data** collected from a limited number of subjects that cover many locations.

Using stationary data it is possible to extract a rich description of the location and situation, including the exact geometry of the intersection, the phases of the signal control and the behaviour of surrounding road users. However, it is difficult to collect information about each individual bicyclist (age, gender, level of experience, etc.) or long-term behaviour, such as average speed over the entire trip. In contrast, floating bicycle data offer detailed information about the bicyclist and his or her long-term behaviour, but little data describing the situation or interactions with other road users.

For the purposes of this dissertation, stationary data is advantageous because interactions with other road users and a comprehensive description of the situation are readily available, both of which are imperative for model development. The intention of this thesis is to develop generic models of bicyclist behaviour that can be calibrated and applied to many different types of bicyclists. For this reason, all bicyclists are included in one inclusive population. Data describing the personal attributes of the observed bicyclists, including gender and age, or characteristics of the bicycle, such as electric motor support, are therefore not necessary to collect or analyse at this stage of model development.

There are many methods for collecting stationary data, including manual observation, fixed detectors for motor vehicles, pedestrians and bicycles (e.g. inductive loop detectors and radar and laser detectors) and aerial view data collection. The most common parameters that are measured from stationary data include traffic volume, speed, road occupancy and vehicle characteristics (type and length). However, in order to develop and calibrate detailed models for bicyclist behaviour, detailed descriptions of the movement and interactions of bicyclists are required. For this reason, aerial data collection is selected for the data collection method. Video data, which provides a wealth of information describing the entire road system, is selected as the type of aerial data collection.

A sampling approach is needed to identify a representative group of bicyclists from which it is possible to draw conclusions that apply to the entire population of bicyclists in Munich. To this aim, an observational study with partial variable control by means of intersection selection is implemented. The independent variables, which are introduced in the next section, describe the geometry of the intersection, the traffic flow and the signal control. In accordance with the experimental design, research intersections are selected to provide variation in the geometric and average traffic characteristics. Once the research intersections are selected, the approach transitions to an observational study in which the remaining variables, including the momentary traffic situation, the signal phase and the behaviour of road users, are not controlled by the experimenter. Behaviour data from all bicyclists crossing the intersection

during the observational study are analysed, which eliminates a potential selection bias. The particular considerations and requirements for the data needed to specify and calibrate the operational and tactical models are elaborated in the following sections.

### 3.1.1 Data specification for tactical models

Tactical behaviours can be represented using a finite number of choice outcomes and therefore models that predict categorical variables can be developed. The data required to specify and calibrate discrete choice models are therefore not derived from the trajectories of the observed bicyclists themselves, but rather from discrete tactical choices made by bicyclists while crossing an intersection. Logistic regression and discrete choice models are typically used to predict tactical behaviour. According to VAN DER PLOEG ET AL. [2014], the minimum number of observations necessary to estimate a logistic regression model with sufficient ability to predict the outcome of dichotomous and multinomial choice situations is ten observations per predictor (independent variable). More stable estimations of the beta parameters in the logistic regression are achieved with 20 observations per predictor. As data from each of the observed bicyclists can only be used to estimate the models coinciding with the choice situations met by that bicyclist, the total number of observations must include a sufficient number of bicyclists faced with each of the tactical choices. For example, in order to estimate a choice model for the reaction to a red signal, the number of bicyclists who encounter a red signal must be greater than 20 times the number of independent variables. The duration of the observational studies needed to obtain a sufficient sample size is given by the following equations.

$$N = 20\ IV \qquad \text{Eq. 3.1}$$

$$N = T\ q\ P_{min} \qquad \text{Eq. 3.2}$$

$$T = \frac{20\ IV}{q\ P_{min}} \qquad \text{Eq. 3.3}$$

where $N$ is the number of observations, $IV$ is the number of independent variables, $T$ is the duration of the observation period, $q$ is the average flow of bicycle traffic at the research intersections, and $P_{min}$ is the estimated proportion of bicyclists faced with the least frequent tactical choice.

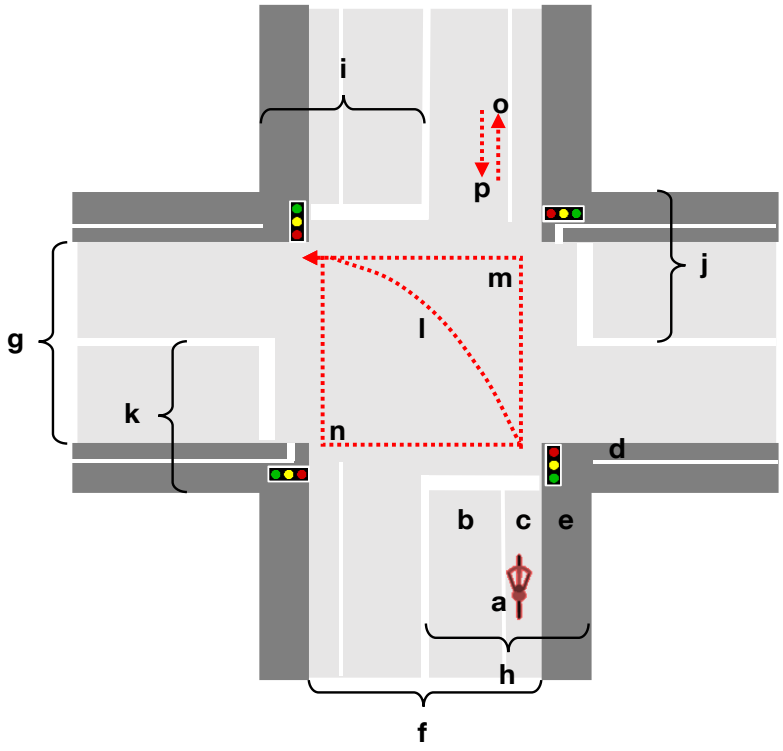
Two types of variables are necessary for the behavioural analyses and subsequent model development; variables describing the tactical behaviour of the bicyclist (dependent variables) and variables describing the situation (independent variables). In this research, automated methods for determining the manoeuvre, response to a red signal and pathway selection are developed and implemented. High-resolution data describing the geometry of the intersection

is required to relate the position of the bicyclist from the trajectory data to the physical infrastructure. Polygons are used to specify the contours of the various components of the intersection geometry, including bicycle facilities, roadways, sidewalks, crosswalks, grass areas and stop lines. The relation between the trajectory data and the geometric information enables the analysis of bicyclist position and infrastructure use. In addition to geometric information, data indicating the signal phase and timing are necessary. The timing data are synchronised with the video data to eliminate potential misalignment in the time stamps. This data makes it possible to relate the operational and tactical behaviour of the bicyclists to the state of the traffic signal.

The situational variables found in the literature review or hypothesised to influence the tactical behaviour of bicyclists are clustered into three categories:

- **Geometric variables** that describe the layout of a given intersection.
- **Traffic variables** that describe the average traffic flow and composition as well as the speed and position of other road users at a given time $t$.
- **Signal control variables** that describe the type of signal control used and the phase of the signal at a given time $t$.

A visual representation of selected independent and dependent variables is shown in Fig 3.2.




a = bicyclist
b = roadway
c = bicycle lane (on roadway)
d = bicycle lane (on sidewalk)
e = sidewalk
f = current road
g = crossing road
h = approach
i = opposite approach
j = right approach
k = left approach
l = direct left turn
m = indirect left turn
n = indirect left turn (wrong way)
o = with direction of travel
p = against direction of travel

**Fig 3.2** Selected tactical choices and situational factors

A total of 37 independent variables describing the geometry of the intersection, the momentary and average traffic and the signal control are specified. Complete lists of the categorical and continuous independent variables analysed along with descriptive statistics of the collected data set are given in Tab. 3.5 and Tab. 3.6 (page 70 and 71). Adequate variation in the variables is assured through the selection of research intersections with approaches (16 approaches in total) differing from one another in terms of traffic flow and geometry. Variables describing the presence of other road users as well as the state of the signal control at the moment a given bicyclist arrives at the intersection provide further variation between the cases.

As noted previously, data describing the personal characteristics of the observed bicyclists, including observable attributes such as age and gender and non-observable attributes such as bicycling experience and aggressiveness, are not collected. Similarly, the type of bicycle (e.g. electric motor support, cargo bicycles, bicycles with trailers, etc.) is not observed or recorded. Although these attributes play a significant role in the operational and tactical behaviour of bicyclists, they are omitted here because of the goal of creating generic bicycle behaviour models that can be calibrated for different types of bicyclists in the future.

### 3.1.2 Data specification for operational models

In order to calibrate and validate behavioural models on the operational level, data that catalogue the trajectories of a sample of bicyclists as they cross an intersection are required. Here, a trajectory is defined as a set of position coordinates $S_i = \{(x_i, y_i)_{t=0} \quad (x_i, y_i)_{t=1} \quad \cdots \quad (x_i, y_i)_{t=T_i}\}$, where $(x_i, y_i)_t$ is the position coordinate of the road user $i$ at a given point in time $t$ after tracking of the road user begins. For each tracked road user, a trajectory is created that contains a position coordinate for each sequential video frame (25 fps) in which the road user can be seen in the video. The length of the trajectory $S_i$ depends on the duration $T_i$ of the intersection crossing, both of which vary between trajectories depending on the pathway taken and the speed travelled by road user $i$. Speed and acceleration are derived from the position data in every time step. By collecting trajectory data from all road users, not only the movement of a specific individual can be studied, but also the interactions between individuals and surrounding road users.

Operational models are built using a set of input variables $X$, which can include, for example, the position and speed of an individual bicyclist and interacting road users as well as the traffic signal state, and a set of parameters that control the magnitude of influence of the given input variables. Once the model form is specified, the model parameters are estimated using Maximum Likelihood Estimation (MLE) to best fit a set of observed data. The sample size used for this calibration must be sufficiently large to produce statistically viable parameter estimates. The desired sample size is therefore dependent on the desired statistical power, the variance in the input variables, the inter-correlation between the model parameters and

the number of parameters to be calibrated. Without prior knowledge of the variance of the input variables in a population of bicyclists and the inter-correlation between model parameters, a general rule of thumb is followed to ensure an adequate sample size [LONG, 1997]. According to LONG [1997], calibrating models with fewer than 100 observations leads to unstable results and a lower threshold of 10 observations per parameter should be observed. Here, the number of observations available from each trajectory is related to the field of view of the camera (overall length in meters of the trajectory), the observation frequency and the speed of the bicyclist. Assuming an average speed of 5 m/s, a field of view of approximately 50 m by 50 m and a frame rate of 25 fps, an average of 250 observations per bicyclist is expected. This is more than double the recommended 100 observations required to achieve stable results.

## 3.2 Research intersections

The research intersections are selected with the goal of maximising the variety of situations with which the observed bicyclists are faced. One of the most important factors affecting bicyclist behaviour is the availability and type of bicycle infrastructure. Twelve types of bicycle infrastructure are defined in the German Road Traffic Regulations, Straßenverkehrs-Ordnung (StVO) [BUNDESMINISTERIUM FÜR VERKEHR BAU UND STADTENTWICKLUNG, 2013] and are listed in the right column of Tab. 3.1 (translation from German by the author). For this research, the types defined in the StVO are clustered into three groups based on the degree of separation between bicyclists and motorised vehicles (left column of Tab. 3.1).

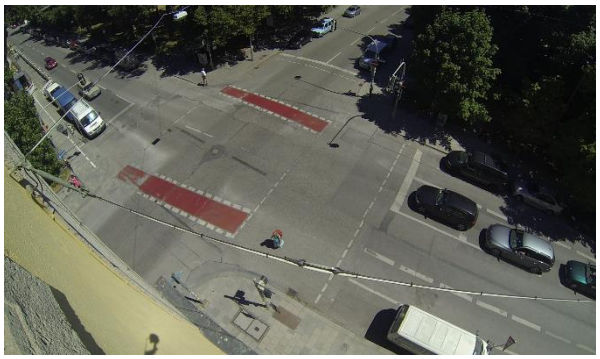
| Cluster | Type of Infrastructure |
|---|---|
| Type 1: None | Roadway without bicycle facility |
| | Restricted speed zone without bicycle facility (30 km/h) |
| | Roadway with traffic calming (without bicycle facility) |
| Type 2: On-road | Roadway with narrow bicycle lane (1.25 m – 1.5 m wide) |
| | Roadway with bicycle lane |
| Type 3: Separated | Physically separated bicycle path |
| | Bicycle path on sidewalk (bicycles and pedestrians separated) |
| | Bicycle path on sidewalk (bicycles and pedestrians mixed) |
| Type 4: Other | Bicycle path in green area |
| | Bicycle facility against the direction of travel on a one-way road |
| | Two-way bicycle facility |
| | Bicycle street (road with bicycle priority) |

**Tab. 3.1** Types of bicycle infrastructure defined by the StVO clustered based on degree of separation between bicyclists and motor vehicles

In the German specialised guidelines for bicycle traffic, the Empfehlungen für Radverkehrsanlagen (ERA) [FORSCHUNGSGESELLSCHAFT FÜR STRAßEN- UND VERKEHRSWESEN, 2010], five main categories of bicycle facilities are described in detail: roadway without bicycle facility, roadway with narrow bicycle lanes (Schutzstreifen), roadways with bicycle lanes, physically separated bicycle paths and two-way bicycle facilities. Special consideration is given in the ERA to combined bicycle and pedestrian facilities, bicycle facilities in bottlenecks, steep grades and bicycle facilities with bus or tram line considerations (tracks or stops). The clustering in the following table reflects the classification of bicycle infrastructure types in the ERA as well as in the StVO.

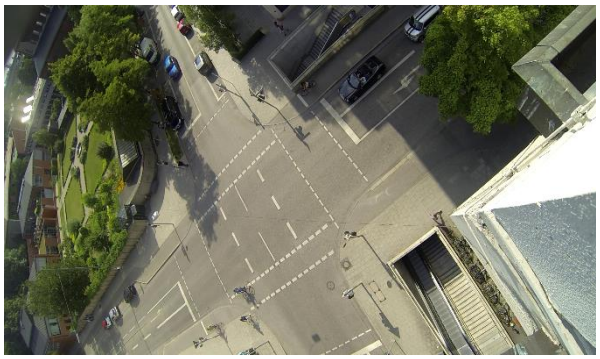
The first selection criteria for the research intersections is the type of bicycle infrastructure. For the purpose of model development, it is necessary to collect data at intersections with an approximately uniform representation of bicycle infrastructure of Type 1, Type 2 and Type 3. Type 4 is not investigated in this dissertation due to the specialised nature of these options and the wide variety of infrastructure forms. The second requirement for intersection selection is the variation of motor vehicle, bicycle and pedestrian traffic volumes between the research intersections. Finally, the research intersections are selected based on practical requirements concerning the video data collection. Such requirements include the presence of a tall building as near as possible to the intersection and access to the roof or a high window of that building. The characteristics of the four research intersections are given in Tab. 3.2 and the resulting descriptive statistics of the independent variables are given in Tab. 3.5 and Tab. 3.6 (page 70 and 71). The layout of the intersections can be found in Fig 3.3, where the camera views from the four research intersections are shown. The red areas that can be seen in Intersection 1 – Arcisstraße and Theresienstraße are road markings commonly used in Munich and other places in Germany to highlight the presence of bicyclists to motor vehicle drivers. The intersections are located within the city centre of Munich as shown in Fig 3.4.

1) Arcisstraße and Theresienstraße



2) Arnulfstraße and Seidlstraße

3) Karlstraße and Luisenstraße



4) Marsstraße and Seidlstraße

**Fig 3.3** Camera view from the research intersections (numbered as in Fig 3.4 and Tab. 3.2)

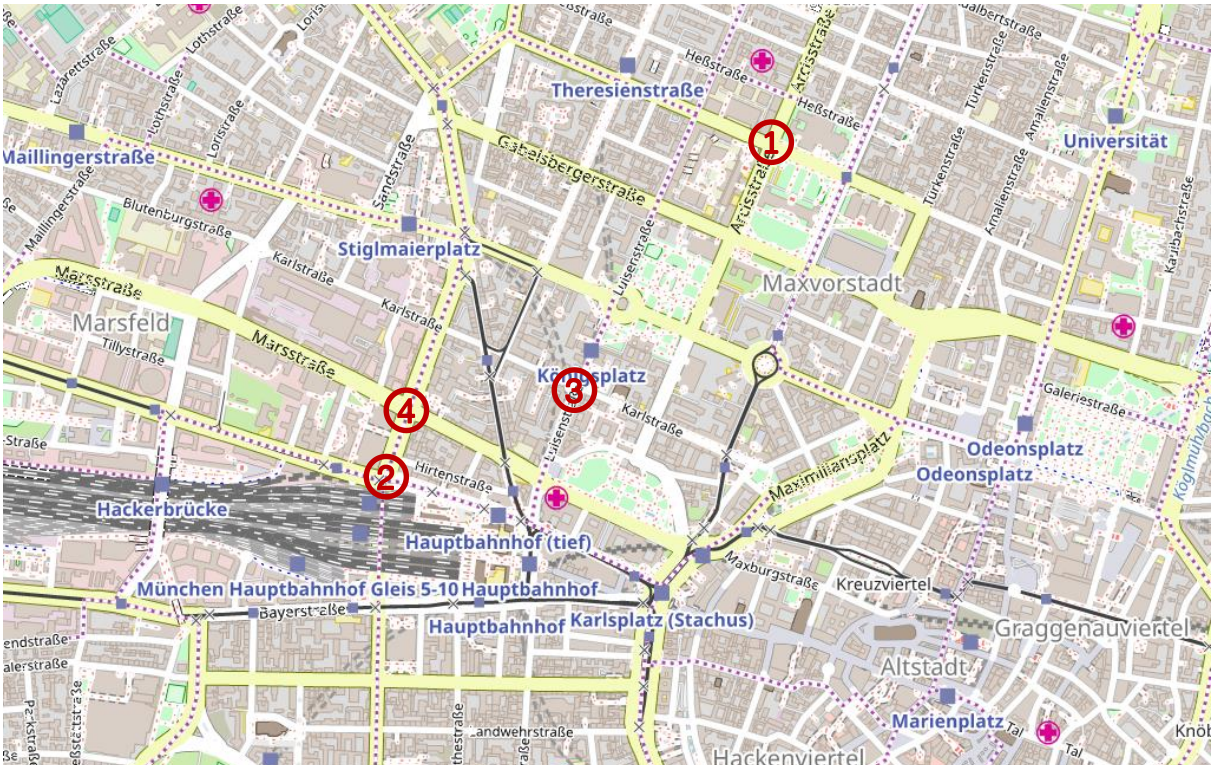



**Fig 3.4** Location of the research intersections (background: https://www.openstreetmap.org/)

| Approach | | Bicycle infrastructure | | Traffic Flow* (veh/h) | % HDV** and buses* | Pedestrian flow* (ped/h) | Bicyclist flow* (bicycle/h) |
|---|---|---|---|---|---|---|---|
| | | Type | Width (m) | | | | |
| Inter. 1 | North | Separated | 1.3 | 420 | 0 | 100 | 220 |
| | East | None | 0.0 | 780 | 6 | 40 | 80 |
| | South | On road | 1.8 | 170 | 7 | 80 | 55 |
| Inter. 2 | North | On road | 2.2 | 900 | 1 | 520 | 550 |
| | East | None | 0.0 | 240 | 0 | 1160 | 200 |
| | South | Separated | 1.6 | 910 | 4 | 40 | 660 |
| | West | Separated | 2.0 | 430 | 8 | 500 | 230 |
| Inter. 3 | North | Separated | 2.0 | 130 | 0 | 260 | 150 |
| | East | None | 0.0 | 140 | 0 | 160 | 70 |
| | South | Separated | 2.2 | 300 | 16 | 120 | 100 |
| | West | None | 0.0 | 120 | 9 | 100 | 100 |
| Inter. 4 | North | On road | 1.6 | 1080 | 9 | 240 | 410 |
| | East | Separated | 1.9 | 780 | 12 | 100 | 100 |
| | South | On road | 2.0 | 580 | 6 | 320 | 370 |
| | West | Separated | 1.6 | 530 | 5 | 100 | 120 |

* Extracted from collected video data

** Heavy Duty Vehicle

**Tab. 3.2** Characteristics of the research intersections

## 3.3 Video data collection

Video data were collected using a GoPro Hero3 Black Edition with a full HD resolution (frame size of 1920x1080 pixels) and a frame rate of 25 fps. A wide-angle lens was used to collect trajectory and situation data from the largest area possible using one camera. The distortion introduced by the use of a wide-angle lens was later rectified by post processing the trajectory data (see Section 3.4.2). Videos were recorded during the spring and summer months of 2013 and 2014 for between two and four days per intersection. The observation period began at approximately 7:00 am and ended at about 7:00 pm, depending on the access to the roof of the buildings. Video data was saved on SD card with 128 GB of storage space, which provided enough space for approximately 14 hours of full HD video data. Power was provided by an external battery pack with a capacity of 13 000 mAh, which was sufficient for the 14 hours of data collection. The data collection system is shown in Fig 3.5.

**Fig 3.5** Data collection system

From the total volume of data, a shorter duration of video data was selected for trajectory extraction and data processing. The duration of the observational study at each intersection is set in consideration of the number of tactical choice observation necessary to calibrate the logistic regression models.

$$T = \frac{20\ IV}{q\ P_{min}} = \frac{20 \cdot 37}{3415 \cdot 0.2} = 1.08\ h \qquad \text{Eq. 3.3}$$

where the total bicycle flow per hour from Tab. 3.2 is 3415 bicycle/h (sum across all approaches of all intersections) and the estimated average left turning rate is 0.2. This tactical behaviour is selected to calculate the duration of the observation study (Eq. 3.3) because it is the least frequently observed tactical choice in the video data. In order to guarantee a sufficiently large sample size, two hours of video data (almost double the minimum duration from Eq. 3.3) from the morning peak hour, about 7:30 am to 9:30 am, were selected for each intersection for trajectory extraction. Video data were collected for between two to four days at each intersection. From the available morning peak data, a segment of video was selected based on the camera angle, which differed slightly for each data collection day, the lighting conditions, which were deemed best if the sun did not create distinct shadows, and the absence of wind disturbances.

The City of Munich provided data from the traffic signals at three of the four research intersections for the data collection period. Each of these signals is traffic actuated and information regarding the time of each phase transition is automatically catalogued (1-second precision). Intersection 4 is controlled by a fixed-time signal control and therefore the phase change time data is not recorded.

## 3.4 Automated video analysis[4]

There are two methods for extracting trajectories from raw video data, manual processing or automated extraction. In manual processing, the position of each road user is marked by an analyst at a given frequency. Proponents of this method argue that the resulting trajectory data is highly accurate because issues with automated processes, such as occlusion, false classification and over and under grouping of features, are avoided. In addition, qualitative assessments of the situation can be incorporated into the data set. For example, the analyst can watch the video data and extract trajectories only in situations deemed critical in terms of safety. A number of tools exist to aid analysts in the manual extraction of trajectories. An example of which is the tool T-Analyst developed at Lund University [FACULTY OF ENGINEERING LTH, 2015]. However, manual extraction of trajectory data is extremely time-consuming and researchers are limited in the number of observations that can be generated using this method. As in all manual data collection processes, human error is also an issue.

Automated methods for trajectory extraction use image processing techniques to identify and track moving objects in sequential video frames. Proponents of automated approaches for trajectory extraction point to the significantly larger sample of trajectories that can be extracted and processed as well as the elimination of human error in the extraction process. Several issues with automated video processing, including occlusion (full or partial blocking of one object by another object), problems with shifting shadows and lighting, as well as difficulties in extracting the trajectories of single road users when a group moves together (particularly pedestrians and bicyclists), account for high variance in the accuracy of the resulting trajectories. However, improvements in the field of computer vision offer increasing accuracy in automated extraction methods.

In consideration of the large sample of trajectories required for model development, automated trajectory extraction is used for this research.

### 3.4.1 Trajectory extraction

There are four main approaches for detecting and tracking moving objects in video data:

1. **Region Based Tracking** in which an area of pixels is identified, often using background subtraction, and tracked through the video frame.

[4] Sections of this text were originally published in the paper:

TWADDLE, H.; SCHENDZIELORZ, T.; FAKLER, O. & AMINI, S. [2014]: Use of automated video analysis for the evaluation of bicycle movement and interaction. In Proc. SPIE 9026, Video Surveillance and Transportation Imaging Applications 2014. San Francisco. [TWADDLE, SCHENDZIELORZ, FAKLER, ET AL., 2014]

2. **Contour Based Tracking,** which is similar to region based tracking except that only the outline, or contour, of a group of moving features is tracked, reducing the computational effort.

3. **3D Model Based Tracking** is a method in which the algorithm locates objects in the video frame based on their similarity to a provided road user model.

4. **Feature Based Tracking** in which distinctive attributes (features), such as corners and edges, that move incrementally between subsequent video frames are tracked. The tracked features are grouped into road user hypotheses in a second step[5].

These approaches have been implemented in a wide variety of software tools developed for trajectory extraction. The majority of these tools are developed independently by researchers and traffic analysis teams and are not explicitly available for use by other analysts. Two companies were identified in the review that offer video processing and trajectory extraction as a service, MIOVISION [2016] and DATAFROMSKY [2015]. Two research institutions were found that offer trajectory extraction as a service, the University of British Columbia [SAYED ET AL., 2016; PIN ET AL., 2015] and the German Aerospace Center [LOCE & SABER, 2015; LEICH ET AL., 2015; JUNGHANS ET AL., 2015]. A single open source software for trajectory extraction and processing was found during the review, *Traffic Intelligence* [JACKSON ET AL., 2013; SAUNIER & SAYED, 2006]. This software implements a feature-based algorithm developed by Nicolas Saunier and his collaborators and was first published in 2006 [SAUNIER & SAYED, 2006]. It is currently available in a Bitbucket code repository called *Traffic Intelligence* [SAUNIER, 2016], which is continually updated as Nicolas Saunier and his team at the Polytechnique Montréal improve and extend the code.

*Traffic Intelligence* is selected based on an evaluation of the functionality of the software. The cost effectiveness of using an open source software and the potential for adapting and extending the software also play an important role in the selection.

The main functionality of the software, feature-tracking and feature-grouping, is available as a C++ code, which is developed using the open source library OpenCV (Open Source Computer Vision Library). The trajectory data extracted from the videos are stored in an SQLite database with four tables. The first table *objects* contains the road user identification number, the type of road user and the number of road users grouped into the identification

[5] A detailed summary of image processing and trajectory processing is out of the scope of this dissertation. However, if the reader is interested in obtaining more information the book *Image processing, analysis and machine vision* [SONKA ET AL., 2014] and the paper *Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis* [SIVARAMAN & TRIVEDI, 2013] are recommended.

number. The second table *objects_features* contains the feature-grouping information. Each of the tracked features is associated with an object id number listed in the *objects* table. The third table *positions* and the fourth table *velocities* contain the position and velocity vectors, respectively, for each feature listed in the *objects_features* table for each video frame in which it was tracked. Many Python scripts are available in the *Traffic Intelligence* code repository that access the data stored in the SQLite database and calculate a variety of parameters describing the behaviour and interactions between road users. On the right side of Fig 3.6 (a) the trajectories of all the tracked features belonging to one pedestrian are displayed. The object trajectory that results when the trajectories of the features belonging to the same object are grouped together is shown in Fig 3.6 (b).

a) Tracked features: b) Grouped features (objects):

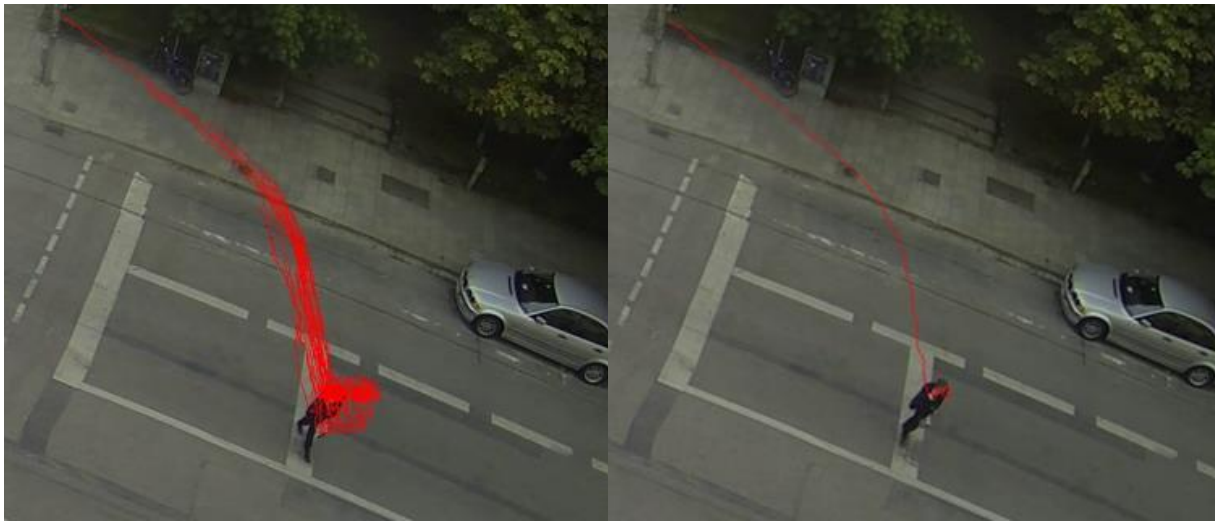


**Fig 3.6** An example of tracked (a) and grouped (b) features

The feature motion is measured from the video data in pixels. The position and speed vectors are given using an arbitrary coordinate system based on the video frame, where the point $0\ px, 0\ px$ is located in the upper left hand corner of the frame. This information is translated to a real world local coordinate system using a homography matrix that rotates, scales and translates the points in the video frame to real world points. More information about using homography matrices to find a perspective transformation between two plans can be found on the Open CV website [OpenCV, 2015]. The units are changed from pixels to meters (projection from video frame to real world coordinates) using a meters per pixel factor from the real world image.

The feature-based tracking algorithm is controlled using 22 parameters, 16 of which are used for feature-tracking, while the other six direct feature-grouping[6]. The feature-tracking parameters control the quality and dynamic characteristics of the features in the video frame that are to be tracked. In this context, a feature is a corner or a small patch of pixels with a strong gradient in the video frame that can be re-identified in subsequent frames. In road user tracking, features can be heads and hands of pedestrians, license plates and side mirror of cars and pedals and wheels of bicycles, amongst other distinguishable points or lines. Through the systematic altering of the control parameters, six parameters were found to play a particularly important role in feature-tracking: the maximum number of features tracked, the minimum feature quality, the size of the search window at each pyramid level, the displacements to test minimum feature motion, the minimum displacement to keep features and the maximum feature acceleration. The parameters must be adjusted in response to the height and angle of the video camera, the lighting conditions as well as the physical and dynamic characteristics of the road users to be tracked.

Optimal parameters were found through the systematic adjustment of the parameters and evaluation of the tracking results. The results of the feature tracking were assessed based on two measures; the portion of bicycles, pedestrians and motorised road users tracked with a minimum of three features, and the delay between the instant the road user began moving and the start of trajectory tracking. The parameter set in Tab. 3.3 was found to be optimal for the four research intersections. Using the given values for the feature-tracking parameters, at least three features are tracked for the vast majority of road users. Values for slightly over sensitive tracking are selected in order to gain information from even the hardest to track road users. For this reason, most road users are tracked not only by one feature but by hundreds of features.

| Parameter | Value |
|---|---|
| Maximum number of features tracked | 1200 |
| Minimum feature quality | 0.005 |
| Size of the search window at each pyramid level | 8 |
| Displacements to test minimum feature motion | 2 |
| Minimum displacement to keep features | 0.005 |
| Maximum feature acceleration | 4 |

**Tab. 3.3** Selected parameters for feature tracking

[6] Refers to the parameters used in the version of *Traffic Intelligence* available during data analysis in 2013.

In order to analyse the resulting trajectory data in a meaningful way, the feature trajectories must be clustered into road user hypotheses. Many qualities of the feature trajectories, including the distance between the features at the first instants of simultaneous tracking, the minimum speed and the segmentation distance (the difference between the maximum and minimum distance between the features over time), are taken into account in feature grouping. In situations with many different types of road users, it is very difficult to find a set of feature grouping parameters that provide sufficient results for all types of road users. The minimum feature speed and the maximum distance between features are significantly different when tracking pedestrians as opposed to large motor vehicles such as buses and trucks. In addition, the rigid nature of motorised vehicles enables one to achieve very good grouping results by allowing only very small differences in the direction of movement of the features. In contrast, different features belonging to a single pedestrian or bicyclist can move in significantly different directions. Consider, for example, the swaying arms of a pedestrian or the incongruent circular motion of a bicyclist's feet as they pedal.

As the analyses are carried out at busy urban intersections with considerable personal vehicle, heavy vehicle, pedestrian and bicycle traffic, a method for resolving this issue is proposed and implemented. The tracked features are pre-classified into two groups, probable motor vehicles and probable bicyclists or pedestrians, based on the positions along the entire trajectory. This is done using two types of polygons, touch-at-any-instant and touch-in-all-instances polygons, as shown in Fig 3.7.

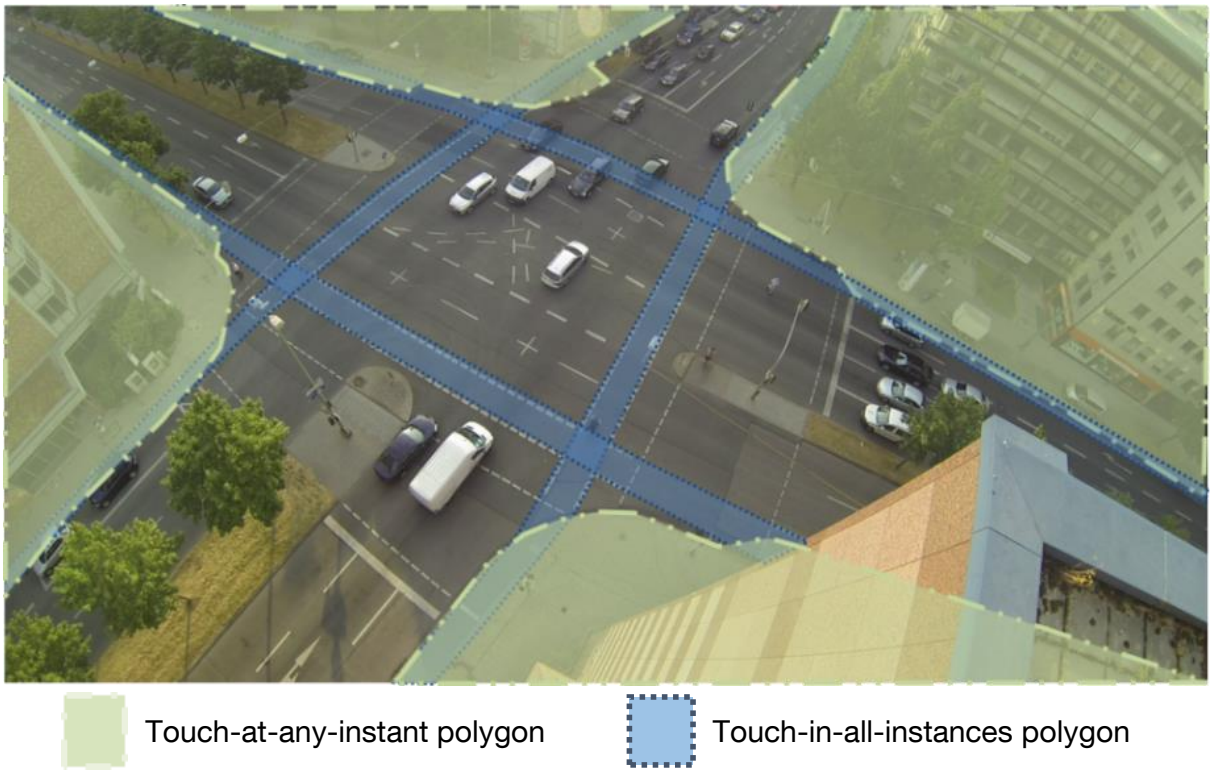



**Fig 3.7** Polygons for pre-classification of tracked features

As the names suggest, if a feature touches one of the touch-at-any-instant polygons at any time during the trajectory it is classified as a feature belonging to a bicyclist or pedestrians. The second type of polygon, touch-in-all-instances, is used to classify road users that are tracked entirely within the bicycle facility as bicyclists.

Pre-classified features are filtered into two separate SQLite databases, one for probable motor vehicles and one for probable pedestrians and bicyclists. Sets of feature-grouping parameters can be subsequently specified for each of the two groups. Feature-grouping parameters are selected using the same systematic altering of parameters that was used for the feature-tracking parameters. Five parameters were identified that had the most evident effect on the grouping results (Tab. 3.4). Each of these parameters was adjusted independently, starting at an extreme value (high or low) and then increasing or decreasing that value until the optimal results were achieved on a one-minute test video segment. The quality of the results was determined qualitatively by assessing the number of over-segmented or over-grouped road users. Over-grouping refers to the grouping of features belonging to multiple road users into a single road user, while over-segmentation occurs when the features from one road user are grouped as multiple road users. Unfortunately, parameter adjustments that reduce over-grouping tend to increase over-segmentation. Inversely, if over-segmentation is minimised, over-grouping will tend to increase. This is particularly true when objects with different physical and dynamic characteristics are analysed together. Although the separation of features into two databases with objects of similar size made it possible to correctly group features and minimise the occurrence of over-grouping and over-segmenting, such errors could not be completely eliminated. As the manual correction of over-segmentation is much easier than that of over grouping, parameters were set to favour over-segmentation.

| Parameter | Bicycle/ Pedestrian | Motor Vehicle |
|---|---|---|
| Minimum number of frames to consider a feature for grouping | 40 | 30-45 |
| Connection distance (distance at first instant) (m) | 0.8 | 1.9-2.1 |
| Segmentation distance (difference between max. and min.) (m) | 0.4-0.5 | 0.5-0.7 |
| Maximum distance (m) | 0.8 | 1.9-2.1 |
| Minimum average number of features per frame | 1 | 1-3 |

**Tab. 3.4** Selected parameters for feature grouping

Once feature-grouping is complete, the bicyclists and pedestrians stored in a single SQLite database are classified based on the maximum speed travelled during the intersection crossing. A maximum speed for pedestrians of 9 km/h and a maximum speed of bicycles of 30 km/h is used. The accuracy of the methodology used to track and classify cars, bicycles and pedestrians varies widely between the research intersections. The percentage of tracked

road users ranges from 72%-93%, while the percentage of correctly classified road users ranges from 28%-97%. The most accurate results were obtained for cars and bicycles at Intersection 4 Marsstraße and Seidlstraße. A number of factors contributed to the better results obtained at this intersection. First, the camera was installed at a relatively high vantage point, which enables better tracking of moving features. Second, the existence of bicycle facilities in all directions, as well as the high proportion of bicyclists that use this infrastructure made the hypothesis of road user type based on the position of the moving features quite accurate. Finally, the relatively large flow of vehicular traffic was found to restrict bicycle movement and behaviour to predictable patterns (use of bicycle facilities, crossing at intended times and locations), which again made road user classification more accurate.

The percentage of missed bicycles and pedestrians at Intersection 1 Arcisstraße and Theresienstraße (28%) is much higher than at the other intersections. This could be due to the relatively low vantage point (14 m). Another problem at this intersection is the presence of a cable used to hang a street light over the intersection. This cable hangs directly between the camera and the intersection and causes issues regarding the grouping of features and classification of road users. Additionally, during data collection, a construction site on Arcisstraße blocked the use of the bicycle facility on the north approach. As a result, many bicyclists used the road with the motorised traffic and were therefore categorised as cars.

### 3.4.2 Trajectory post processing

A wide-angle lens was used to collect information from road users while approaching, crossing and exiting the intersections. The image distortion associated with wide-angle lenses, however, leads to the analogous distortion in the trajectories, which must be rectified before or after extraction. A method for post-processing trajectory data that has already been extracted from the video data is developed and implemented in this dissertation. The intrinsic and extrinsic parameters of the GoPro camera within the waterproof plastic casing are estimated using OpenCV. The resulting Camera Matrix ($K$) and Distortion Coefficients ($d$) are given below.

$$K = \begin{bmatrix} 2129.02613 & 0.0 & 950.917979 \\ 0.0 & 2102.04718 & 535.95449 \\ 0.0 & 0.0 & 1.0 \end{bmatrix} \quad \text{Eq. 3.1}$$

$$d = [-0.898305625 \quad 0.606958917 \quad 0.0 \quad 0.0 \quad 0.0] \quad \text{Eq. 3.2}$$

The video frame is rectified using the estimated parameters. The rectified frame (Fig 3.8 b) is used to create a second homography matrix that describes the perspective transformation between the rectified video image and the real world map. The trajectories of the tracked road users are corrected using the following steps:

1. Return the trajectory positions to the original, distorted image positions using the inversion of the original homography matrix for all road users.

2. Remap the position coordinates from the original, distorted image to the rectified camera image using the estimated camera parameters (Eq. 3.1 and Eq. 3.2). This is done using the *undistortPoints* function belonging to the OpenCV Image Processing library [OPENCV, 2015]. This function takes an array of points in the original image and remaps them to the rectified image.

3. Transform the positions back to real-world coordinates using the second homography matrix relating the rectified camera image to the real world image.

4. Recalculate the velocity vectors using the rectified position points.

a) Distorted camera view b) Rectified camera view

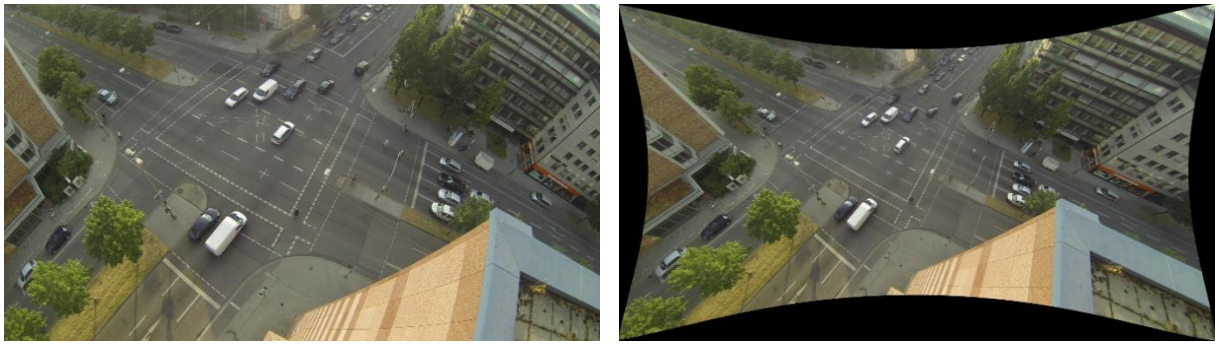


**Fig 3.8** View from the video camera a) before and b) after rectification of distortion

In addition to the trajectory rectification method, an algorithm was developed for merging segmented trajectories belonging to the same road user[7]. *Traffic Intelligence* tracks moving road users and as such, when a road user stops, which occurs frequently at signalised intersections, the track is temporarily lost. When the road user begins moving again, *Traffic Intelligence* again tracks the road user but using a different identification number. Trajectories are also disconnected when a road user passes behind an obstacle or another road user, temporarily blocking it from the view of the camera (occlusion). Even seemingly insignificant occlusion instances, such as when a road user passes under a thin electric cable, can lead to segmented trajectories. A simple rule-based algorithm that identifies trajectories that start or end within the inner region of the video frame and matches these trajectories based on

[7] Many of the functionalities developed here, such as rectification and classification, are now available from *Traffic Intelligence* (implemented slightly differently). These features were not yet available at the time of this work (2013).

temporal, spatial and dynamic similarity was developed and tested. The testing results showed that the algorithm performed very well for motor vehicles and lone pedestrians and bicyclists. However, in cases where two or more pedestrians or bicyclists bunched together, which often occurred at stop lines in the collected video data, the algorithm was not capable of correctly re-identifying the same road user.

A final manual correction of the trajectory databases was deemed necessary despite the efforts in automated data post processing. Although the developed classification algorithm was quite successful at identifying the road users as motorised vehicles, bicyclists or pedestrians, incorrect classifications skew the results of the subsequent behavioural and interaction analyses. The automated method developed for merging segmented trajectories for the same road user was found to be unsatisfactory, particularly in situations with high-density pedestrian or bicycle traffic. Therefore, the extracted trajectories were displayed over the raw video data and manually controlled for inconsistencies. Misclassified road users were corrected, fragmented trajectories belonging to the same road user were merged and erroneous or superfluous trajectories were deleted. This process was found to be tedious and time-consuming (200 hours for 8 hours of trajectory data and 5146 bicycle trajectories) but worthwhile for the increased accuracy and usability of the trajectory data. The effort required to correct the trajectory database is deemed considerably less than the effort that would be necessary to manually trace trajectories for the same number of bicyclists.

## 3.5 Database extension

The resulting trajectory database was extended to include qualitative behavioural variables as well as variables to describe the current situation that were difficult to collect using automated methods. This was done by watching the videos and recording the variables for each bicyclist.

The exact timings of the signal phase transitions at three of the research intersections were obtained from the City of Munich. These three intersections are controlled with a traffic-actuated signal and the data from the traffic sensors as well as the signal timing changes are recorded by the City of Munich. An older signal controller, which uses a fixed signal plan, is installed at Intersection 4, Marsstraße and Seidlstraße. It is not possible for the City of Munich to record the timing of the phase changes from this type of controller. It was, however, possible to obtain the fixed signal plan for the intersection and manually synchronise this information with the behaviour observed in the video data. This manual synchronisation is associated with a minor error because it is not possible to determine exactly when the traffic signal turns but only infer when this change occurs from when the road users start moving after they are given a green signal. The exact and inferred phase change timing information is linked with the trajectory database by matching the time log of the video data and the signal phase timing information. For this dissertation, the phase timing information is stored in a

comma separated value file for each of the intersections. The signal phase timing information, which is given in hours, minutes and seconds, is translated to the associated video file name and frame number. The given signal phase for any frame can then be extracted from the database.

Complete lists of the independent variables collected using automated and manual data processing measures are shown below. This information is used in Section 5.1 to estimate logistic regression models for predicting the tactical choices of bicyclists. The categorical independent variables along with the number and percentage of observations in each category are listed in Tab. 3.5. The continuous independent variables with descriptive statistics measured from the collected video data are shown in Tab. 3.6.

| Independent variable | Category 1 | Category 2 | Category 3 |
|---|---|---|---|
| Strategic / prior tactical choice | | | |
| Manoeuvre | Straight<br>N=4040 (80.4%) | Right<br>N=454 (9.0%) | Left<br>N=534 (10.6%) |
| Infrastructure selection | Bicycle facility<br>N=3532 (94.8%) | Roadway<br>N=67 (1.8%) | Sidewalk<br>N=128 (3.4%) |
| Geometry | | | |
| Bicycle facility | None<br>N=634 (12.4%) | Bicycle facility<br>N=4485 (87.6%) | |
| Bicycle facility type | None<br>N=634 (12.4%) | On-road<br>N=2070 (40.4%) | Separated<br>N=2415 (47.2%) |
| Parking | None<br>N=2268 (44.3%) | Parking<br>N=2851 (55.7%) | |
| Left turn lane | None<br>N=2484 (48.5%) | Left turn lane<br>N=2635 (51.5%) | |
| Centre island | None<br>N=1729 (33.8%) | Centre island<br>N=3390 (66.2%) | |
| Traffic | | | |
| Right lane occupancy | No<br>N=1617 (33.5%) | Yes<br>N=3214 (66.5%) | |
| Signal control | | | |
| Signal phase | Red/Yellow<br>N=2817 (55.6%) | Green<br>N=2253 (44.4%) | |
| Specific bicycle signal | Shared signal<br>N=2438 (47.6%) | Bicycle signal<br>N=2681 (52.4%) | |

**Tab. 3.5** Description of categorical independent variables

| Independent variable | Unit | Mean | Std.Dev. | Min. | Max. |
|---|---|---|---|---|---|
| Geometry | | | | | |
| Bicycle facility width | m | 1.5 | 0.7 | 0.0 | 2.2 |
| Sidewalk width | m | 3.5 | 1.6 | 0.9 | 9.3 |
| Roadway width (approach) | m | 7.9 | 3.0 | 0.0 | 12.0 |
| Roadway width (opposite approach) | m | 5.3 | 1.7 | 0.0 | 10.9 |
| Driving lanes (approach) | - | 2.4 | 0.9 | 0 | 4 |
| Driving lanes (opposite approach) | - | 1.6 | 0.6 | 0 | 2 |
| Total roadway width (current road) | m | 18.3 | 7.0 | 8.8 | 28.6 |
| Total roadway width (crossing road) | m | 16.8 | 5.5 | 8.8 | 28.6 |
| Total driving lanes (current road) | - | 4.0 | 1.4 | 2 | 6 |
| Total driving lanes (crossing road) | - | 3.6 | 1.2 | 2 | 6 |
| Traffic | | | | | |
| Cars in approach | - | 2.5 | 2.0 | 0 | 10 |
| Trucks in approach | - | 0.1 | 0.4 | 0 | 3 |
| Pedestrians in approach | - | 1.3 | 1.8 | 0 | 20 |
| Bicyclists in approach | - | 1.6 | 2.1 | 0 | 16 |
| Traffic volume (approach) | veh/h | 646.6 | 309.7 | 0 | 1800 |
| Traffic volume (crossing road) | veh/h | 523.9 | 286.6 | 0 | 1800 |
| Percentage HDV and buses (approach) | % | 4.9 | 4.0 | 0 | 16 |
| Percentage HDV and buses (crossing road) | % | 5.6 | 4.2 | 0 | 16 |
| Bicyclist volume (approach) | bicycle/h | 337.7 | 201.3 | 0 | 660 |
| Bicyclist volume (crossing road) | bicycle/h | 199.4 | 159.2 | 0 | 660 |
| Pedestrian volume (approach) | ped/h | 255.9 | 238.2 | 0 | 1160 |
| Pedestrian volume (crossing road) | ped/h | 329.5 | 358.3 | 0 | 1160 |
| Signal control | | | | | |
| Time since last phase change (red) | s | 25.5 | 16.2 | 0 | 98 |
| Time since last phase change (green) | s | 14.8 | 11.8 | 0 | 102 |
| Time until next phase change (red) | s | 28.0 | 17.9 | 0 | 91 |
| Time until next phase change (green) | s | 17.2 | 16.0 | 0 | 103 |
| Phase length | s | 43.3 | 17.2 | 7 | 104 |

**Tab. 3.6** Description of continuous independent variables

# 4. Trajectory clustering

Although bicyclists employ different strategies to navigate across an intersection, it is presumed that common pathway types or shapes emerge when the trajectories of many bicyclists are analysed together. Two clustering approaches are developed here to identify common pathway types for subsequent use in the modelling of the operational and tactical behaviour of bicyclists. First, a clustering algorithm that is approach specific is used to identify the various pathways used by bicyclists at a given approach to cross the intersection. The number and form of these pathways depend on many factors, including the layout of the intersection, the availability and type of bicycle infrastructure, characteristics of the traffic signal, static and dynamic traffic attributes and the strategic routes of bicyclists. These clustered pathways are used in Section 5.2 to simulate operational behaviour. A second clustering approach is used to identify the general strategies used by bicyclists to carry out their desired manoeuvre. For example, bicyclists turning left may implement a direct or an indirect left turn. This clustering approach is generic and does not depend on the specific characteristics of the given approach or intersection. The results are used in Section 5.1 to develop tactical behaviour models.

Although it is possible to manually examine trajectory data and designate a structure to classify the pathways, this process is subjective and can be work intensive. Furthermore, the classification structure designated for one intersection may not be transferable to other intersections. Once the classification structure is defined, the observations can be classified using manual or supervised learning approaches. The effectiveness of both strategies is highly dependent on the suitability of the classification structure.

Another method for clustering the pathways used by bicyclists to cross an intersection is unsupervised learning. Clustering algorithms, otherwise known as unsupervised learning algorithms, assess the similarity of observations in a dataset and use various methods to separate these observations into clusters. The main advantage of clustering is that the practitioner does not have to make prior assumptions about the classification structure, resulting in data driven and non-subjective clusters. However, the quality of the clustering results is highly dependent on the amount of data, data accuracy, the shape, type and scale of the features used and the selection of an appropriate clustering algorithm. There are five main steps in clustering outlined by JAIN & DUBES [1988]:

1. **Pattern representation:** Each observation is described using a vector of features. Feature selection is carried out by determining an effective number, type and scale of the original features of the observation. Feature extraction is the transformation of one or more original features to create salient features that have added value in describing the pattern in the observations.

2. **Proximity:** The measure used to evaluate the similarity between observations in the dataset is selected in this step. A common metric used to discern similarity is the Euclidean distance, but many other metrics exist.
3. **Clustering:** A wide variety of algorithms exist to assess the similarity between the vector representations of the observation set and form cluster hypotheses. Observations are assigned to a cluster using a hard or fuzzy approach. Hard clustering assigns the observation to one of the clusters while fuzzy clustering estimates the probability of an observation belonging to each of the clusters.
4. **Abstraction:** Here, a simplified representation of the clustering structure, which is useful for further machine learning applications or human interpretation, is taken from the clustering results or is generated.
5. **Evaluation:** The effectiveness of the clustering approach is assessed using an externally defined classification structure, an internal examination of the validity of the resulting clusters or a relative comparison between two clustering approaches.

In order to avoid the subjective, specific and work-intensive manual definition of a classification structure for the pathway selection of bicyclists crossing intersections, an unsupervised learning approach is implemented and assessed. A first method is developed in Section 4.1 to cluster the pathways used by bicyclists at a given approach of a given intersection. A similar method is used to generically cluster the type of pathways selected by bicyclists to carry out their manoeuvre, such as direct or indirect left turns, regardless of the approach (Section 4.2). The steps and the notation defined by JAIN & DUBES [1988] will be used in the subsequent sections to develop the clustering approaches.

## 4.1 Approach specific pathway clustering

The approach specific clustering method examines the trajectories from bicyclists at a specific approach of a specific intersection. The objective is to identify representative pathways that are used by bicyclists to carry out their desired manoeuvres. For example, bicyclists turning left at a given approach may typically use one of three pathways to perform the turn. At another approach, only two pathways may be used typically by bicyclists to turn left. The specific shape of these pathways differs depending on the geometry and possibly other characteristics of the approach and intersection. The steps in the clustering approach are described in detail in Section 4.1.1 to 4.1.5. The representative trajectories of the resulting clusters are used subsequently in Section 5.2 to provide a desired pathway to the simulated bicyclists.

### 4.1.1 Pattern representation

Pattern representation involves selecting or extracting features that describe a particular observation, where $f_{ij}^*$ is feature $j$ for observation $i$. Here, $N$ is the total number of observations in the dataset and $d$ is the total number of features selected or extracted from each of the observations. Features can be continuous or categorical variables that define particular aspects of an observation. Feature selection involves choosing directly measured variables. Feature extraction is the augmentation of original features to provide information in another form to the clustering algorithm. The selection and extraction of features that adequately describe the data and include important attributes that make it possible to separate the observations into desirable clusters is a crucial step in the clustering process. All features pertaining to observation $i$ are arranged in a one dimensional feature vector $F_i^*$ (the asterisk denotes that the features are raw and have not been normalised). The feature vectors from all observations in a dataset are combined to form a pattern matrix $\mathcal{A}^*$.

$$\mathcal{A}^* = [F_{i=1}^* \quad F_{i=2}^* \quad \cdots \quad F_{i=N}^* ]^T = \begin{bmatrix} f_{11}^* & f_{12}^* & \cdots & f_{1d}^* \\ f_{21}^* & f_{22}^* & \cdots & f_{2d}^* \\ \vdots & \vdots & \ddots & \vdots \\ f_{N1}^* & f_{N2}^* & \cdots & f_{Nd}^* \end{bmatrix} \quad \text{Eq. 4.1}$$

In this clustering application, the observations include raw trajectory data, which have the form:

$$S_i = \{(x_i, y_i)_{t=0} \quad (x_i, y_i)_{t=1} \quad \cdots \quad (x_i, y_i)_{t=T_i}\} \quad \text{Eq. 4.2}$$

where $(x_i, y_i)_t$ is the position coordinate of the road user $i$ at a given point in time $t$ after tracking of the road user began. Each trajectory contains a position coordinate for each video frame during the tracking of that road user (25 fps in this case). The length of the trajectory $S_i$ depends on the duration $T_i$ of the trajectory, which varies between trajectories depending on the pathway taken and the speed travelled by road user $i$. In addition, the starting and ending point of the trajectories differ in space depending on where in the video frame the automated tracking of the road user began and ended.

In a first step, the trajectory data is filtered and reshaped to increase the comparability of the trajectories. Short trajectories, which occur when a road user is only tracked during a portion of the intersection crossing, are removed from the dataset. This is done using a polygon representing the middle section of the intersection (blue polygons in Fig 4.1). Trajectories that start or end within this polygon, as well as trajectories with a cumulative length (Eq. 4.4) less than 10 m, are removed. The remaining trajectories are trimmed using a second polygon, shown in green in Fig 4.1. Position points at the beginning and end of the trajectory that are outside polygon are removed from the trajectory. The first and last position points that fall inside the polygon act as the new start and end of the trajectory. Position points along this trimmed trajectory that fall outside the polygon are not removed. This trimming enables

comparability between the start and end points of the trajectories. The feature selection and extraction is subsequently carried out using the filtered and trimmed trajectory dataset.

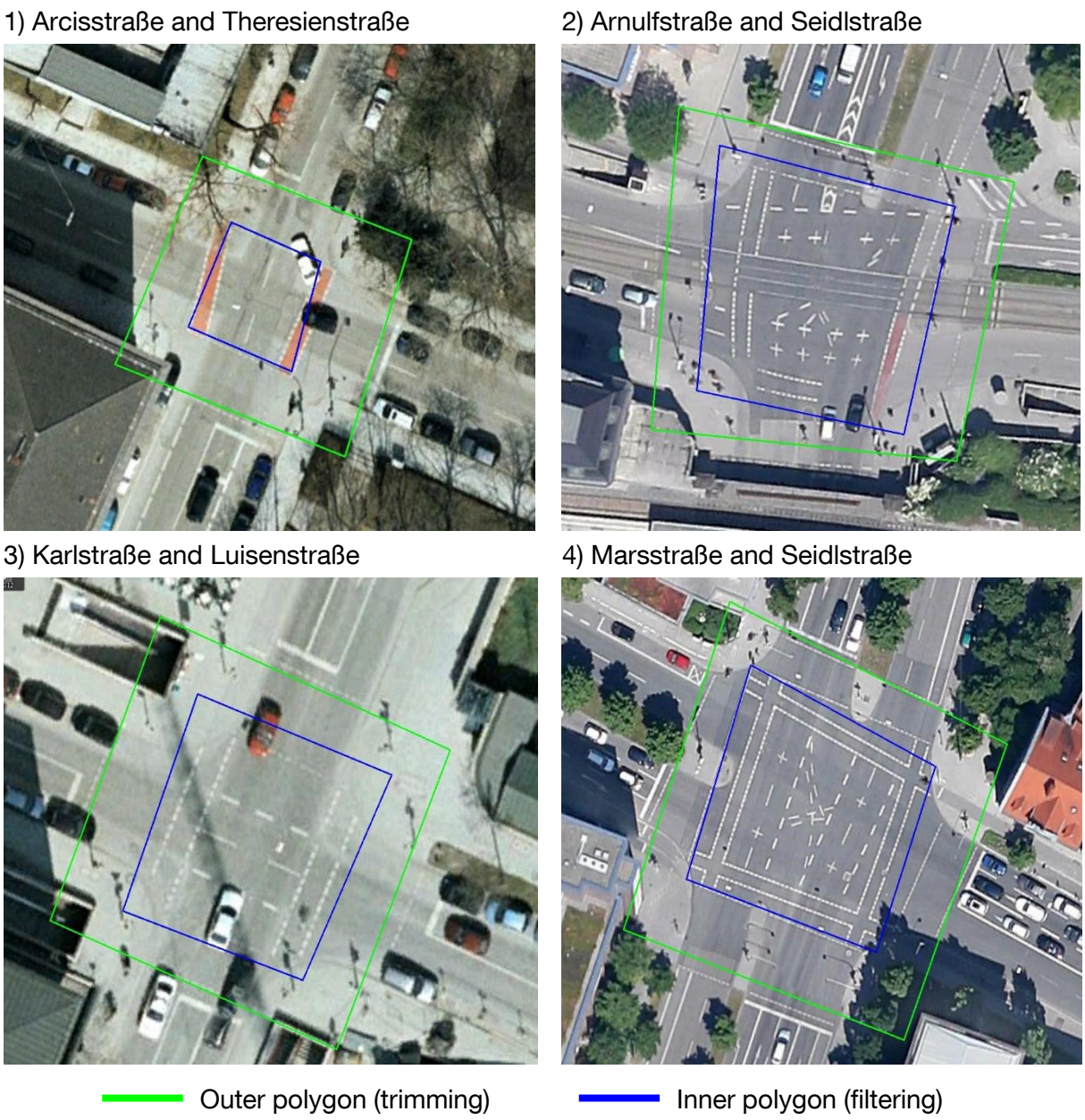



**Fig 4.1** Polygons for trajectory filtering and trimming (background images: Google Earth 2013)

In order to implement standard unsupervised clustering techniques, the number of features in the observation vectors must be uniform across the set of observations. This is not true for the trajectories, which range in length depending on the duration of the intersection crossing. To create observations with the same number of features, trajectories are reduced to a one-dimensional vector containing the position data from a given number of position points. There are many options for selecting the position points from the complete trajectory $S_i$. The most evident of which is to divide the trajectory into equal temporal or spatial segments. To derive

equal segments based on the temporal progression of the road user, the total duration of the trajectory $T_i$ is divided by $n$, yielding a vector with the form:

$$F_i^* = [x_{i\,0} \quad y_{i\,0} \quad x_{i\,T_i/n} \quad y_{i\,T_i/n} \quad x_{i\,2T_i/n} \quad y_{i\,2T_i/n} \quad \cdots \quad x_{i\,T_i} \quad y_{i\,T_i}] \qquad \text{Eq. 4.3}$$

The number of features in this vectors is given by $d = 2 * (n + 1)$. The number of coordinates used in the feature vector is equal to $n + 1$. This is multiplied by two because each point $(x_i, y_i)$ is divided into two features, the $x$ and the $y$ component.

If $F_i^*$ is derived based on the temporal progression, features $f_{ij}^*$ can be included multiple times if the road user stops during the crossing. Although this definition of the feature vector can be useful in clustering based on speed and stopping manoeuvres, a more general clustering based on the pathway followed can be achieved using the spatial progression of the road user. This is done by first calculating the cumulative distance travelled $D_i$ by road user $i$.

$$D_i = \sum_{t=1}^{T_i} \sqrt{(x_{i\,t} - x_{i\,t-1})^2 + (y_{i\,t} - y_{i\,t-1})^2} \qquad \text{Eq. 4.4}$$

To derive equal segments based on the spatial progression of the road user, the cumulative distance $D_i$ is divided by $n$. Cut-off cumulative distances are calculated based on $D_i$ and $[0 \quad D_i/n \quad 2 * D_i/n \quad 3 * D_i/n \quad \cdots \quad D_i]$. Position points nearest to each of the resulting cut-off cumulative distances are selected. The coordinate $(x_{D_i/n}, y_{D_i/n})$, for example, is the coordinate in trajectory $S_i$ that is located at the nearest cumulative distance from the first point of the trimmed trajectory to the calculated cut-off cumulative distance $D_i/n$.The resulting trajectory vector has the form:

$$F_i^* = [x_{i\,0} \quad y_{i\,0} \quad x_{i\,D_i/n} \quad y_{i\,D_i/n} \quad x_{i\,2D_i/n} \quad y_{i\,2D_i/n} \quad \cdots \quad x_{i\,D_i} \quad y_{i\,D_i}] \qquad \text{Eq. 4.5}$$

Where the indices of the $x$ and $y$ coordinates indicate the respective cut-off cumulative distance. The resulting feature vectors describe the spatial progression of the bicyclist rather than the temporal progression. The $F_i^*$ vectors from all the bicyclists in the dataset are combined to form a pattern matrix $\mathcal{A}^*$, as shown in Eq. 4.11 with the shape $N\ x\ d$ where $d = 2 * (n + 1)$ and $N$ is the number of observations.

### 4.1.2 Proximity

Proximity is the means of measuring similarity between observations in the dataset and is used in the clustering algorithm to identify groups within the data. There are a number of methods for measuring proximity, the most common of which are shown in Eq. 4.6 through Eq. 4.8 [LAROSE & LAROSE, 2015]. An appropriate measure must be selected by the practitioner to reflect the type of features included in $F_i^*$ and in consideration of the character of the clustering problem.

Euclidean distance: $$d_{Euclidean}(x,y) = \sqrt{\sum_i (x_i - y_i)^2} \quad \text{Eq. 4.6}$$

Manhattan distance: $$d_{Manhattan}(x,y) = \sum_i |x_i - y_i| \quad \text{Eq. 4.7}$$

Different (categorical variables): $$different(x_i, y_i) = \begin{cases} 0 & if\ x_i = y_i \\ 1 & otherwise \end{cases} \quad \text{Eq. 4.8}$$

Euclidean distance (Eq. 4.6) is the most commonly used proximity measure and is applicable in cases in which the direct distance has an objective meaning. The Manhattan distance is useful in situations that are best represented using a grid (e.g. chessboard and city blocks) and in which the direct distance does not have a tangible meaning. The different measure reflects proximity for categorical variables for which physical distance cannot be measured. Another measure of proximity is the Hamming distance, which is the number of symbols that are different between two strings of equal length. This measure is useful for determining the similarity between words or sentences. There are a number of more complex methods for measuring the distance between trajectories and identifying similar trajectories. Examples of which include Dynamic Time Warping based measures, Edit Distance based measures and Longest Common Subsequence (LCSS) based measures. WANG ET AL. [2013] provide an overview and evaluation of six commonly used measures.

In this application, direct distance precisely describes the clustering problem, which involves clustering the trajectories based on their similarity in space. No categorical or string variables are included in the features vector $F_i^*$ for which the different measure must be incorporated.

Depending on the characteristics of the data included in the pattern matrix $\mathcal{A}^*$, it may be advantageous to normalise the dataset. This is typically necessary if there are considerable discrepancies between the magnitudes of the features in the feature vector $F_i^*$. For example, if one wishes to cluster crop output observations based on field size and annual precipitation, where $f_{i1}^*$ represents the field size in square kilometres (~$10^1$) and $f_{i2}^*$ contains the annual precipitation in millimetres (~$10^3$), the scales will have a large effect on the resulting proximity measure. In addition, the magnitude of the features relative to one another will influence the weighting of the individual features in the clustering process. To mitigate this effect, all features in $\mathcal{A}^*$ can be normalised. This is usually done using min-max normalisation (Eq. 4.9) or Z-score standardisation (Eq. 4.10).

$$x_{ij} = \frac{x_{ij}^* - \min(x_j^*)}{\max(x_j^*) - \min(x_j^*)} \quad \text{Eq. 4.9}$$

$$Z\ score = \frac{X - \mu}{\sigma} \quad \text{Eq. 4.10}$$

Where $\mu$ is the mean and $\sigma$ is the standard deviation of the attribute $j$. However, in this application, the features in $\mathcal{A}^*$ are all described using the same units (m) and therefore no single feature or subset of features will dominate clustering. On the contrary, normalising the data would result in a loss of information about the actual distance between the observations. Therefore, the following pattern matrix is used for observation clustering:

$$\mathcal{A}^* = \begin{bmatrix} x_{1\,0} & y_{1\,0} & x_{1\,D_1/n} & y_{1\,D_1/n} & x_{1\,2D_1/n} & y_{1\,2D_1/n} & \cdots & x_{1\,D_1} & y_{1\,D_1} \\ x_{2\,0} & y_{2\,0} & x_{2\,D_2/n} & y_{2\,D_2/n} & x_{2\,2D_2/n} & y_{2\,2D_2/n} & \cdots & x_{2\,D_2} & y_{2\,D_2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{N\,0} & y_{N\,0} & x_{N\,D_N/n} & y_{N\,D_N/n} & x_{N\,2D_N/n} & y_{N\,2D_N/n} & \cdots & x_{N\,D_N} & y_{N\,D_N} \end{bmatrix} \quad \text{Eq. 4.11}$$

### 4.1.3 Clustering

Clustering is a well-developed field of data analytics and as such a wide selection of clustering algorithms is available. The most well-known types of clustering algorithms, including their strengths and weaknesses, are introduced briefly below. A complete overview of clustering algorithms can be found in *Introduction to Data Mining* [PANG-NING ET AL., 2006] or *Data Mining and Predictive Analytics* [LAROSE & LAROSE, 2015].

- **K-means** is one of the most straightforward and commonly used clustering algorithms. The number of clusters is provided as input to the algorithm and that number of cluster centroids is randomly selected from the data sample. Each of the observations is assigned to the nearest cluster centroid based on the selected proximity measures. A new centroid for each cluster is selected which is nearest to the central point of the observations assigned to that cluster. All observations are then reassigned based on these new centroids. This is repeated until there is no change in the clusters or centroids. This algorithm is computationally inexpensive and can often provide excellent results if the number of clusters is known. However, poor results can emerge for clusters with differing sizes and densities and for non-globular clusters. Globular, or convex, is a term for clusters in which any line drawn in $d$ dimensional space between an observation and the cluster centre, or between two observations in one cluster, is within the boundary of that cluster.

- **Agglomerative Hierarchical Clustering** is a means of clustering in which the observations are first included as individual clusters with one observation in each cluster. At each subsequent step, the closest pair of clusters is combined based on the definition of proximity. Clusters are combined in this fashion until the predefined number of clusters is reached. Alternatively, all observation can be included in one cluster in a first step and then split at each subsequent step based on dissimilarity (Divisive Hierarchical Clustering). These algorithms are particularly useful when the application requires a hierarchy (e.g. taxonomy). However, hierarchical clustering algorithms are computationally expensive, a number of clusters must be predefined

and the finality of the merges can be a source of error if the data is noisy or contains many dimensions.

- **Density-based clustering** algorithms define clusters by identifying regions of high density that are separated by low-density regions (e.g. DBSCAN). In a first step, each of the observations is labelled as a core, border or noise observation based on the number of observations within a predefined radius defined using the proximity measure. The noise observations are deleted and the border observations are assigned to the most suitable cluster core. The advantages of this approach include high resistance to noise and the ability to cluster irregularly shaped data (non-globular). However, clusters with widely varying densities and observations with many dimensions are difficult to analyse using density based approaches.

In selecting a clustering algorithm, the nature of the observations to be clustered must be considered. Here, the expected pathway clusters vary with regard to the number of observations in each cluster (cluster size) and the degree of similarity between observations in the same cluster (cluster density). The resulting clusters are expected to be globular. Additionally, because the number of different pathways across the intersection is not known in advance, the number of clusters cannot be designated as an input to the algorithm.

Based on the attributes of the clustering problem, the Affinity Propagation [FREY & DUECK, 2007] clustering algorithm is selected. This algorithm is similar to K-means in that cluster centroids are identified and observations are assigned to the centroids based on their similarity. However, unlike the K-means algorithm, the number of clusters does not have to be specified in advance. All observations are considered simultaneously as cluster centroids, which are referred to as exemplars. An optimal set of exemplars is identified by exchanging two types of messages between data points:

> "The "responsibility" $r(i,k)$, sent from data point $i$ to candidate exemplar $k$, reflects the accumulated evidence for how well-suited point $k$ is to serve as the exemplar for point $i$, taking into account other potential exemplars for point $i$." [FREY & DUECK, 2007, P.972]

> "The "ability" $a(i,k)$, sent from candidate exemplar $k$ to point $i$, reflects the accumulated evidence for how appropriate it would be for point $i$ to choose point $k$ as its exemplar, taking into account the support from other points that point $k$ should be an exemplar." [FREY & DUECK, 2007, P.972]

Two parameters must be specified in the Affinity Propagation algorithm. The first of which, *preference*, which describes the suitability of each observation to be used as an exemplar. If the suitability of each observation is not known in advance, *preference* can be set to a common value such as the mean or median of all observation similarities. *Preference* carries the same unit as similarity, or negative proximity, which is quantified here using Euclidean

distance. The developers of the Affinity Propagation method recommend using the minimum, median or maximum similarity as an initial presumption for cluster creation [PROBABILISTIC AND STATISTICAL INFERENCE GROUP, 2009]. The magnitude of this value controls the number of clusters, where lower values lead to a larger number of clusters. The second parameter, *damping* $\lambda$, prevents numerical oscillations by controlling the portion of each message taken from the previous iteration and the prescribed updated value ($0 < \lambda < 1$) [FREY & DUECK, 2007].

The route across the intersection is identified for each of the observed bicyclists prior to clustering. A trail without first determining the route was initially carried out. However, this approach led to the clustering of bicyclists with similarly shaped trajectories but different start and end points. An example is a bicyclist who turned indirectly against the given direction of travel and a bicyclist who made a similar manoeuvre but continued riding straight, but on the opposite side of the road. If route is not first determined, the trajectories from these two bicyclists would likely be clustered together. To identify the route, the first position point $(x_i, y_i)_{t=0}$ and last position point $(x_i, y_i)_{t=T_i}$ of the original, untrimmed trajectory $S_i$ are used in combination with geometric information from the intersection. The method used to radially divide the intersection and assign the start and end approach is shown in Fig 4.2.

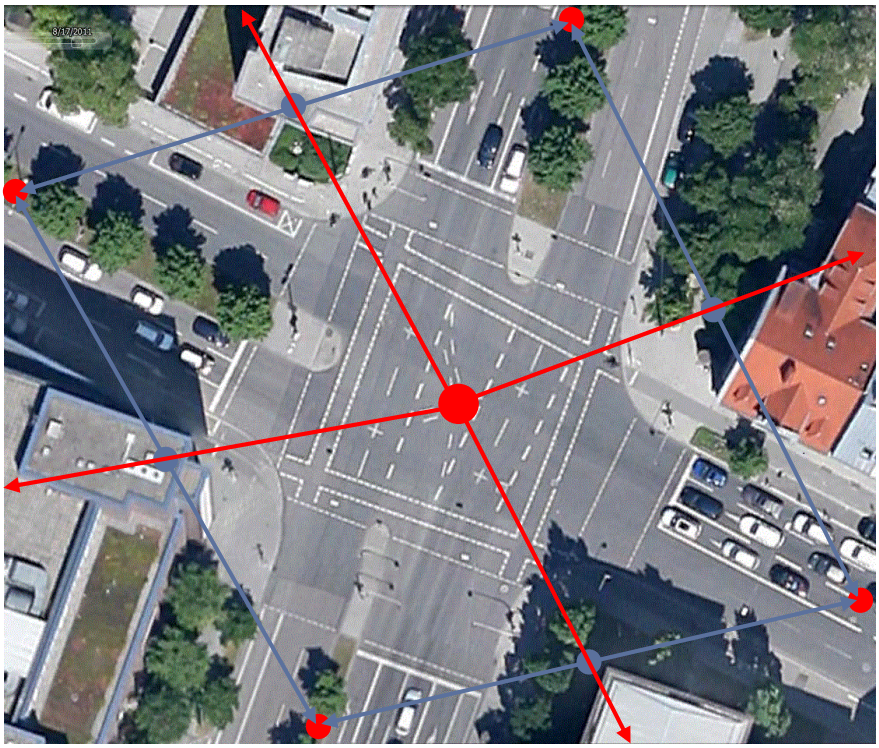


**Fig 4.2** Radial division of the intersection for route determination (background image: Google Earth 2013)

The coordinates of the centre point of the intersection (large red point in Fig 4.2) as well as a point on the centre line of each arm of the intersection (smaller red points in Fig 4.2) are identified from the intersection plans. The points representing adjacent arms are connected

and points are located at the centre of the connecting lines (blue points in Fig 4.2). Finally, these coordinates and the intersection centre are used to divide the intersection radially into polygons representing each arm.

The polygons within which the first and last position points are located dictate the approach and exit arm of the intersection. The final route label is a combination of the approach and exit arm. For example, if the first point of the observed trajectory is located within the polygon representing the North approach and the last point is within the East polygon, the trajectory is assigned the route NE. Bicycles that approach and exit on the same arm (e.g. NN) are excluded from the analysis.

The route labels are used to divide the observations into a total of $R$ $\mathcal{A}^*$ matrices, where $R$ is the number of routes identified at each intersection. The scikit learn implementation of Affinity Propagation [SCIKIT-LEARN DEVELOPERS, 2014] is used to cluster the trajectory data in each $\mathcal{A}^*$ matrix separately. The input parameter *preference* is systematically altered between the maximum and minimum similarity between the observations and the *damping* parameter is altered between $\lambda = 0.5$ and $\lambda = 1.0$ (limits defined by scikit learn). The resulting clustering structure of each *preference*/*damping* combination is evaluated using the mean Silhouette Score $\bar{s}$ (see Eq. 4.12). The combination resulting in the highest $\bar{s}$ is saved and used for the final clustering. The best parameter set for each of the pattern matrices is identified independently, meaning each route at each approach may have a different parameter set.

Once the clusters are developed, the similarities between the observations included in each cluster are examined and observations found to be very dissimilar to the cluster are removed. To this aim, the Euclidean distance between all of the observations in a cluster and the cluster exemplar are calculated and used to calculate the $Z - Score$ (Eq. 4.10) for each observation. This measure quantifies the number of standard deviations above or below the sample mean on which an observation lies in terms of distance from the cluster exemplar. Here, observations are removed from the cluster if $Z - Score > 2.0$ or $Z - Score < -2.0$ (where $\mu$ is the mean Euclidean distance from the cluster exemplar and $\sigma$ is the standard deviation of the Euclidean distances). This is done to remove outliers in the clusters. These are observations that are not similar to the other trajectories in the cluster and tend to be unique behaviours that were observed for only one bicyclist.

It was hypothesised that the number of position points extracted from the original trajectory $n$ influences the quality of the clustering results. However, the clustering approach delivers high quality results for $n > 2$. If $n = 2$, the algorithm cannot differentiate between routes that start and end at the same position but follow different pathways across the intersection, such as different types of left turns. If $n = 10$, important differences in the start and end point become less pronounced in comparison to the overall shape of the pathway, which was found to lead to over-grouping in some cases. Although the clustering results vary slightly for different values of $n$, this parameter is not found to play an important role in developing high

quality clusters. After assessing the Silhouette Score, over-segmentation and over-grouping for $n = [2,10]$ for all the intersections, $n = 5$ is selected for final clustering and is shown on the right side of Fig 4.3.

### 4.1.4 Abstraction

Abstraction is an optional step in which clustering results are compacted or simplified, typically with the goal of either subsequent machine based analyses or the human assessment of the clusters. Humans are limited in their capacity to visualise more than three dimensions. If multi-dimensional observations form the basis of the clustering problem, it becomes difficult to subjectively assess the quality of the clusters. Abstraction in the form of simplification makes it possible for humans to visualise and assess the clusters. If the goal is subsequent analysis, a compact representation of the clustered dataset is often useful. This can be done using representative observations, or if available, centroid observations from each cluster.

Here, despite the high dimensionality of the observations $d = 2(n + 1)$, the clusters can be easily assessed by humans by displaying the $F_i$ vectors as trajectories composed of $(x, y)$ coordinates. The original trajectory $S_i$ from the observations can be used to display all details from the trajectory (Fig 4.3, left). The dataset can be compacted for subsequent analyses by extracting the centroid, or exemplar, for each cluster. This is an advantage of centroid based clustering algorithms such as K-means and Affinity Propagation. The $F_i$ vectors shown as trajectories of the cluster exemplars for the same example are shown in Fig 4.3 (right).

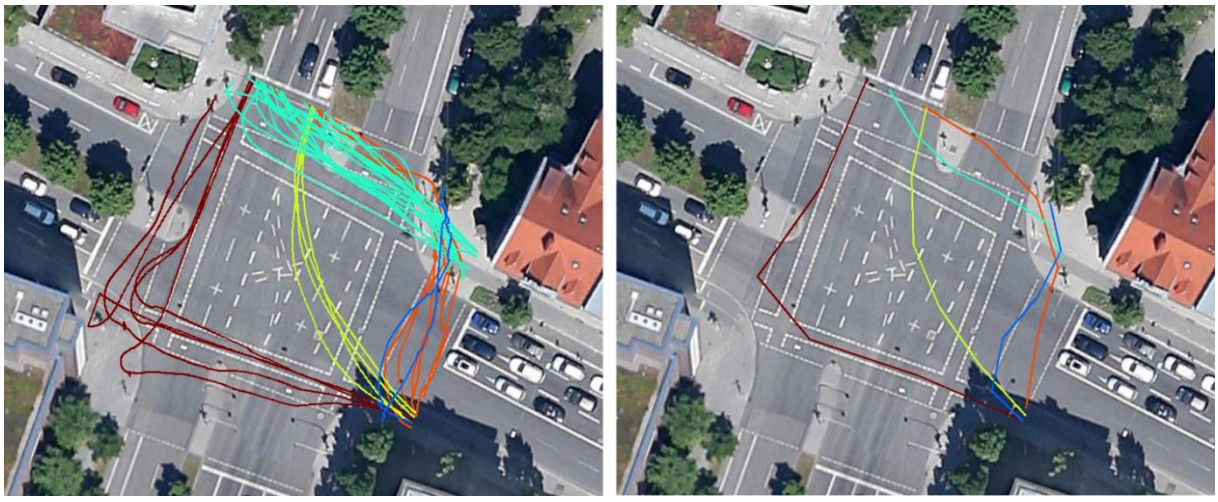


**Fig 4.3** Clustered trajectories from bicyclists with the route NE at the intersection Marsstr. / Seidlstr. (left) and the cluster exemplars (right) with $n = 5$ (background images: Google Earth 2013)

### 4.1.5 Evaluation

Trajectory clustering is an unsupervised learning task and as such, there is no set of classified trajectories with which to compare the resulting clusters. The quality of the clusters must nevertheless be assessed to ensure that they reflect valid clusters in the data [LAROSE & LAROSE, 2015]. One useful measure for evaluating clusters is the Silhouette Score $s_i$, which compares the proximity $a_i$ of observation $i$ to its assigned exemplar $k_i$ with the proximity to the next nearest exemplar $b_i$:

$$s_i = \frac{b_i - a_i}{\max(b_i, a_i)} \qquad \text{Eq. 4.12}$$

The resulting $s_i$ value ranges between $0$ and $1$, with higher values indicating better quality clusters. In addition to the disaggregate $s_i$ measure, the mean Silhouette Scores $\bar{s}$ is useful in assessing the overall quality of the clustering structure. According to ROUSSEEUW & KAUFMAN [1990], $\bar{s} > 0.70$ indicates a strong clustering structure, $0.50 < \bar{s} \leq 0.70$ indicates a reasonable structure, $0.25 < \bar{s} \leq 0.50$ points to a weak structure that may be a mathematical construct and $\bar{s} \leq 0.25$ shows that no substantial structure was found.

In order to assess the accuracy and feasibility of this trajectory clustering approach, the results are also evaluated subjectively. This is done by identifying instances of over-segmentation and over-grouping. Over-segmentation describes the clustering of observations in different clusters while they subjectively belong to the same cluster. Over-grouping, on the other hand, is the clustering of observations that are subjectively different from one another into a single cluster. The sum of the over-segmented and over-grouped observations yields the total error in the resulting clustering structure. Outliers are the trajectories that are deleted from the clusters based on their proximity to the cluster centroid ($Z - Score > 2.0$ or $Z - Score < -2.0$). The results for the four research intersections are shown in Tab. 4.1.

| | **Arcisstr. / Theresienstr.** | **Arnulfstr. / Seidlstr.** | **Karlstr. / Luisenstr.** | **Marsstr. / Seidlstr.** |
|---|---|---|---|---|
| # of observations | 502 | 488 | 395 | 1128 |
| # of routes | 9 | 10 | 9 | 12 |
| # of clusters (total) | 19 | 25 | 20 | 34 |
| Mean Silhouette Score $\bar{s}$ | 0.60 | 0.68 | 0.68 | 0.81 |
| # of outliers | 25 (5.0%) | 36 (7.4%) | 16 (4.1%) | 55 (4.9%) |
| # of over-segmentation errors | 3 (0.6%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) |
| # of over-grouping errors | 5 (1.0%) | 3 (0.4%) | 10 (2.5%) | 10 (0.9%) |
| # of total errors | 8 (1.6%) | 3 (0.4%) | 10 (2.5%) | 10 (0.9%) |

**Tab. 4.1** Evaluation of the approach specific clustering approach

The number of observations is the number of trajectories included in the clustering procedure after removing short trajectories and those that started or ended within the centre region of the intersection. The route across the intersection includes the approach and exit arm of the intersection (e.g. NE), with a maximum of 12 potential routes. This maximum was only observed at Intersection 4 Marsstraße and Seidlstraße. Only a portion of the routes was observed at the other intersections due to various factors, including differing flows and turning rates and unique patterns of movement on the one-way road (Theresienstraße).

The number of pathway clusters per route ranges between 2.1 and 2.8 for all the intersections. This makes logical sense as there are fewer pathway possibilities for the manoeuvres travelling straight and executing a right-hand turn than there are for carrying out a left-hand turn. There is no maximum number of theoretical pathway options as this will always depend on the geometry of the intersection.

The mean Silhouette Score $\bar{s}$ ranges from a low of 0.60 at Arcisstraße and Theresienstraße to a high of 0.81 at the intersection Marsstraße and Seidlstraße. The lowest score surpasses the cut-off for a reasonable structure while the remaining three scores are near or meet the requirement for a strong structure. In addition, the extremely low total error percentages, which range between 0.4% and 2.5%, suggest high quality clusters. The slight tendency towards over-grouping could be addressed by reducing the *preference* parameter.

The resulting cluster centroids are used in Section 5.2 to model the operational behaviour of bicyclists as they cross the intersection. In this approach, simulated bicyclists require a desired pathway across the intersection, which they attempt to follow while reacting to obstacles and other road users. The cluster centroids act as these desired pathways.

## 4.2 Generic pathway clustering

The pathway clustering approach described in the previous section is useful for determining the physical pathways followed by bicyclists arriving on specific approach of a specific intersection. Although the shapes of the trajectories in the pathways clusters are comparable for approaches with similar geometrical characteristics, the classification structure found in Section 4.1 does not offer any insight into the general pathfinding behaviour of bicyclists. In this section, a generic method for clustering the pathways used by bicyclists to carry out desired manoeuvres (left turn, right turn and riding straight across the intersection) is developed. This generic approach combines and clusters trajectories from bicyclists approaching from all directions at multiple intersections, allowing for the analysis of the pathfinding strategies of bicyclists in general. The clustering results are used in the development of tactical behaviour models (Section 5.1).

The methodology used to generically cluster the pathways of bicyclists is very similar to the method presented in Section 4.1. The main difference in comparison to the approach specific clustering method is that here the trajectories from bicyclists at all approaches of all the research intersections are collected in one large sample. Additional augmentations are necessary in the pattern representation step to enable the comparison between pathway forms with differing sizes (due to the geometry of the intersection) and orientations (due to the approach direction). Once again, the five steps in unsupervised learning outlined by JAIN & DUBES [1988] are used to guide the development of the generic clustering approach. In cases where the method does not diverge from that of the approach specific method, readers are referred to Section 4.1.

### 4.2.1 Pattern representation

As described in detail in Section 4.1.1, trajectories are denoted using an ordered set of position coordinates, the length of which depends on the duration of the intersection crossing. For clustering, trajectories are reduced to a one-dimensional feature vector $F_i^*$ containing data from selected position coordinates. If $F_i^*$ is regarded as a two dimensional spline in which each position observation is a spline coordinate, the size and orientation of the pathway become apparent. These two characteristics are important because they distinguish observed trajectories from different approaches and different intersections. The trajectories are transformed in order simulate a situation in which the origin of all trajectories is the same point at the same intersection. Three transformations are applied here to assimilate trajectories differing in size and orientation, enabling clustering based only on the shape or form of the trajectories:

1. **Rotation**: Least squares linear regression is carried out to find the line of best fit for all position coordinates from the start of the trajectory to the mid-point of the trajectory. The method also works if the line is fit using all points, but better results were achieved using only the first half. The angle of the slope $\alpha$ of the resulting line is used to rotate the entire vector $F_i^*$ by $\frac{\pi}{2} - \alpha$ clockwise. The first position coordinate of the trajectory serves as the origin for the rotation.

2. **Translation:** The rotated vectors are translated by $-x_{i\,0}$ in the x direction and $-y_{i\,0}$ in the y direction such that all features vectors begin at the point $(0{,}0)$.

3. **Scaling** – the rotated and translated vectors $F_i^*$ are scaled based on the length and shape of all the filtered and trimmed trajectories from the same approach of an intersection. The intention is to gain information about the shape of the given approach and intersection based on all trajectories observed at that given approach. The maximum y-coordinates of all rotated and translated trajectories from each approach $y_{\max\_app}$ are determined and the feature vectors are scaled on the x and y axis by a factor of ${y_{\max\_app}}^{-1}$. The resulting

trajectories have a maximum y coordinate of one and the original aspect ratio is maintained. This is repeated for all approaches of all four intersections.

### 4.2.2 Proximity

As in the approach specific clustering approach (Section 4.1.2), Euclidean distance is used as a proximity measure in the generic clustering approach. The pattern matrix is normalised through the transformation process described in the previous section. Therefore, $\mathcal{A}^*$ is not further processed using min-max normalisation (Eq. 4.9) or Z-score standardisation (Eq. 4.10).

### 4.2.3 Clustering

The Affinity Propagation (Frey & Dueck 2007) algorithm, which is introduced in Section 4.1.3, is used once again for generic clustering. The conditions for the approach specific pathway clustering problem also hold true for generic clustering and include imbalanced cluster sizes and varying cluster density (see Section 4.1.3). Once again, it is advantageous that the number of clusters is not required as an input parameter by the Affinity Propagation algorithm. The number of different pathways used by bicyclists is an outcome of the clustering algorithm and is therefore not know in advance. As in Section 4.1, clusters are expected to be globular.

In a first step, the manoeuvre of the bicyclist is identified using the method developed to extract route information in the approach specific pathway clustering approach. Each trajectory is assigned a starting and ending arm of the intersection (e.g. NE for a bicyclist approaching from the North and exiting on the East arm). This route information is generalised here to denote the manoeuvre associated with the route. Three manoeuvres are defined; travelling straight across the intersection (NS, EW, SN, WE), turning right (NW, EN, SE, WS) and turning left (NE, ES, SW, WN). Bicyclists that approach and exit on the same arm of an intersection are excluded from the dataset. Three pattern matrices $\mathcal{A}^*$ are built from the rotated, translated and scaled feature vectors $F_i^*$, one for each of the manoeuvres. Trajectories from bicyclists observed on all approaches of the four research intersections are included in each of the three pattern matrices. Effectively, the 40 pattern matrices generated for the approach specific clustering method are combined here into three generalised manoeuvre matrices.

Once the pattern matrices are created, the *preference* and *damping* parameters of the Affinity Propagation algorithm are selected using the same systematic variation of parameters that is described in Section 4.2.3. The mean Silhouette score $\bar{s}$ (Eq. 4.12) is used as an evaluation criterion and is maximized by adjusting the *preference* and *damping* $\lambda$ parameters. The parameter set leading to the highest $\bar{s}$ value is used to create the clusters for each manoeuvre. The best parameters set is found for each of the pattern matrices independently, meaning that different parameters are used for the clustering of each of the manoeuvres.

Once the cluster structures are developed for the three manoeuvres, observations are removed from the clustered dataset based on the $Z - Score$. As in Section 4.1.3, observations are removed from the cluster if $Z - Score > 2.0$ or $Z - Score < -2.0$ (where $\mu$ is the mean Euclidean distance from the cluster exemplar and $\sigma$ is the standard deviation of the Euclidean distances).

As in the approach specific clustering method, the number of position points $n$ included in reduced feature vectors is found to play a marginal role in the quality of the generic clusters. In order to gain sufficient information about the cluster shape to differentiate between different forms of left turn $n = 3$ is recommended as a lower boundary. A range of position points, $10 < n < 25$, is found to produce the best pathway clusters.

### 4.2.4 Abstraction

Abstraction enables human assessment of the resulting cluster structure or the simplification of the dataset for subsequent use. Here, the reduced feature vectors are plotted as splines overlaid over the plans of the research intersections. Splines are used here to display a smoothed trajectory. This can also be done as in Section 4.1.4, where the pathways are displayed as line segments with the coordinates from the feature vector $F_i^*$. The rotated, scaled and translated feature vectors from bicyclists observed at all the research intersections were combined in mutual pattern matrices (one per manoeuvre) for clustering. The clustered observations are then returned to their original form and are plotted on the plan of their respective intersection. This allows for the assessment of the clustering results under the original conditions at the observation intersection. Agreeance between intersections can be evaluated by comparing the plotted trajectories belonging to the same cluster on the different intersection plans. Examples of the resulting four left turn clusters at the research intersections are shown in Fig 4.4.

These examples are used in the qualitative evaluation of the approach along with the quantitative evaluation using the Silhouette Score. Additionally, the clustered transformed trajectories from all the bicyclists carrying out a given manoeuvre (from all intersections) were examined. This made it possible to identify potential errors or problems in the clustering approach that are difficult to recognise once the trajectories are returned to their original orientation and shape.

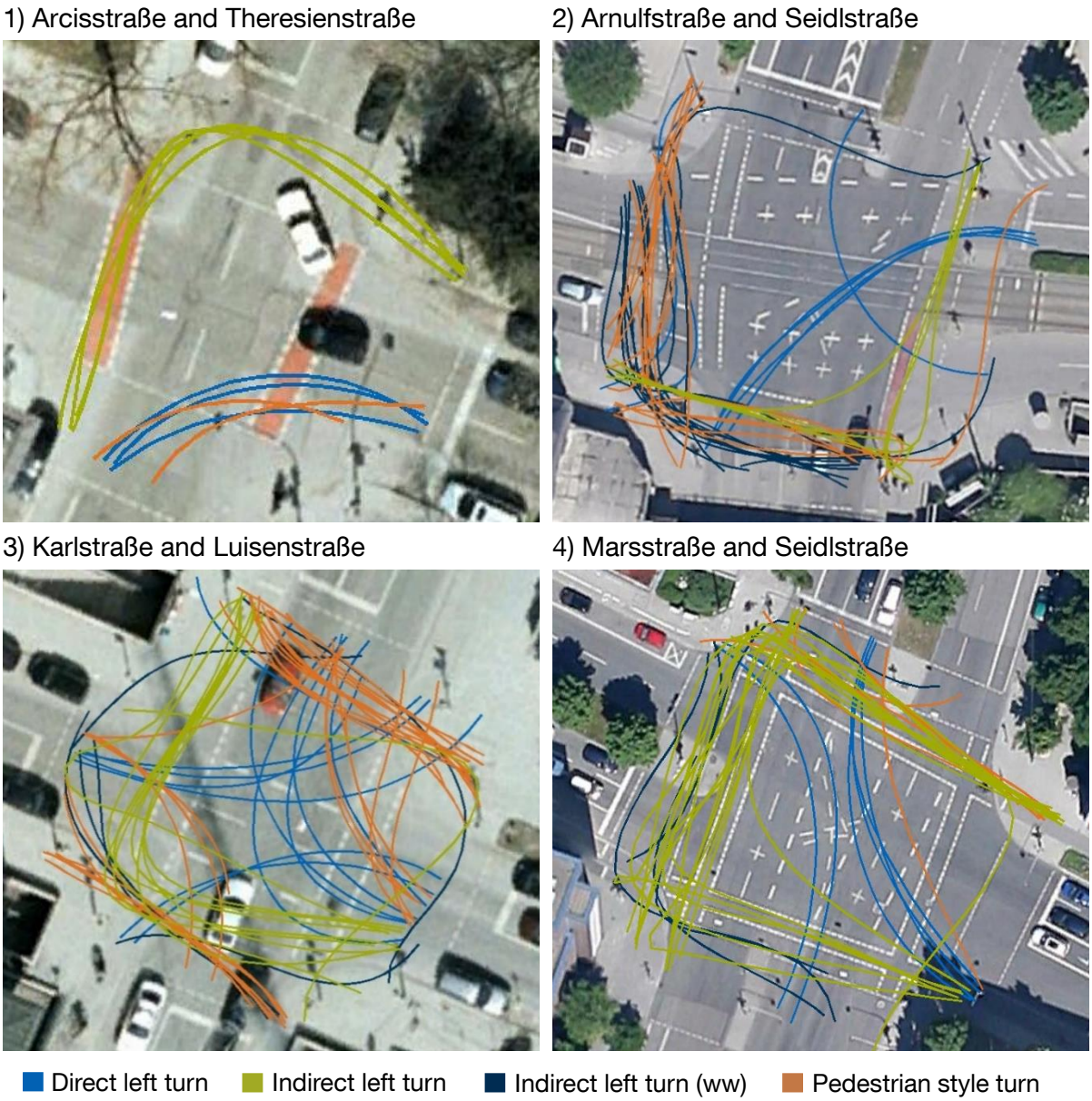



**Fig 4.4** Resulting left turn clusters (background images: Google Earth 2013)

### 4.2.5 Evaluation

The generic clustering approach is evaluated using the mean silhouette score $\bar{s}$, the number of outliers removed from the trajectory set for each manoeuvre and the total number of over-segmentation and over-grouping errors, which are determined qualitatively. These are the same measures used in Section 4.1.5 to evaluate the approach specific pathway clustering method. The evaluation results are shown for each of the three manoeuvres in Tab. 4.2.

| | Straight | Left turn | Right turn |
|---|---|---|---|
| # of observations | 1922 | 149 | 152 |
| # of clusters | 2 | 4 | 2 |
| Mean Silhouette Score $\bar{s}$ | 0.88 | 0.47 | 0.77 |
| # of outliers | 118 (6.1%) | 12 (8.1%) | 16 (10.5%) |
| # incorrect group | 0 (0.0%) | 14 (9.4%) | 4 (2.6%) |
| # of over-segmentation errors | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) |
| # of over-grouping errors | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) |
| # of total errors | 0 (0.0%) | 14 (9.4%) | 4 (2.6%) |

**Tab. 4.2** Evaluation of the generic clustering approach

The cluster structure resulting from the generic approach proves to be effective. The mean Silhouette Score $\bar{s}$ for the right turning and straight manoeuvres exceed the threshold for a strong cluster structure of $\bar{s} = 0.7$ [ROUSSEEUW & KAUFMAN, 1990]. The score for the left turn manoeuvre is considerably lower and lies slightly below the threshold for a reasonable structure. Nevertheless, the qualitative assessment of the cluster structure for all three manoeuvres indicates excellent results that agree almost completely with the presumed manoeuvre types. For example, the three types of left turns identified in the review of the literature, direct, indirect and indirect against the given direction of travel (see Fig 3.2), are identified by the clustering approach. An additional type of left hand turn in which the road user ends the manoeuvre on the wrong side of the roadway and travels against the given direction of travel is identified. Trajectories of this type are shown in orange in Fig 4.4 and are referred to as a pedestrian type turn.

The number (and corresponding percentage) of observations in each of the clusters for the three manoeuvres (right turn, left turn and straight) are shown in Fig 4.5. The Silhouette Score $s_i$ for each of the observations in the clusters can be read along the x axis. The highest Silhouette Score is plotted on the top of the cluster globule and the lowest Silhouette Score at the bottom. A globule indicating perfect clustering would be a rectangle with a width of 1.0. As can be seen in Fig 4.5, the clusters for the straight and right turn manoeuvres are quite well defined, while those for left turning bicyclists tend to have lower Silhouette Scores. The small group of bicyclists that made an indirect straight manoeuvre arrived at the intersection travelling in the mandatory direction of travel and then switched sides of the road to ride against the mandatory direction of travel. It is hypothesised that these bicyclists were reaching their destination on the other side of the road.

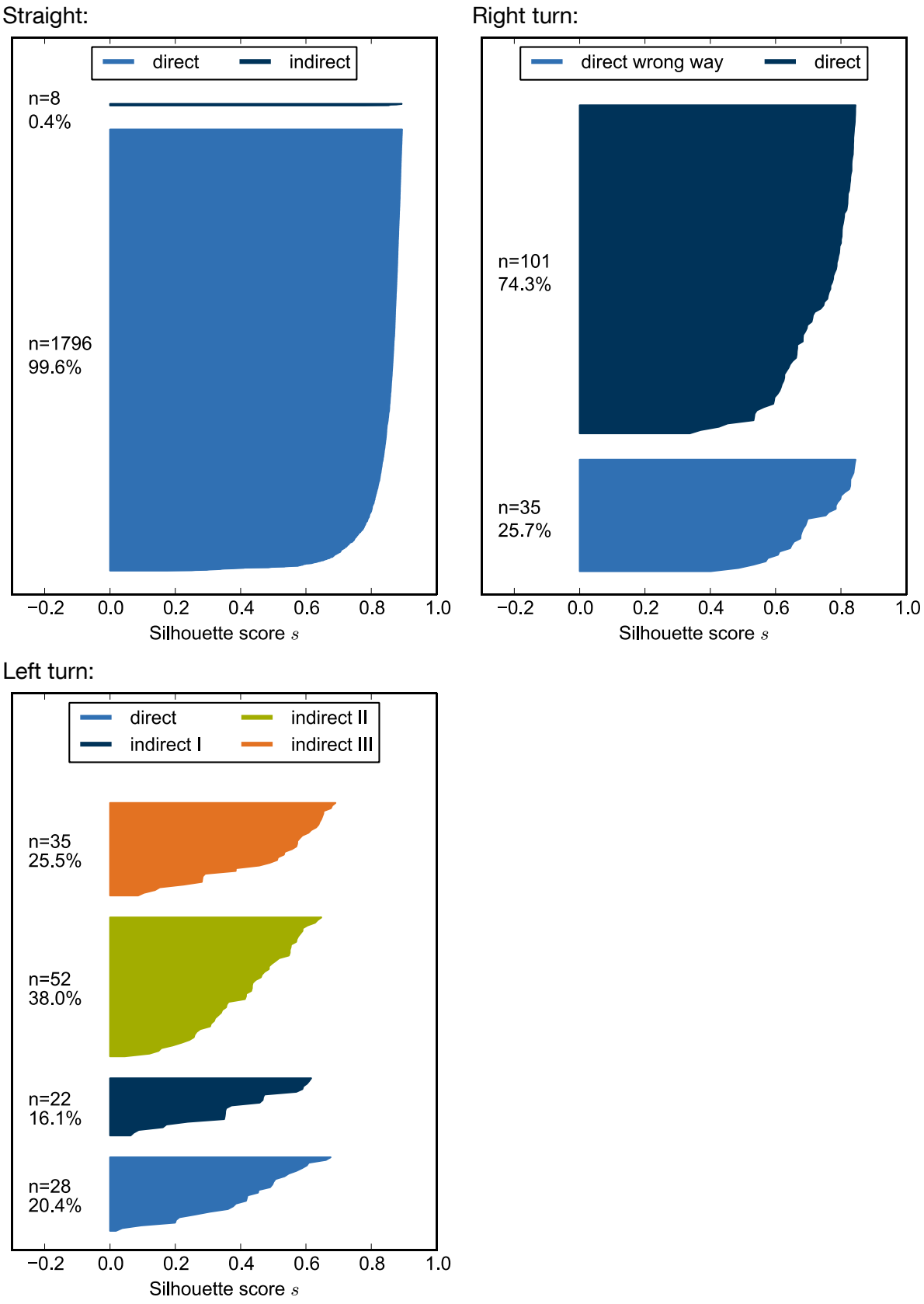



**Fig 4.5** Silhouette score distribution and number of observations in each cluster

# 5. Modelling bicyclist behaviour

In order to microscopically simulate traffic, it is necessary to implement models representing behaviour at the three levels of the hierarchy suggested by MICHON [1985]; strategic behaviour, tactical behaviour and operational behaviour. Within the context of this dissertation, models are developed to reflect the unique operational and tactical behaviour of bicyclists. Strategic behaviour, such as mode and route choice, is outside of the scope of this work. In this section, the modelling approaches for the tactical (Section 5.1) and operational (Section 5.2) behaviour are presented. Both sections begin with a detailed explanation of model specification followed by the methodology used to calibrate and validate the models using trajectory and situational data. The results of the model calibration and validation are presented at the end of each section.

## 5.1 Tactical models

In Section 2.3, Research needs assessment, the need to develop an approach for simulating bicyclists using and switching between different parts of infrastructure (bicycle facility, roadway and sidewalk), riding with and against the mandatory direction of travel and selecting unique pathways across the intersection was identified. To this aim, tactical models are estimated in this section to reflect four unique tactical behaviours of bicyclists; infrastructure selection, reaction to a red traffic signal, direction of travel and selection of a pathway turning left. The left turn manoeuvre was selected based on the results of the generic clustering carried out in Section 4.2. While bicyclists travelling straight across the intersection were found to use mainly one type of manoeuvre, bicyclists turning left were found to implement four types of turns with relatively similar frequencies. Bicyclists turning right were found to carry out a normal right turn 75% of the time and a right hand turn to travel in the wrong direction once completing their manoeuvre 25% of the time. These two types are not distinguished in this section as the manoeuvres are very similar and the direction of travel is analysed as a separate tactical choice.

In order to predict the outcome of these tactical choices, binomial and multinomial logistic regression models are estimated using revealed preference data from the 4710 bicyclists observed at the four research intersections. This number is larger than the number of trajectories clustered in the previous section because bicyclists not tracked successfully throughout the video frame using *Traffic Intelligence* are also included here. Where possible, data were extracted using automated methods and the results of the generic manoeuvre clustering approach presented in 4.2. However, in cases where the trajectory data were incomplete, data describing the behaviour of the bicyclists and the situation were manually collected.

Logistic regression models are used to describe the relationship between a response variable and a set of explanatory variables. In this case, the explanatory variables describe the average and momentary traffic situation, the geometry of the intersection and the phase timing of the signal control. In addition, tactical decisions made previously by the bicyclist are taken into consideration as explanatory variables. This includes, for example, if a bicyclist arrives at the intersection using a bicycle facility, the sidewalk or the roadway. These variables are only included if they are always observable before the modelled tactical choice is made. All of the variables listed in Tab. 3.5 and Tab. 3.6 are used here as potential explanatory variables.

### 5.1.1 Model specification

Logistic regression models are used to predict categorical variables based on a set of observable independent variables, which can be continuous or categorical. The variable to be predicted can be binary (binary logistic regression) or have more than two possible outcomes (multinomial logistic regression). Logistic regression follows the principles of linear regression, such that the expected value of a given outcome $Y$ is linearly related to a vector of explanatory variables $x$. Linear regression is defined below:

$$E(Y|x) = \beta_0 + \beta_1 x \qquad \text{Eq. 5.1}$$

where $x$ is a vector of explanatory variables, $\beta_1$ is a vector of parameters estimated from observed data that represent the relative influence of each of the explanatory variables and $\beta_0$ is the constant coefficient associated with the expected value of $Y$.

According to HOSMER ET AL [2013], $\pi(x) = P(Y = i|x)$ is used to represent the conditional mean of $Y$ given $x$ for a logistic regression model. This is expressed below:

$$\pi(x) = P(Y = i|x) = \frac{e^{\beta_0+\beta_1 x}}{1 + e^{\beta_0+\beta_1 x}} \qquad \text{Eq. 5.2}$$

The equation above represents a dichotomous choice situation where Y can take on a value of 0 or 1. In a multinomial choice situation in which Y can take on $n > 2$ values, the equation can be generalized to:

$$\pi(x) = P(Y = i|x) = \frac{e^{\beta_{i0}+\beta_{i1} x_i}}{\sum_{i=0}^{n} e^{\beta_{i0}+\beta_{i1} x_i}} \qquad \text{Eq. 5.3}$$

The logit transformation is used to simplify Eq. 5.3 and capitalise on the useful properties of linear regression. This transformation is given by:

$$g(x) = ln\left[\frac{\pi(x)}{1 - \pi(x)}\right] = \beta_0 + \beta_1 x \qquad \text{Eq. 5.4}$$

In linear regression, the outcome variable is often given as $y = E(Y|x) + \varepsilon$, where $\varepsilon$ is an error term that is typically assumed to follow the Gaussian distribution with a mean of zero and a variance that is constant across variables. The outcome of a dichotomous variable is expressed by $y = \pi(x) + \varepsilon$, where the error term $\varepsilon$ follows a distribution with a mean of zero and a variance $VAR(\epsilon) = \pi(x)[1 - \pi(x)]$. Log likelihood maximisation is then applied to identify the set of $\beta$ parameters that produce the best fit for a set of observed explanatory variables and choice outcomes[8].

### 5.1.2 Model calibration

A logistic regression model for each of the identified tactical choices is specified and calibrated using recursive feature elimination [GUYON & ELISSEEFF, 2003], which combines K-fold cross validation and predictor selection based on the log likelihood of the model. Binomial and multinomial regression models are estimated and evaluated using the statistics software package *R* [THE R FOUNDATION, 2016]. The main effects and two-way interactions between the situational variables listed in Tab. 3.5 and Tab. 3.6 are used as an initial set of explanatory variables $x$. The following steps are taken to identify the optimal set of explanatory variables for each of the tactical choice models and estimate the corresponding $\beta$ parameters:

1. **Data pre-processing:** Relevant cases are extracted from the complete dataset to analyse each of the tactical choices. For example, to estimate a regression model for predicting the response to a red signal, only cases where the bicyclist encounters a red signal are selected (N=1935). The data subset for each of the tactical model is, therefore, unique and must be pre-processed individually prior to model estimation. Variables are removed if they are constant or have very low variance in the subset. The pair-wise correlations between the remaining variables are assessed to identify inter-correlated variables. If a correlation greater than 0.6 is identified, the variable with the largest mean correlation with all other variables is removed from the dataset. Data pre-processing is carried out in two phases. In the first phase, the individual variables listed in Tab. 3.5 and Tab. 3.6 are assessed and variables with near to zero variance and high correlations with other variables are removed from the dataset. Pair-wise interaction terms for the remaining variables are created and the pre-processing procedure is repeated.

2. **Feature selection:** There are a number of methods available for reducing the number of variables, including backwards and forwards elimination as well as grouping the variables into components (principle component analysis or PCA). In order to maintain the

[8] A detailed description of logistic regression is outside the scope of this dissertation. However, if the reader is interested, the book *Applied Logistic Regression* [HOSMER ET AL., 2013] provides an excellent overview of the topic.

interpretability of the model results and to avoid potential difficulties with including categorical data in PCA, it was decided to use the individual variables as well as two-way interactions of the variables themselves without first carrying out a principle component analysis. The main effects of the 37 variables in Tab. 3.5 and Tab. 3.6 and the two-way interactions between all variables generate an extremely large number of explanatory variables, even after a number of variables are removed in step 1.

Here, recursive feature elimination is used to select the optimal set of explanatory variables (also know as predictors). This is done by dividing the resulting dataset into $k = 10$ roughly equally sized subsets for a recursive feature elimination that incorporates resampling (see KUHN [2008]). The model is estimated using $k - 1$ of the subsets and is evaluated using the remaining subset. This is repeated $k$ times using each of the data subsets once for evaluation. The backwards elimination process for feature selection is carried out within each of these folds.

The model is estimated using all of the variables remaining after the pre-processing step. The predictive power of the model, which is assessed using the Area Under the Curve (AUC) for binomial logistic regression and accuracy (Eq. 5.5) for multinomial regression, is assessed using the held back dataset and the explanatory variables are ranked based on their importance. The least important variable is removed and the model is re-estimated with the remaining variables. This is repeated until only one variable remains in the model. The optimal set of predictors (largest AUC or accuracy) is identified for each fold. The performance profiles of the variable subsets are calculated over all the samples, held back in turn in each of the $k$ folds of the cross-validation, and the optimal set of predictors is determined.

The *recursive feature elimination* function of the classification and regression training package *caret* [KUHN, 2016] is used to identify the most powerful set of predictors.

3. **Full model estimation**: The entire data subset is used to estimate the $\beta$ values for the identified optimal set of predictors. In order to improve the interpretability of the regression models, the main effects of both variables in two-way interaction terms are added to the optimal set of predictors for the final model. This is done even if the main effects do not improve the predictive power of the model.

4. **Simplified model estimation:** The $\beta$ values are re-estimated for a reduced model comprised of only the predictors found to be statistically significant ($p < 0.01$) in the full model. In cases where the main effect can replace an interaction term, the main effect predictor is given preference to maintain model simplicity.

In this section, the calibrated simplified models are presented. The detail offered by these models is deemed sufficient for application in microscopic traffic simulation. However, if the reader is interested in the full model results, they are included in Appendix 1.

### 5.1.3 Model validation

The Receiver Operating Characteristic (ROC) curve, which compares the true positive rate (Sensitivity, Eq. 5.6) with the false positive rate (1- Specificity, Eq. 5.7) of a binary predictor at various classification thresholds, is used to assess the binomial logistic regression models and identify the optimal classification threshold. According to HOSMER ET AL. [2013, P.174], "this measure has now become the standard for evaluating a fitted model's ability to assign, in general, higher probabilities of the outcome to the subgroup who develop the outcome (y=1) than it does to the subgroup who do not develop the outcome (y=0)". The predictive power of the model can be deduced from the Area Under the Curve (AUC), which is the area that falls under the ROC curve. AUC values range between 0.5 and 1.0, where 0.5 indicates that the model is no better at predicting the outcome than random chance and 1.0 indicates a perfect prediction. In general, AUC values between 0.5-0.7 indicate poor discrimination, 0.7-0.8 indicates acceptable discrimination, 0.8-0.9 signifies excellent discrimination and above 0.9 shows outstanding discrimination [HOSMER ET AL., 2013].

In addition to evaluating the power of the logistic regression model, the ROC Curve is useful for selecting a well-suited cut-off point for the classification. Typically the cut-off point for a classification model is set at 0.5 such that if $P(y = 1) \geq 0.5$, the outcome is predicted to be one. This value can be shifted, however, to maximize the sensitivity (Eq. 5.6) and specificity (Eq. 5.7) of the regression model. Here, a cut-off point is selected for each of the models that is plotted on the upper most left corner of the ROC Curve.

Along with AUC, the confusion matrix, an example of which is shown in Tab. 5.1, is used to evaluate the logistic regression models.

| | | True class | |
|---|---|---|---|
| | | **p** | **n** |
| Predicted class | **Y** | **T**rue **P**ositives | **F**alse **P**ositives |
| | **N** | **F**alse **N**egatives | **T**rue **N**egatives |

**Tab. 5.1** Confusion matrix concept (adapted from FAWCETT [2006])

The following metrics are derived from the confusion matrix and are used here to evaluate the predictive power of the full and simplified models (abbreviations are taken from the bolded letters in Tab. 5.1):

| | | |
|---|---|---|
| Accuracy: | $\frac{TP + TN}{p + n}$ | Eq. 5.5 |
| Sensitivity (True positive rate): | $\frac{TP}{p}$ | Eq. 5.6 |
| Specificity (True negative rate): | $\frac{TN}{n}$ | Eq. 5.7 |
| Positive predictive value (Positive precision): | $\frac{TP}{Y}$ | Eq. 5.8 |
| Negative predictive value (Negative precision): | $\frac{TN}{N}$ | Eq. 5.9 |

To evaluate multinomial logistic regression models, these evaluation parameters are generalised to the mean sensitivity, mean specificity, mean positive predictive value and mean negative predictive value across all choice categories. The confusion matrix for each of the multinomial logistic models is presented in addition.

### 5.1.4 Results

The simplified regression models estimated for the four tactical choices are presented in this section. For each of the tactical choices, the optimal set of predictors identified using recursive feature elimination and K-fold cross validation is given. The $p$ value is used to test the null hypothesis that the $\beta$ factors in the predicted model are equal to zero. The smaller the $p$ value is, the less likely that the null hypothesis is true, meaning that the $\beta$ factor in question is more likely to be different from zero. In this section, only the predictors found to be statistically significant ($p \leq 0.01$) are shown (simplified model). The full models with the complete set of predictors are given in Appendix 1. The predictors are sorted by their predictive power within the main effects and interaction effects. The most important predictors in each model are discussed and compared with the findings of previous studies. In total, data describing the tactical behaviour of 4710 bicyclists were collected. The tactical choices selected for analysis, all of which are described using nominal variables with two or three categories, are listed in Tab. 5.2 along with the number and percentage of bicyclists observed carrying out each of the choice outcomes.

| Tactical choice | Category 1 | Category 2 | Category 3 |
|---|---|---|---|
| Infrastructure selection (no bicycle facility) N=451 | Roadway N=428 (94.9%) | Sidewalk N=23 (5.1%) | n.a. |
| Infrastructure selection (bicycle facility) N=3727 | Bicycle facility N=3532 (94.8%) | Roadway N=67 (1.8%) | Sidewalk N=128 (3.4%) |
| Response to red signal N=1935 | Stop N=1552 (80.2%) | Violate N=383 (19.8%) | n.a. |
| Direction of travel N=4710 | With direction N=4651 (98.7%) | Against direction N=59 (1.3%) | n.a. |
| Left turn manoeuvre N=426 | Direct turn N=66 (15.5%) | Indirect turn N=166 (39.0%) | Indirect turn (wrong way) N=194 (45.5%) |

**Tab. 5.2** Tactical choices with categories and observed counts and percentages

**Infrastructure selection without a bicycle facility**

Over 95% of the observed bicyclists use the roadway on approaches with no bicycle facility. A small optimal set of predictors was found for this tactical choice (Appendix 1). A simplified model consisting of only two predictors is found to provide acceptable predictive power (AUC = 0.76). Traffic attributes on the approach have an important influence on infrastructure choice; the ratio of probability (odds ratio) of using the roadway decreases by 1.72 ($0.58^{-1}$) times for each additional car in the approach. This finding echoes that of KULLER ET AL. [1986], who found that high traffic volumes discourage roadway use. The manoeuvre of the bicyclist also affects the choice outcome. The manoeuvre (left turn) is included separately in the simplified model because the large effect of the interaction term stems mainly from the manoeuvre predictor. According to this model, the ratio of probability for using the roadway decreases by 6.25 ($0.16^{-1}$) for bicyclists turning left. This finding seems counterintuitive but is because many bicyclists turning left ride against the mandatory direction of travel (route simplification) and therefore use the sidewalk rather than the roadway.

Although roadway use is predicted with considerable success, the prediction of sidewalk use proves to be less reliable. This could indicate that bicyclists choose to use the sidewalk for reasons that are unobservable, such as a feeling of safety or the anticipation of upcoming manoeuvre. Additionally, the low number of sidewalk use observations limits the potential to identify patterns between the independent variables and this choice outcome. A high classification threshold of 0.95 is identified, which addresses the observed skewness in

choices by shifting predictions into the sidewalk category. In this case, the full model offers a substantially better prediction of infrastructure use without adding excessive complexity.

| **N = 451**<br>**Sidewalk use = 0, Roadway use = 1** | $\beta$ | **Odds ratio** | $p$ |
|---|---|---|---|
| Intercept | 4.24 | 69.20 | 0.000 |
| Cars in approach | -0.54 | 0.58 | 0.000 |
| Manoeuvre (left turn) | -1.81 | 0.16 | 0.000 |
| | | | |
| Classification threshold: | | 0.95 | |
| AUC | | 0.76 | |
| Accuracy | | 0.73 | |
| Sensitivity | | 0.73 | |
| Specificity | | 0.78 | |
| Positive predictive value | | 0.98 | |
| Negative predictive value | | 0.13 | |

**Tab. 5.3** Simplified binomial logistic regression model with evaluation for infrastructure selection without bicycle facility

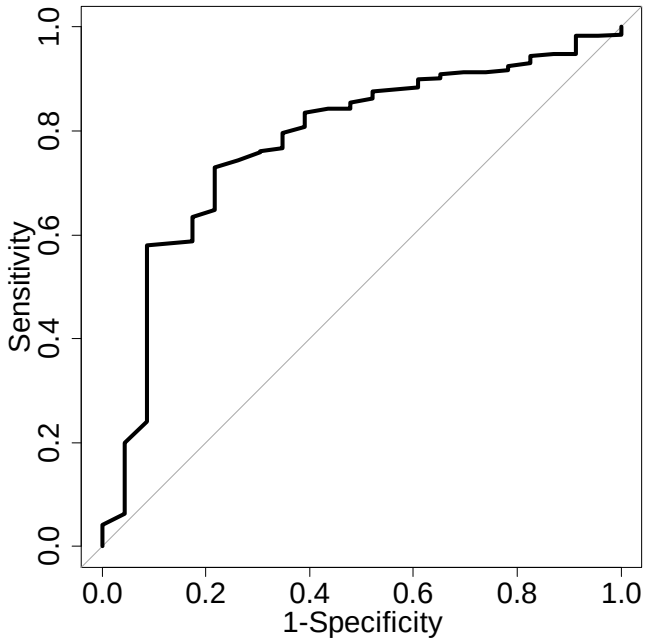



**Fig 5.1** ROC curve for the simplified binomial logistic regression model for infrastructure selection without bicycle facility

**Infrastructure selection with a bicycle facility**

If a bicycle facility is provided, bicyclists tend to use this facility. Over 95% of bicyclists observed on approaches with a bicycle facility selected this infrastructure, which is slightly higher than the 90% found by ALRUTZ ET AL. [2009]. Infrastructure selection can be framed as a discrete choice with three possible outcomes, bicycle facility, roadway or sidewalk. Initially, a multinomial logistic regression model was estimated to predict infrastructure use. Although correlations were found between the predictors and the choice outcome, these correlations were not strong enough to estimate a model capable of predicting roadway or sidewalk use.

In order to capitalise on the simplicity of binomial logistic regression as well as the adjustable classification threshold, the model shown in Tab. 5.4 is developed to predict whether a bicyclist will use an available bicycle facility or not. The strongest predictor of bicycle facility use is a right turn manoeuvre, which decreases the ratio of probability of using the bicycle facility by 6.03 ($0.17^{-1}$) times due to increased sidewalk use. The width of the bicycle facility plays an important role in the choice, with the ratio of probability of bicycle facility use increasing by 1.25 times for each additional cm of width ($e^{\frac{21.56}{100}}$). This effect is moderated by the volume of bicycle traffic on the approach, which decreases the probability ratio of bicycle facility use by 1.05 times for each additional bicycle per hour. The presence of other road users in the approach has an interesting effect on bicycle facility use. If there are only cars or only pedestrians present, the ratio of probability of bicycle facility use is reduced. If both are present, however, the interaction term increases the probability of bicycle facility use. This make intuitive sense as the presence of other road users on the sidewalk and roadway likely push bicyclists into an available bicycle facility. The presence of other bicyclists on the other hand, pushes bicyclists from the bicycle facility, particularly on separated facilities.

The simplified model predicts bicycle facility use with acceptable overall accuracy. However, the prediction of bicycle facility use is more reliable than that of not using the bicycle facility. This is likely due to the overrepresentation of bicycle facility observations in the sample and the potential role of personal attributes and unobservable factors in the choice to use the roadway or sidewalk when a bicycle facility is available. The high classification threshold of 0.96 coerces the prediction of not using the bicycle facility, but these predictions are often incorrect (low negative predictive value).

The findings of previous studies indicate that the width and type of the bicycle facility are decisive in infrastructure selection while the traffic conditions do not play an important role [ALRUTZ ET AL., 2009; GUO ET AL., 2013]. According to the findings here, the number and type of road users in the approach have a strong influence on infrastructure choice. Previous studies found that wider bicycle facilities have a higher rate of acceptance, which is confirmed. However, unlike ALRUTZ ET AL. [2009], on-road bicycle lanes are found here to have a higher acceptance than physically separated facilities.

| **N = 3727**<br>**No bicycle facility use = 0, Bicycle facility use = 1** | $\beta$ | **Odds ratio** | $p$ |
|---|---|---|---|
| Intercept | -30.86 | 0.00 | 0.000 |
| Manoeuvre (right turn) | -1.80 | 0.17 | 0.000 |
| Bicyclist volume – approach (bicycle/h) | 0.09 | 1.09 | 0.000 |
| Bicycle facility width (m) | 21.56 | 2.32e9 | 0.000 |
| Bicycle facility type (separated) | -5.65 | 0.00 | 0.001 |
| Driving lanes (same direction) | 1.24 | 3.47 | 0.000 |
| Sidewalk width (m) | -1.50 | 0.22 | 0.002 |
| Centre island | 1.81 | 6.10 | 0.010 |
| Parking | -1.67 | 0.19 | 0.000 |
| Pedestrians in approach | -0.16 | 0.85 | 0.033 |
| Bicyclists in approach | -0.04 | 0.96 | 0.421 |
| Cars in approach | -0.01 | 0.99 | 0.828 |
| Bicycle facility width (m) * Bicyclist volume – approach (bicycle/h) | -0.05 | 0.95 | 0.000 |
| Bicycle facility type (separated) * Bicyclists in approach | -0.28 | 0.76 | 0.000 |
| Bicycle facility type (separated) * Sidewalk width (m) | 0.75 | 2.11 | 0.056 |
| Cars in approach * Pedestrians in approach | 0.06 | 1.06 | 0.047 |
| | | | |
| Classification threshold: | | 0.96 | |
| AUC | | 0.76 | |
| Accuracy | | 0.73 | |
| Sensitivity | | 0.73 | |
| Specificity | | 0.72 | |
| Positive predictive value | | 0.98 | |
| Negative predictive value | | 0.13 | |

**Tab. 5.4** Simplified binomial logistic regression model with evaluation for infrastructure selection with bicycle facility

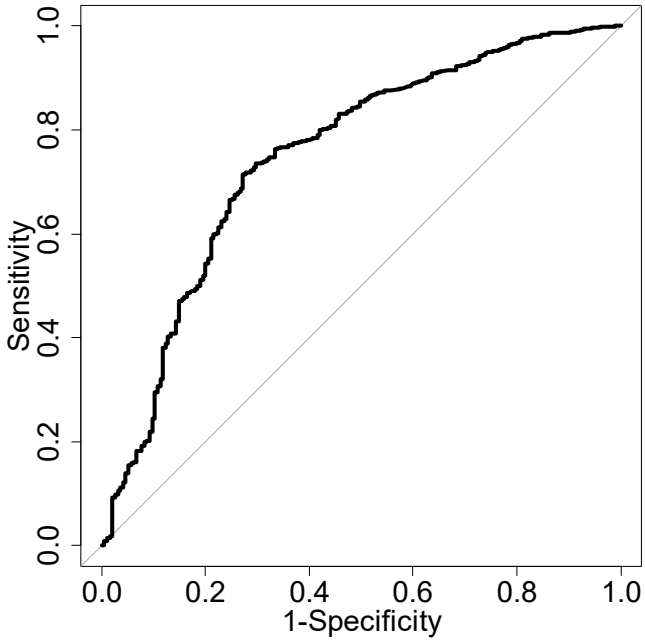



**Fig 5.2** ROC curve for the simplified binomial logistic regression model for infrastructure selection with bicycle facility

**Response to a red signal**

When faced with a red traffic signal, roughly a fifth of the observed bicyclists violated the signal. A small subset of predictors is found to provide the best prediction of red light violation (full model in Appendix 1). The manoeuvre carried out by a bicyclist plays a very strong role in whether or not he or she will stop at a red light. Bicyclists turning right have a 134.15 times increase in the ratio of probability to run a red light than those riding straight across the intersection. On one-way roads, bicyclists turning left have a 13.76 times increase in the ratio of probability to violate a red light (Roadway width – opposite = 0). For each meter of roadway width in the opposite direction, this probability increases by 1.33 times. This is because, here, carrying out an indirect left turn against the mandatory direction of travel usually includes violating the first traffic signal (the signal that would have been waited for if the bicyclist had carried out an indirect left turn with the mandatory direction of travel). The time elapsed since the signal became red has a deterring effect on red light violations; bicyclists have a 1.35 ($\frac{1}{e^{-0.03*10}}$) times decrease in the ratio of probability to violate the signal for each ten seconds passed since the signal change.

| **N = 1935**<br>**Stop = 0, Violate = 1** | $\beta$ | **Odds ratio** | $p$ |
|---|---|---|---|
| Intercept | -1.22 | 0.29 | 0.000 |
| Manoeuvre (right turn) | 4.90 | 134.16 | 0.000 |
| Time since signal change – red (s) | -0.03 | 0.97 | 0.000 |
| Manoeuvre (left turn) | 2.62 | 13.76 | 0.000 |
| Roadway width – opposite (m) | -0.28 | 0.76 | 0.000 |
| Manoeuvre (left turn) * Roadway width – opposite (m) | 0.28 | 1.33 | 0.008 |
| | | | |
| Classification threshold: | | 0.46 | |
| AUC | | 0.92 | |
| Accuracy | | 0.91 | |
| Sensitivity | | 0.85 | |
| Specificity | | 0.93 | |
| Positive predictive value | | 0.74 | |
| Negative predictive value | | 0.96 | |

**Tab. 5.5** Simplified binomial logistic regression model with evaluation for response to red signal

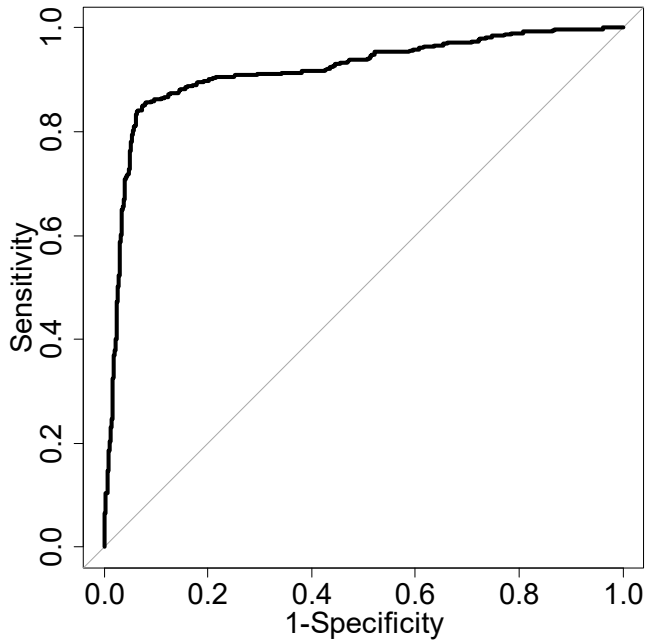



**Fig 5.3** ROC curve for the simplified binomial logistic regression model for response to red signal

The estimated binomial logistic regression model predicts the choice outcome with high accuracy. The prediction of signal compliance is slightly more reliable than that of signal violation. However, the prediction rates for both suggest that this behaviour is highly influenced by observable situational factors. In this case, the simplified model is found to provide excellent predictions with greater simplicity than the full model.

The resulting model supports previous studies that found that turning right increases the probability of violating a red light [JOHNSON ET AL., 2011]. The influence of additional parameters, such as the signal timing, infrastructure selection and left turn manoeuvre, are identified here.

**Direction of travel**

The vast majority of the observed bicyclists ride in the mandatory direction of travel (98.7%). An extensive full model and a greatly simplified model are estimated to predict this heavily skewed tactical decision (full model in Appendix 1 and simplified model in Tab. 5.6). According to the simplified model, bicyclists turning left have a 9.25 times increase in the ratio of probability to travel against the direction of travel than those carrying out other manoeuvres. The availability of a left turn lane discourages travelling against the direction of travel to a large extent (6.36 ($\frac{1}{e^{-1.85}}$) times decrease in ratio of probability). The presence of parking facilities increases the ratio of probability of riding against the mandatory direction of travel by 2.67 times while the presence of a separated bicycle path decreases the ratio of probability by 9.87 ($\frac{1}{e^{-2.29}}$) times. Together, these two factors increase the ratio of probability of riding the wrong way by 2.31 ($e^{-2.29+0.98+2.15}$) times.

The cross-validation indicates a high success rate for predicting riding with the direction of travel and a low success rate for predicting those riding against the given direction of travel. The very low classification threshold of 0.02 reflects the skewing in the choice observations and the low positive predictive value of 0.04 reflects the inaccuracy caused by manipulating the classification threshold. The full model does not offer a strong advantage over the simplified model.

Far fewer bicyclists were observed riding against the mandatory direction of travel in this study than reported by KULLER ET AL. [1986] and WACHTEL & LEWISTON [1994]. The only finding that could be verified in this study is that bicyclists turning left are more likely to ride the wrong way, reflecting the goal of path simplification that was noted by KULLER ET AL. [1986].

| **N =4710**<br>**With direction = 0, Against direction = 1** | **β** | **Odds ratio** | $p$ |
|---|---|---|---|
| Intercept | -4.50 | 0.01 | 0.000 |
| Manoeuvre (left turn) | 2.23 | 9.25 | 0.000 |
| Left turn lane | -1.85 | 0.16 | 0.000 |
| Bicycle facility type (separated) | -2.29 | 0.10 | 0.019 |
| Parking | 0.98 | 2.67 | 0.202 |
| Parking * Bicycle facility type (separated) | 2.15 | 8.60 | 0.035 |
| | | | |
| Classification threshold: | | 0.02 | |
| AUC | | 0.77 | |
| Accuracy | | 0.76 | |
| Sensitivity | | 0.78 | |
| Specificity | | 0.76 | |
| Positive predictive value | | 0.04 | |
| Negative predictive value | | 0.99 | |

**Tab. 5.6** Simplified binomial logistic regression model with evaluation for direction of travel

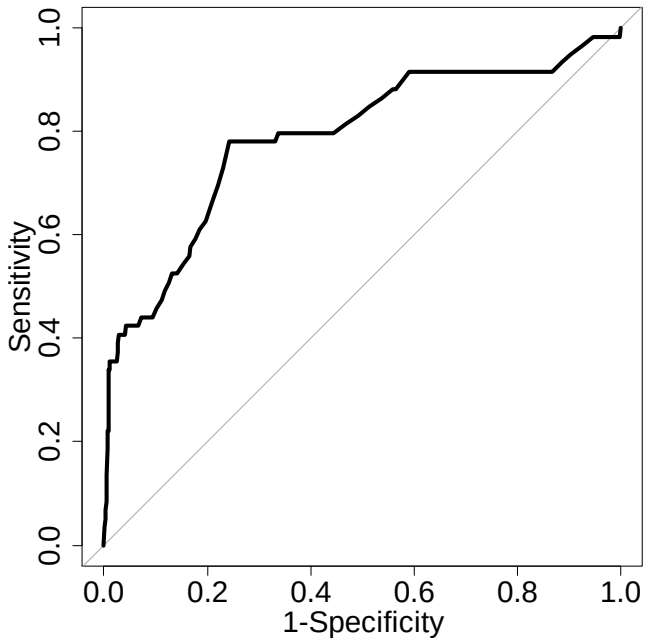



**Fig 5.4** ROC curve for the simplified binomial logistic regression model for direction of travel

**Left turn manoeuvre**

A multinomial regression model with three choice outcomes, direct left turn, indirect left turn and indirect left turn against the mandatory direction of travel, is estimated to predict the type of manoeuvre carried out by the bicyclist. The three choice outcomes are described in Section 2.1.1 and are shown graphically in Fig 3.2. The estimated $\beta$ parameters are in reference to the base category, which is the direct left-hand turn.

The most important predictor for the type of left turn is roadway use; bicyclists using the roadway have a more than 60 ($\frac{1}{e^{-4.12}}$) times decrease in the ratio of probability to carry out an indirect left turn and 20.5 ($\frac{1}{e^{-3.02}}$) times decrease in the ratio of probability to carry out an indirect left turn against the given direction of travel. Two characteristics of the infrastructure design, the type of bicycle facility and the presence of car parking, influence the choice outcome. The probability ratio of an indirect left turn increases by 3.42 times if there is only parking available and 6.05 times if only a separated bicycle path is present. If both these features are present, the ratio of probability of this manoeuvre increases by 4.43 ($e^{1.80+1.23-1.54}$) times. A similar mechanism is at play for the choice to execute an indirect left turn against the mandatory direction of travel (5.18 times increase with parking only, 9.54 times increase with a separated bicycle path only and 9.21 ($e^{2.26+1.64-1.68}$) times with both features). The signal phase and the presence of other road users also play important role in the left turn choice. A green signal encourages an indirect left turn (3.23 times increase in probability ratio) and discourages an indirect left turn against the mandatory direction of travel (8.41 ($\frac{1}{e^{-2.13}}$) times decrease in the probability ratio). For each bicyclist in the approach, the probability ratio of carrying out an indirect left turn decreases by 1.21 ($\frac{1}{e^{0.53+-0.73}}$) times and the probability ratio of carrying out an indirect left turn against the direction of travel decreases by 2.25 ($\frac{1}{e^{0.30+-1.11}}$) times.

In contrast to the multinomial regression model for infrastructure selection, which failed to predict roadway and sidewalk use, the multinomial regression model for the left turning manoeuvre provides exceptional predictions for all three types of turns. The predictive power of this model suggests that this decision is greatly influenced by observable situational factors at the intersection. The simplified model provides a slightly less accurate prediction of the turning manoeuvre but enables a more straightforward interpretation of the predictors.

The resulting model supports the findings of a previous study that found that bicyclists using the roadway often carry out direct left turns while bicyclists on the sidewalk and bicycle facility do not [AMINI ET AL., 2016]. The signal phase and the time since the last phase change are also found to play a role in manoeuvre selection, as previously found.

| **N = 426**<br>**Base category = Direct left turn** | | $\beta$ | **Odds ratio** | $p$ |
|---|---|---|---|---|
| **Indirect left turn** | Intercept | -0.40 | 0.67 | 0.704 |
| | Infrastructure selection (roadway) | -4.12 | 0.02 | 0.000 |
| | Bicycle facility type (separated) | 1.80 | 6.05 | 0.052 |
| | Parking | 1.23 | 3.42 | 0.165 |
| | Signal phase (green) | 1.17 | 3.23 | 0.023 |
| | Bicyclists in approach | 0.53 | 1.69 | 0.008 |
| | Parking * Bicycle facility type (separated) | -1.54 | 0.22 | 0.158 |
| | Signal phase (green) * Bicyclists in approach | -0.72 | 0.49 | 0.003 |
| **Indirect left turn (wrong way)** | Intercept | 0.56 | 1.74 | 0.606 |
| | Infrastructure selection (roadway) | -3.02 | 0.05 | 0.000 |
| | Bicycle facility type (separated) | 2.26 | 9.54 | 0.020 |
| | Parking | 1.64 | 5.18 | 0.076 |
| | Signal phase (green) | -2.13 | 0.12 | 0.000 |
| | Bicyclists in approach | 0.30 | 1.34 | 0.121 |
| | Parking * Bicycle facility type (separated) | -1.68 | 0.19 | 0.136 |
| | Signal phase (green) * Bicyclists in approach | -1.11 | 0.33 | 0.022 |
| Accuracy | | | 0.73 | |
| Mean Sensitivity | | | 0.70 | |
| Mean Specificity | | | 0.85 | |
| Mean Positive predictive value | | | 0.73 | |
| Mean Negative predictive value | | | 0.86 | |

**Tab. 5.7** Simplified multinomial regression model with evaluation for left turn manoeuvre

| | | True class | | |
|---|---|---|---|---|
| | | **Direct** | **Indirect** | **Indirect (ww)** |
| Predicted class | **Direct** | **41** | 6 | 8 |
| | **Indirect** | 13 | **104** | 17 |
| | **Indirect (ww)** | 12 | 56 | **169** |

**Tab. 5.8** Confusion matrix for the simplified multinomial regression model for left turn manoeuvre

## 5.2 Operational models

Operational behaviour models are required to move simulated bicyclists along their desired pathway across the intersection while reacting to other road users, obstacles and the signal control. This is realised by generating an acceleration vector $a(t)$ for each simulated bicyclist in every simulation time step. In car-following approaches, acceleration is reduced to a scalar quantity in the longitudinal direction. Longitudinal acceleration is calculated based on the position and velocity of a leading road user using the same link. The lateral position remains constant. An extension of this approach is typically used for simulating bicycle traffic in commercially or publically available simulation software such as PTV Vissim [FELLENDORF & VORTISCH, 2010] and SUMO [KRAJZEWICZ ET AL., 2014]. Traditional car-following models are extended to allow road users to vary their lateral position within a lane to optimise their progression in the longitudinal direction. However, using extended car-following models to simulate bicycle traffic has a number of disadvantages, most of which are caused by problems recreating the higher flexibility of bicyclists in comparison to vehicular traffic.

Here, a social force model for pedestrian movement, the NOMAD model [HOOGENDOORN, 2001], is adapted with the aim of simulating more flexible and realistic bicycle movement. The main advantage of using a force model over a car-following model is the potential for simulated bicyclists to interact with all other road users in a given vicinity rather than only the leading road users using the same one-directional link. However, in comparison with pedestrians, the dynamics of riding a bicycle limit the possible changes in direction and speed in each time step. It is, therefore, necessary to adapt social force models developed for pedestrians to represent realistic bicyclist movement.

The original NOMAD model is described in detail in the report *Normative pedestrian flow behavior theory and applications* [HOOGENDOORN, 2001] and a simplified version is presented, calibrated and validated in the paper *Microscopic Calibration and Validation of Pedestrian Models: Cross-Comparison of Models Using Experimental Data* [HOOGENDOORN & DAAMEN, 2007]. The simplified version of the NOMAD model for pedestrian movement specified by HOOGENDOORN & DAAMEN [2007] is given in Eq. 5.10.

$$a_p(t) = \frac{v_p^0 - v_p(t)}{T_p} - A_p \sum_{q \in Q_p} u_{pq}(t)\, e^{\frac{-d_{pq}(t)}{R_p}} \qquad \text{Eq. 5.10}$$

where:

$$d_{pq}(t) = \left\| r_q(t) - r_p(t) \right\| \qquad \text{Eq. 5.11}$$

$$u_{pq}(t) = \frac{r_q(t) - r_p(t)}{d_{pq}(t)} \qquad \text{Eq. 5.12}$$

The desired velocity, which is a two-dimensional vector that points to the desired destination of pedestrian $p$, is denoted as $v_p^0$, the velocity at time $t$ as $v_p(t)$ and the position as $r_p(t)$. The set of pedestrians within a certain vicinity of pedestrian $p$ is given by $Q_p$. Four pedestrian specific parameters, which remain constant for each pedestrian, are included in the model. These include the desired speed $V_p^0$, the necessary time for acceleration $T_p$, the interaction factor $A_p$ and radius of interaction $R_p$.

An extension to the basic model that accounts for the anisotropic behaviour of pedestrians was proposed by HOOGENDOORN & DAAMEN [2007] and is given by Eq. 5.13. Anisotropy describes the tendency for pedestrians to react mainly to other pedestrians directly in their desired path of travel. Pedestrians positioned behind the simulated pedestrian have no influence on the movement while those ahead of the simulated pedestrian but off to the side of the desired path have a relatively small influence.

$$a_p(t) = \frac{v_p^0 - v_p(t)}{T_p} - A_p \sum_{q \in Q_p} u_{pq}(t)\, e^{\frac{-d_{pq}^*(t)}{R_p}} 1_{u_{pq}(t)\cdot v_p(t)>0} \qquad \text{Eq. 5.13}$$

where:

$$d_{pq}^*(t) = \frac{u_{pq}(t)\cdot v_p(t)}{\|v_p(t)\|} + \eta_p \frac{u_{pq}(t)\cdot w_p(t)}{\|v_p(t)\|}\,[9] \qquad \text{Eq. 5.14}$$

The vector $w_p(t)$ is perpendicular to $v_p(t)$ with the same magnitude. Pedestrian $p$ only responds to other pedestrians in front of him or herself, which is denoted by $1_{u_{pq}(t)\cdot v_p(t)>0}$, an indicator function that takes the value 1 if the interacting pedestrian is in front of pedestrian $p$ and zero if it is behind. The factor $\eta_p$ is constant and pedestrian specific and denotes the reaction difference of pedestrian $p$ to obstacles or interacting road users in front or to the side of the pedestrian ($\eta_p > 1$). Higher values of $\eta_b$ indicate a stronger relative importance of road users directly in the direction of travel compared to those to the left or right. This is because the adjusted distance $d_{pq}^*(t)$ increases faster for higher values for $\eta_p$ for interacting road users further outside the direction of travel of pedestrian $p$. The vectors and distance scalar $d_{pq}(t)$ of the original and extended NOMAD model for pedestrian movement are shown schematically in Fig 5.5. In this case, pedestrian $p$ will move to the right due to the positions of the destination and pedestrian $q$.

[9] In this disseration, dot products are denoted using a · symbol and cross products are denoted using a × symbol.

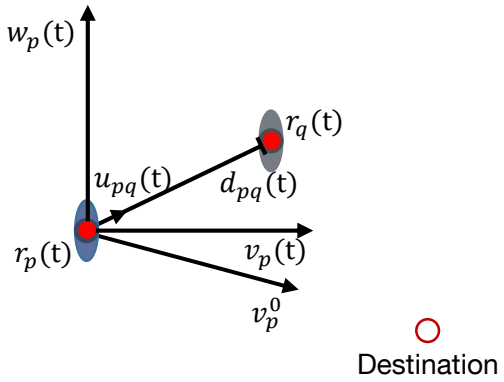



**Fig 5.5** Schematic of the vectors and distance scalar $d_{pq}(t)$ of the original and extended NOMAD model for pedestrian movement

An extension of this model was developed by HOOGENDOORN & DAAMEN [2007] in which the delayed response of pedestrians to the situation is considered. This is realised through the addition of a reaction time $\tau_p$ such that $a_p(t)$ in Eq. 5.10 and Eq. 5.13 is replaced by $a_p(t+\tau_p)$. The three models (basic, anisotropic and delayed) are calibrated and validated using trajectory data from walking experiments on the university campus. This modelling concept and the approach for calibrating and validating the model presented by HOOGENDOORN & DAAMEN [2007] is used as the foundation for model specification, calibration and validation in this section.

### 5.2.1 Model specification

The NOMAD model is adapted here by using the norm and angle representation of the acceleration vector $a(t)$ and developing independent models to determine the change in speed $\Delta V(t) = \| a(t)\|$ and a change in direction $\Delta\theta(t) = \tan^{-1}[a_y(t)/a_x(t)]$ in each time step. This enables the practical restriction of both $\Delta V(t)$ and $\Delta\theta(t)$ to reflect the dynamics of bicycling. Please note that vectors are signified using lower case letters and scalar quantities are shown using capital letters in this dissertation. The vector sign is omitted for simplicity.

**Basic model:**

The formulation of the separated norm and angle model for bicyclist behaviour is given in Eq. 5.15 and Eq. 5.16, respectively.

$$\Delta V_b(t) = \frac{V_b^0 - V_b(t)}{T_{vb}} - A_{vb}\, e^{\frac{-min_i\{D_s, D_{bi}(t)\}}{R_{vb}}} 1_{\phi<\frac{\pi}{2}} \qquad \text{Eq. 5.15}$$

$$\Delta\theta_b(t) = \frac{\theta_b^0(t) - \theta_b(t)}{T_{\theta b}} - A_{\theta b} \sum_{i\in IRU} U_{bi}(t)\, e^{\frac{-D_{bi}(t)}{R_{\theta b}}} 1_{\phi<\frac{\pi}{2}} \qquad \text{Eq. 5.16}$$

where:

$$U_{bi}(t) = \frac{v_b(t) \times d_{bi}(t)}{D_{bi}(t)\, V_b(t) \sin\phi} \qquad \phi = \cos^{-1}\frac{d_{bi}(t) \cdot v_b(t)}{D_{bi}(t)\, V_b(t)} \qquad \text{Eq. 5.17}$$

In the change in speed equation (Eq. 5.15), the parameter $V_b^0$ is the desired speed of bicyclist $b$, $V_b(t)$ is the current speed at time $t$, $T_{vb}$ is a speed relaxation parameter unique to bicyclist $b$ and $R_{vb}$ is the radius of interaction for bicycle $b$ regarding the speed adjustment. The set of other road users within a predefined radius (e.g. 10 m) is given by $IRU$. The distance between road user $i$ and bicyclist $b$ is given by the vector $d_{bi}(t) = p_i(t) - p_b(t)$, the scalar quantity of which is $D_{bi}(t) = \|d_{bi}(t)\|$. A graphical representation of the vectors and angles used in Eq. 5.15 – Eq. 5.17 is shown in Fig 5.6.

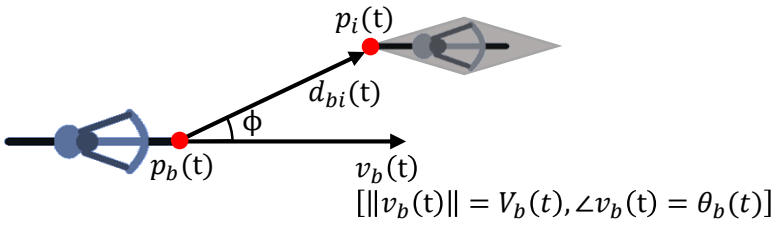



**Fig 5.6** Graphical representation of the vectors and angles included in the specified models

In response to the presence of the interacting road user $i$ in Fig 5.6, the depicted bicyclist $b$ will reduce their speed ($\Delta V_b(t) < 0$) and will change direction away from the interacting road user in the clockwise direction ($\Delta\theta_b(t) < 0$).

In addition to separating the model into the norm and angle representation of the velocity vector, the original model presented by HOOGENDOORN & DAAMEN [2007] is adapted in two important ways. First, the reaction to a traffic signal is included directly in the change in speed model. This is done by mimicking the reaction of a bicyclist to a large interacting road user that cannot be passed. The outline of the signalised intersection is denoted by a polygon that connects the stop lines of all the approaches. The nearest point on the stop line polygon to $p_b(t)$ is selected as $p_s(t)$. The distance vector $d_s(t) = p_s(t) - p_b(t)$ and norm $D_s(t) = \|d_s(t)\|$ are analogous to the variables defined for interacting road users. This is only done for the change in speed model because bicyclists cannot manoeuvre around traffic signals.

Second, instead of using a constant bicycle specific parameter to control the interaction response $A_{vb}$, a variable is defined based on the current speed $V_b(t)$, the desired speed $V_b^0$ and the speed relaxation parameter $T_{vb}$ that ensures simulated bicyclists are able to stop in any simulation second. If the current speed $V_b(t)$ is large, the interaction response $A_{vb}$ also increases in magnitude to allow for a stronger deceleration response to avoid collisions in all possible situations. In addition, this conversion reduces the number of parameters in the model by one, enabling a more stable prediction of the remaining model parameters.

Here:

$$A_{vb} = \frac{V_b^0 + (T_{vb} - 1)\, V_b(t)}{T_{vb}} \qquad \text{Eq. 5.18}$$

The equation describing the change in direction $\Delta\theta_b(t)$ (Eq. 5.16) is formulated similarly; $\theta_b(t)$ is the direction of travel of bicyclist $b$ at time $t$, and $T_{\theta b}$, $A_{\theta b}$ and $R_{\theta b}$ are constant bicyclist specific parameters controlling the change of direction specifically. $U_{bi}(t)$ is introduced to specify the position of the interacting road user with respect to the desired path of travel and enables bicyclist $b$ to move to the left in response to a road user on the right and vice versa. $U_{bi}(t)$ can take a value of either $-1$ or $1$ depending on the side of the interacting road user relative to the velocity of bicycle $b$. The main difference between Eq. 5.15 and Eq. 5.16 is that $\theta_b^0(t)$ is not a static parameter, as $V_b^0$ is in Eq. 5.15, but rather changes to guide bicyclist $b$ along his or her desired pathway across the intersection. The desired change in direction based on the desired position of bicyclists $b$ at time $t$ is shown in the following figure.

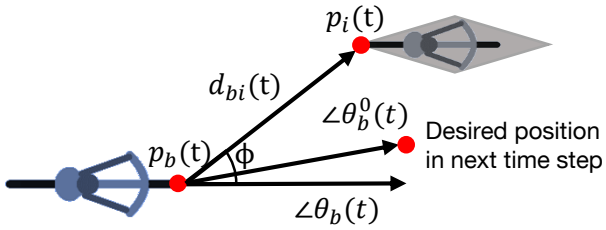



**Fig 5.7** Graphical representation of the definition of desired change of direction

The summation across all interacting road users in the set $IRU$ is restricted to the change in angle equation (Eq. 5.16). This reflects the interacting behaviour of road users. It is presumed that a bicyclist adapts his or her speed based only on the most critical of interacting road users. For example, when riding in a single file platoon of bicyclists, a bicyclist does not ride slower if there are many bicyclists ahead in the platoon than he would if there were only one bicyclist ahead. The speed is determined to prevent a collision with the most critical interacting road user. In contrast, the direction is adapted as a response to many other road users within a given area. This enables the bicyclist to manoeuvre through a group of bicyclists. Furthermore, it is presumed that bicyclists do not react to road users positioned behind themselves and to this effect the indicator function $1_{\phi<\pi/2}$ is deployed in all models including the basic model. This presumption was confirmed through initial evaluations of models including interacting road users in all directions.

**Anisotropic model:**

The first variation of the basic model examined here is an adaptation of the anisotropic model proposed by HOOGENDOORN & DAAMEN [2007], which takes into account the position of the interacting road user $i$ with respect to the direction of travel of bicyclist $b$. Using this approach,

road users directly in the line of travel of bicyclist $b$ have the largest impact on the change in speed $\Delta V_b(t)$ and direction $\Delta\theta_b(t)$. The formulation of the anisotropic model is given in Eq. 5.19 and Eq. 5.20:

$$\Delta V_b(t) = \frac{V_b^0 - V_b(t)}{T_{vb}} - A_{vb}\, e^{\frac{-min_i\{D_s^*, D_{bi}^*(t)\}}{R_{vb}}} 1_{\phi<\frac{\pi}{2}} \quad \text{Eq. 5.19}$$

$$\Delta\theta_b(t) = \frac{\theta_b^0(t) - \theta_b(t)}{T_{\theta b}} - A_{\theta b} \sum_{i \in IRU} U_{bi}(t)\, e^{\frac{-D_{bi}^*(t)}{R_{\theta b}}} 1_{\phi<\frac{\pi}{2}} \quad \text{Eq. 5.20}$$

where:

$$D_{bi}^*(t) = \frac{d_{bi}(t) \cdot v_b(t)}{V_b(t)} + \eta_b \frac{d_{bi}(t) \cdot w_b(t)}{V_b(t)} \quad \text{Eq. 5.21}$$

Here, the distance between bicyclist $b$ and road user $i$ is divided into two components, one parallel and the other perpendicular to the direction of travel of bicyclist $b$. The vector $w_b(t)$ is perpendicular to $v_b(t)$ with the same scalar quantity and is oriented in the direction of road user $i$. The bicycle specific parameter $\eta_b$ describes the weighting of the distance of two components relative to one another ($\eta_b > 1$). Two different values of $\eta_b$ are solved for in the model, one for the $\Delta V_b(t)$ component ($\eta_{vb}$) and one for the $\Delta\theta_b(t)$ component ($\eta_{\theta b}$).

**Velocity anisotropic model:**

A final extension is proposed here to include the direction of travel of the interacting road user in the change of speed and change of direction model. The behaviour hypothesis behind this extension is that bicyclists have a less pronounced response to interacting road users moving with a similar velocity. For example, a bicyclist moving in a platoon of other bicyclists will only make minimal adjustments to his or her speed in response to a leading bicyclist moving in the same direction with a similar speed, even though this leading bicyclist may be very close in terms of distance. In contrast, the response to another road user moving towards bicycle $b$ will be much larger, even if this road user is further away. To account for this aspect of behaviour, the effective distance $D_{bi}^{**}(t)$ is increased or decreased depending on the velocity vector of the interacting road user in relation to that of bicyclist $b$.

$$\Delta V_b(t) = \frac{V_b^0 - V_b(t)}{T_{vb}} - A_{vb}\, e^{\frac{-min_i\{D_s^*, D_{bi}^{**}(t)\}}{R_{vb}}} 1_{\phi<\frac{\pi}{2}} \quad \text{Eq. 5.22}$$

$$\Delta\theta_b(t) = \frac{\theta_b^0(t) - \theta_b(t)}{T_{\theta b}} - A_{\theta b} \sum_{i \in IRU} U_{bi}(t)\, e^{\frac{-D_{bi}^{**}(t)}{R_{\theta b}}} 1_{\phi<\frac{\pi}{2}} \quad \text{Eq. 5.23}$$

where:

$$D_{bi}^{**}(t) = \frac{d_{bi}(t) \cdot v_b(t)}{V_b(t)} + \eta_b \frac{d_{bi}(t) \cdot w_b(t)}{V_b(t)} + \gamma_b \frac{v_i(t) \cdot v_b(t)}{V_i(t)\, V_b(t)} \quad \text{Eq. 5.24}$$

The parameter $\gamma_b$ reflects the bicycle specific adjustment of the effective distance depending on the similarly between the velocity of the interacting road user $i$ and that of bicycle $b$. The cosine of the angle between $v_i$ and $v_b$ ($\frac{v_i(t) \cdot v_b(t)}{V_i(t)\, V_b(t)}$) ranges between $-1$ and $1$ and as such the parameter $\gamma_b$ represents the adjustment of the effective distance in meters. Two different values of $\gamma_b$ are solved for in the model, one for the $\Delta V_b(t)$ component ($\gamma_{vb}$) and one for the $\Delta\theta_b(t)$ component ($\gamma_{\theta b}$).

### 5.2.2 Model calibration

For each of the proposed models, there are a number of bicyclist specific parameters to be calibrated using the observed trajectory data (Tab. 5.9).

| **Model** | $\Delta \boldsymbol{V_b(t)}$ | $\Delta \boldsymbol{\theta_b(t)}$ |
|---|---|---|
| Basic | $V_b^0, T_{vb}, R_{vb}$ | $T_{\theta b}, A_{\theta b}, R_{\theta b}$ |
| Anisotropic | $V_b^0, T_{vb}, R_{vb}, \eta_{vb}$ | $T_{\theta b}, A_{\theta b}, R_{\theta b}, \eta_{\theta b}$ |
| Velocity anisotropic | $V_b^0, T_{vb}, R_{vb}, \eta_{vb}, \gamma_{vb}$ | $T_{\theta b}, A_{\theta b}, R_{\theta b}, \eta_{\theta b}, \gamma_{\theta b}$ |

**Tab. 5.9** Model parameters to be calibrated

Observed values for $p_b(t)$, $v_b(t)$ and $p_i(t)$, $v_i(t)$ for all interacting road users in the set $IRU$ are extracted from the trajectory data for each time step $t$ for each bicyclist $b$. Each trajectory has the form $S_b = [(x_b, y_b)_{t=0} \quad (x_b, y_b)_{t=1} \quad \cdots \quad (x_b, y_b)_{t=T_b}]$, where $(x_b, y_b)_t$ is the position coordinate of bicyclist $b$ at time $t$ and $T_b$ is the duration of the trajectory. Here, the position observations are aggregated into groups of three and therefore the frequency of 25 observations per second is reduced to 8.33 (25/3) observations per second. The noise in the position data was partially resolved through this aggregation. The aggregated position points are smoothed using the Savitzky–Golay filter [SAVITZKY & GOLAY, 1964]. The change in speed and change in direction observations are derived from the smoothed and aggregated trajectories.

The remaining vectors $d_{bi}(t)$ and $w_b(t)$, angle $\phi$ and the value $U_{bi}(t)$ are calculated from the $p_b(t)$, $v_b(t)$, $p_i(t)$ and $v_i(t)$ observations. In each time step, a presumed desired direction $\theta_b^0(t)$ is inferred using the representative trajectory for the cluster to which bicyclist $b$ is found to belong (Section 4.1). The representative trajectory is assumed to embody the desired position points of each bicyclist along their trajectory. Without this assumed desired trajectory, it is not possible to generate a $\theta_b^0(t)$ in each time step. The desired direction in time step $t$ is found by locating the point on the representative trajectory nearest to $p_b(t)$ (point $l_1$

in Fig 5.8). A second point is specified at a distance $\overline{V^0}$ further along the representative trajectory (point $l_1 + \overline{V^0}$ in Fig 5.8). The first point is necessary to act as a start point for measuring the distance $\overline{V^0}$ along the representative trajectory. A vector is drawn between $p_b(t)$ and this second point, the angle of which is taken to be $\theta_b^0(t)$. A graphical representation of this approach is shown in Fig 5.8.

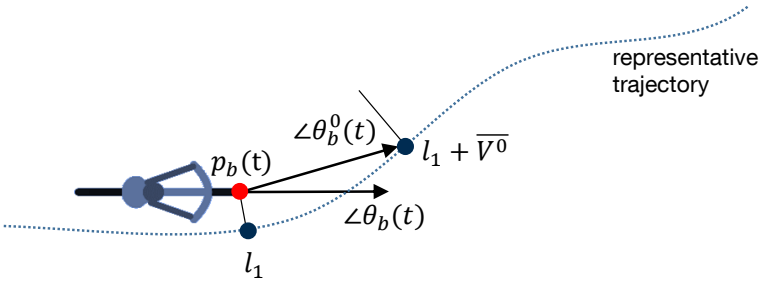



**Fig 5.8** Graphical representation of the approach for finding $\theta_b^0(t)$

The model parameters are calibrated to fit the observed behaviour using Maximum Likelihood Estimation (MLE). This method provides a means for deriving the values of a set of parameters $\beta = \{\beta_0, \beta_1, \beta_2, \dots \beta_m\}$ in a model to best fit a sample of data. This is achieved by expressing the likelihood as a joint probability mass function of the sample of observations as shown Eq. 5.25 and Eq. 5.26. The likelihood $L(\beta)$ is a function of the parameter set $\beta$ and is maximized to find the optimal set of parameters.

$$L(\beta) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \quad \text{Eq. 5.25}$$

$$L(\beta) = f(x_1;\beta) \cdot f(x_2;\beta) \cdots f(x_n;\beta) = \prod_{i=1}^{n} f(x_i;\beta) \quad \text{Eq. 5.26}$$

Assuming that the observations are normally distributed, the probability mass function can be expressed using Eq. 5.27 and the joint probability mass function or likelihood of $\beta$ is given by Eq. 5.28

$$f(x_i;\beta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{(x_i^{pred} - x_i^{obs})^2}{2\sigma^2}\right)} \quad \text{Eq. 5.27}$$

$$L(\beta) = \prod_{i=1}^{n} f(x_i;\beta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\left(-\frac{(x_i^{pred} - x_i^{obs})^2}{2\sigma^2}\right)} \quad \text{Eq. 5.28}$$

For mathematical convenience, the log of the likelihood $\mathcal{L}(\beta)$ is typically maximized:

$$\mathcal{L}(\beta) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(x_i^{pred} - x_i^{obs}\right)^2 \quad \text{Eq. 5.29}$$

The standard deviation $\sigma^2$ must be determined in order to numerically solve for the best fitting parameter set $\beta$. The maximum likelihood estimator of the variance is given in Eq. 5.30.

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i^{pred} - x_i^{obs}\right)^2 \quad \text{Eq. 5.30}$$

The log likelihood function given $\hat{\sigma}^2$ is expressed in Eq. 5.31. Using this function, the parameter set $\hat{\beta}$ can be solved for using numerical optimization.

$$\mathcal{L}(\beta;\hat{\sigma}^2) = -\frac{n}{2}\ln\left(\frac{2\pi}{n}\sum_{i=1}^{n}\left(x_i^{pred} - x_i^{obs}\right)^2\right) - \frac{n}{2} \quad \text{Eq. 5.31}$$

$$\hat{\beta} = \arg\max \mathcal{L}(\beta;\hat{\sigma}^2) \quad \text{Eq. 5.32}$$

This method is used to calibrate the parameters in the models specified in Section 5.2.1 for each of the observed bicyclists at three of the four research intersections. The parameter set for each bicyclist is denoted as $\beta_b$ and includes the parameters shown in Tab. 5.9. The model parameters are calibrated for each observed bicyclist using the following equations:

$$\mathcal{L}(\beta_b;\hat{\sigma}_b^2) = -\frac{n}{2}\ln\left(\frac{2\pi}{n}\sum_{i=1}^{n}\left(a_i^{pred} - a_i^{obs}\right)^2\right) - \frac{n}{2} \quad \text{Eq. 5.33}$$

$$\hat{\beta}_b = \arg\max \mathcal{L}(\beta_b;\hat{\sigma}_b^2) \quad \text{Eq. 5.34}$$

where $n$ is the number of observation points along the aggregated trajectory. To ensure that $n$ is large enough to find stable estimates of $\hat{\beta}_b$, samples with fewer than 50 observations are filtered from the dataset. This reflects the recommendation by LONG [1997] to include at least 10 observations per parameter. The majority of the trajectories include between 100 and 250 observation points.

The maximum log likelihood is numerically solved for using the *SciPy* Python implementation [THE SCIPY COMMUNITY, 2016] of the Constrained Optimization BY Linear Approximation (COBYLA) algorithm proposed by POWELL [1994]. Using this method, it is possible to set constraints that prevent the algorithm from locating an illogical, but mathematically optimal, minimum (e.g. extremely large desired velocity $V_b^0$ and relaxation time $T_b$ pairs). Initial estimates of the parameters are supplied to the algorithm. The values of these initial estimates are determined using an iterative process in which a set of values was specified, the models

were calibrated and the initial estimate was compared to the resulting parameter distribution. The same initial estimates were used for all observed bicyclists. The means of the resulting parameter distributions were used as the new initial estimates and the model was recalibrated. This process was terminated once stable initial estimates and parameter means were achieved. This prevented the algorithm from locating illogical minima to a large extent.

### 5.2.3 Model validation

The calibrated models are validated using K-fold cross-validation. Using this method, the sample of observations ($n$ per bicyclist) is randomly divided into $K$ mutually exclusive sub-samples of approximately equal size. The calibration of the parameter set is repeated $K$ times (or folds). In each fold, a sub-sample is held back from the model calibration and the parameter set $\hat{\beta}_{b\ sub-sample}$ is estimated using the remainder of the dataset. The calibrated model is then validated using the held back sub-sample. The model predictions are made by taking the observed position and velocity of the road users at each time step, extracting or calculating the model vectors and scalars and computing the predicted $\Delta V_b(t)$ and $\Delta\theta_b(t)$. These predictions are compared to the actual change in speed and angle observed at that time step. This type of validation reduces potential influence from the random splitting of the data because each of the observations is used exactly once for validation.

In order to assess the predictive power of the models, the performance is compared to the constant velocity model in which acceleration equals zero in each time step (null case). Although this model is not capable of simulating traffic, it provides a useful method to determine if the predictions made in each time step are significantly better than a prediction of $0$. Two measures are used to compare the developed models with this base model; the average improvement in log likelihood and the log likelihood ratio test. The average improvement in log likelihood, which indicates the overall improvement in model performance, is given by Eq. 5.35:

$$I = \frac{\sum_{b\in B}\mathcal{L}(\beta_b;\hat{\sigma}^2) - \sum_{b\in B}\mathcal{L}_{b\ null}}{\left\|\sum_{b\in B}\mathcal{L}_{b\ null}\right\|} \qquad \text{Eq. 5.35}$$

Where $B$ is the set of observed bicyclists and $\mathcal{L}_{b\ null}$ is the log likelihood of the constant velocity model. The log likelihood ratio test is a statistic that enables the comparison between models of differing complexities. Because model estimations inherently improve with each additional parameter, the magnitude of this improvement with respect to the increase in degrees of freedom must be examined to determine if a complex model is actually better. The test statistic for the log likelihood ratio test is given by:

$$D = 2\ (\mathcal{L}(\beta_b;\hat{\sigma}_b^2) - \mathcal{L}_{b\ null}) \qquad \text{Eq. 5.36}$$

$D$ is compared with the critical value $\chi^2$ from the chi-squared distribution with degrees of freedom $df = df_{alternative} - df_{null}$. The degrees of freedom $df_{alternative}$ is the number of parameter for each model shown in Tab. 5.9. Models are accepted as significantly better if $p < 0.1$. The percentage of models (one model per observed bicyclist) that pass this test is used as an assessment measure for the overall model performance. The calibration and validation algorithm implemented here is shown in pseudo code below. All three model variations are tested using this approach.

**for** *each bicyclist* $b \in B$ **do**
  extract sample $K$ from trajectory ($\Delta V_b(t)$, $\Delta\theta_b(t)$, $p_b(t)$, $v_b(t)$, $p_i(t)$, $v_i(t)$);
  calculate $\mathcal{L}_{bnull}$ for constant velocity model;
  divide sample into $k = 10$ mutually exclusive sub-samples;
  **for** *each sub-sample k* **do**
    find $\hat{\beta}_{bsub-sample} = \arg\max \mathcal{L}(\beta; \hat{\sigma}^2)$ using training sample $K - k$;
    predict outcome for sub-sample $k$ with $\hat{\beta}_{bsub-sample}$;
  **end**
  calculate $\mathcal{L}(\hat{\beta}_{bsub-sample}; \hat{\sigma}^2)$ for all predicted outcomes in $K$;
  perform log likelihood ratio test to compare $\mathcal{L}_{bnull}$ and $\mathcal{L}(\hat{\beta}_{bsub-sample}; \hat{\sigma}^2)$;
  **if** $p < 0.1$ **then**
    find $\hat{\beta}_b = \arg\max \mathcal{L}(\beta_b; \hat{\sigma}^2)$ using entire sample $K$;
    append $\hat{\beta}_b$ to population parameter set;
  **end**
**end**
calculate $I = \frac{\sum_{b\in B} \mathcal{L}(\beta_b;\hat{\sigma}^2) - \sum_{b\in B} \mathcal{L}_{b-null}}{\left\|\sum_{b\in B} \mathcal{L}_{b-null}\right\|}$;

The models are tested using different delay times that represent the reaction times of the observed bicyclists. It is presumed that, like vehicle drivers, bicyclists do not respond immediately to stimuli in the road environment. Instead, a certain amount of time is required to collect sensory information, process this data and decide upon an appropriate response. Mean reaction times for car drivers have been estimated to lie between 0.7 – 1.5 s [GREEN, 2007]. To reflect this reaction time, the acceleration observations are collected at $t + \tau$ for observations from time step $t$. Values of $\tau$ ranging between 0.0 s and 1.5 s are tested and the model quality is assessed using $I$ (Eq. 5.35) and the percentage of models that pass the log likelihood ratio test (Eq. 5.36).

### 5.2.4 Results

The results of the calibration and K-fold cross-validation are presented in this section. The parameter distributions for the calibrated change in speed model are shown and are briefly discussed in the first section followed by the results for the calibrated change in angle model. The best-suited reaction time $\tau$ is located by examining the overall percent increase in the log

likelihood $I$ and the percent of calibrated models that pass the log likelihood ratio test for each reaction time $\tau$ between 0 s and 1.5 s at an interval of 0.12 s. The lowest reaction time $\tau$ leading to markedly improved model performance is selected and the calibrated model parameters for this reaction time are presented.

**Change in speed**

The evaluation measures for the varying reaction time $\tau$ values are shown in Fig 5.9. A reaction time $\tau$ of 1.2 s leads to the best prediction of change in speed values. This lies within the range suggested by GREEN [2007] for car drivers but is considerably larger than the 0.28 s found by HOOGENDOORN & DAAMEN [2007] for pedestrians.

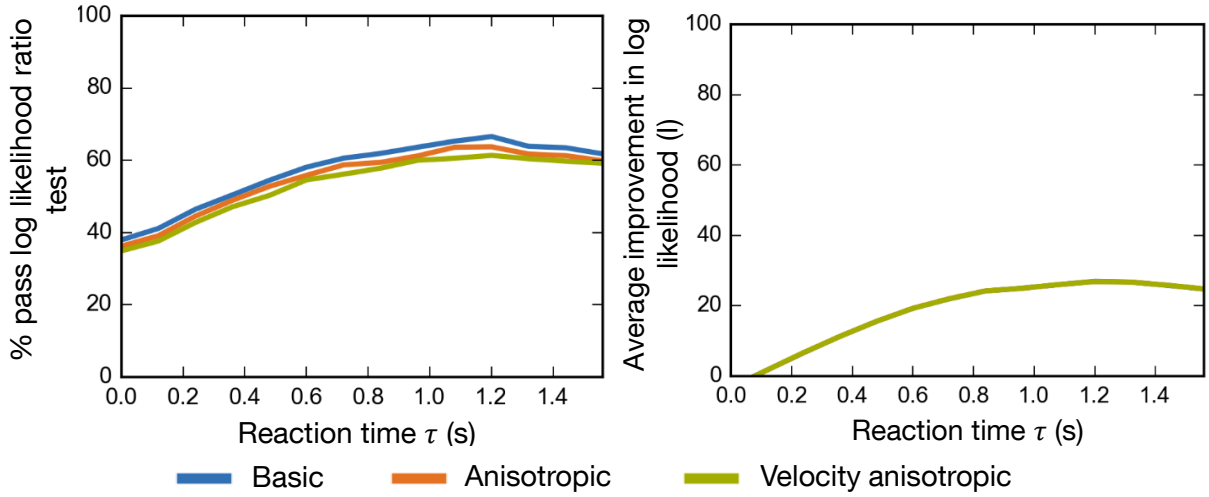



**Fig 5.9** Evaluation of the calibrated change in speed models for varying reaction times $\tau$

Based on a qualitative assessment of the distributions shown in Fig 5.10 (page 122), the calibrated model parameters are deemed to be roughly represented by the normal distribution, although the lognormal distribution may be more appropriate for the parameters $T_{vb}$ and $\gamma_{vb}$. Nevertheless, all parameters here defined using the mean and standard deviation, which are listed in Tab. 5.10 to Tab. 5.12. The magnitudes of these distribution parameters fall within an expected range of values and the consistency between parameters occurring in multiple models, $V_b^0$, $R_{vb}$, $T_{vb}$ and $\eta_{vb}$, indicates stability of the parameters outputs from the maximum likelihood estimation. The resulting parameter distributions of the basic and anisotropic models closely resemble those shown for the velocity anisotropic model shown in Fig 5.10.

The correlations between the model parameters are also important to consider and include in the subsequent traffic simulation to avoid incompatible parameter combinations. The Pearson correlation coefficient between the parameters is calculated using the formula below:

$$R_{ab} = \frac{n\sum ab - \sum a \sum b}{\sqrt{(n\sum a^2 - (\sum a)^2)(n\sum b^2 - (\sum b)^2)}} \qquad \text{Eq. 5.37}$$

where $n$ is the number of observations, and $a$ and $b$ are variables for which the correlation is to be determined. The resulting correlation ranges between -1 and 1, with 0 indicating no correlation between parameters.

| N=634 | | Desired speed $V_b^0$ (m/s) | Radius of interaction $R_{vb}$ (m) | Relaxation time $T_{vb}$ (s) |
|---|---|---|---|---|
| Mean [CI] | | 5.20 [5.07, 5.33] | 3.07 [3.00, 3.15] | 3.75 [3.56, 3.94] |
| Std. | | 1.47 | 0.83 | 2.15 |
| Correlation $R_{ab}$ | $V_b^0$ | 1 | -0.10 | 0.04 |
| | $R_{vb}$ | - | 1 | 0.05 |
| | $T_{vb}$ | - | - | 1 |

**Tab. 5.10** Calibration results for $\Delta V_b(t+\tau)$ basic model (reaction time: $\tau = 1.2\ s$)

| N=634 | | Desired speed $V_b^0$ (m/s) | Radius of interaction $R_{vb}$ (m) | Relaxation time $T_{vb}$ (s) | Anisotropic factor $\eta_{vb}$ |
|---|---|---|---|---|---|
| Mean [CI] | | 5.22 [5.09, 5.36] | 3.13 [3.05, 3.21] | 3.78 [3.59, 3.98] | 2.05 [2.02, 2.09] |
| Std. | | 1.46 | 0.91 | 2.15 | 0.39 |
| Correlation $R_{ab}$ | $V_b^0$ | 1 | -0.07 | 0.05 | 0.24 |
| | $R_{vb}$ | - | 1 | 0.03 | -0.14 |
| | $T_{vb}$ | - | - | 1 | 0.08 |
| | $\eta_{vb}$ | - | - | - | 1 |

**Tab. 5.11** Calibration results for the $\Delta V_b(t+\tau)$ anisotropic model (reaction time: $\tau = 1.2\ s$)

| N=634 | | Desired speed $V_b^0$ (m/s) | Radius of interaction $R_{vb}$ (m) | Relaxation time $T_{vb}$ (s) | Anisotropic factor $\eta_{vb}$ | Velocity direction factor $\gamma_{vb}$ |
|---|---|---|---|---|---|---|
| Mean [CI] | | 5.24 [5.11, 5.37] | 3.10 [3.02, 3.17] | 3.81 [3.61, 4.00] | 2.05 [2.02, 2.08] | 1.03 [1.00, 1.05] |
| Std. | | 1.44 | 0.82 | 2.14 | 0.32 | 0.30 |
| Correlation $R_{ab}$ | $V_b^0$ | 1 | -0.13 | 0.09 | 0.19 | 0.19 |
| | $R_{vb}$ | - | 1 | 0.06 | -0.04 | -0.02 |
| | $T_{vb}$ | - | - | 1 | 0.08 | 0.06 |
| | $\eta_{vb}$ | - | - | - | 1 | 0.52 |
| | $\gamma_{vb}$ | - | - | - | - | 1 |

**Tab. 5.12** Calibration results for $\Delta V_b(t+\tau)$ velocity anisotropic model (reaction time: $\tau = 1.2\ s$)

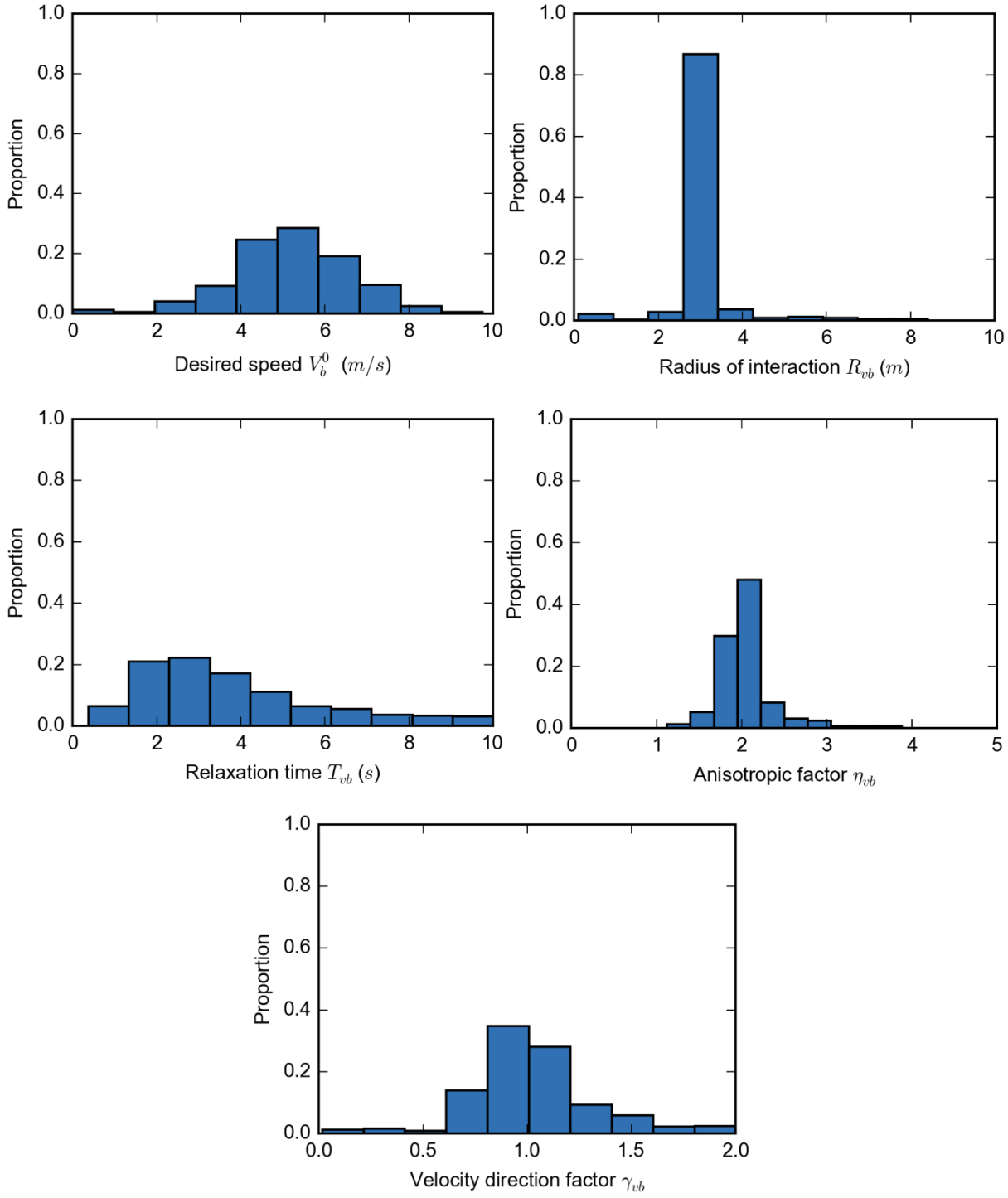



**Fig 5.10** Distributions of calibrated model parameters for the $\Delta V_b(t+\tau)$ velocity anisotropic model

In the meta-analysis of 28 studies that measure the speed of bicyclists (Fig 2.6), a median value of 4.6 m/s (16.5 km/h) was found. The higher value of 5.2 m/s (18.7 km/h) found here is logical because this value represents the desired speed of a population of bicyclists rather than the realised or observed speeds. Realised speeds are per definition lower as the bicyclist must slow to avoid other road users and obstacles and react to the signal control.

The relaxation time is a proxy measure for the maximum acceleration. If the mean relaxation time of 3.8 s is combined with the mean desired speed, a maximum acceleration of 1.4 m/s$^2$ emerges. The acceleration rate is slightly higher than those found in the literature (Section 2.1.2) but falls within a reasonable range. This is also expected as it is based on desired speed and not observed speed. A radius of interaction of 3.1 m appears reasonable but cannot be assessed in comparison to the findings of other research because none were found. The applicability of this value is instead tested in the simulation environment (Section 6). The mean and standard variations of the parameters $\eta_{vb}$ and $\gamma_{vb}$ are realistic and signify an average bicyclist that weights interacting road users directly in the path of travel with roughly twice the importance of those directly to the side. Roughly 1 m is subtracted from the effective distance $D_{bi}^{**}$ if the interacting road user is travelling with the same velocity as bicyclist $b$ and added if the velocity is opposite.

The correlations between the calibrated model parameters are quite small. The most noteworthy correlations are the positive correlation between the desired speed and both the anisotropic and velocity direction factor, the negative correlation between the desired speed and the radius of interaction and the correlation between the anisotropic factor and the velocity direction factor. Together these correlations indicate that bicyclists who aim to travel with a higher speed have a slightly smaller interaction zone and are more focused on interacting road users directly in their planned pathway and those with an opposing velocity vector. This may point to a group of bicyclists who are faster, more confident and have a high risk acceptance.

**Change in angle**

The equation for predicting the change in angle proved to be much more difficult to calibrate than that for the change in speed. This is due to the inherent difficulty in isolating the desired change in angle. Here, the representative trajectory of a cluster of bicyclists is used to approximate the desired direction of travel of a bicyclist at each position while crossing the intersection. The desired direction in each time step $\theta_b^0(t)$ is therefore based on the representative trajectory for the cluster to which bicyclist $b$ is assigned. It is very likely, however, that the actual desired direction $\theta_b^0(t)$ varied (slightly or greatly) from the approximated value. Another difficulty arises due to noise in the trajectory data. Although the high angle of the video camera to the intersections and smoothing the trajectory data reduced the amount of noise, small variations in the tracked centroid likely have a small impact on the quality of the model calibration. Nevertheless, the calibrated models proved to be a significant improvement over the constant velocity model for more than half of the observed bicyclists. Considering that $\Delta\theta_b(t)$ lies very close to zero in each time step of 0.12 s and the lack of concrete data describing the desired direction, the results presented below are exceptional.

The evaluation parameters for the calibrated models, the average improvement in log likelihood $I$ (Eq. 5.35) and percent passing the log likelihood ratio test, for reaction times $\tau$ ranging between 0 s and 1.5 s with an interval of 0.12 s are shown in Fig 5.11.

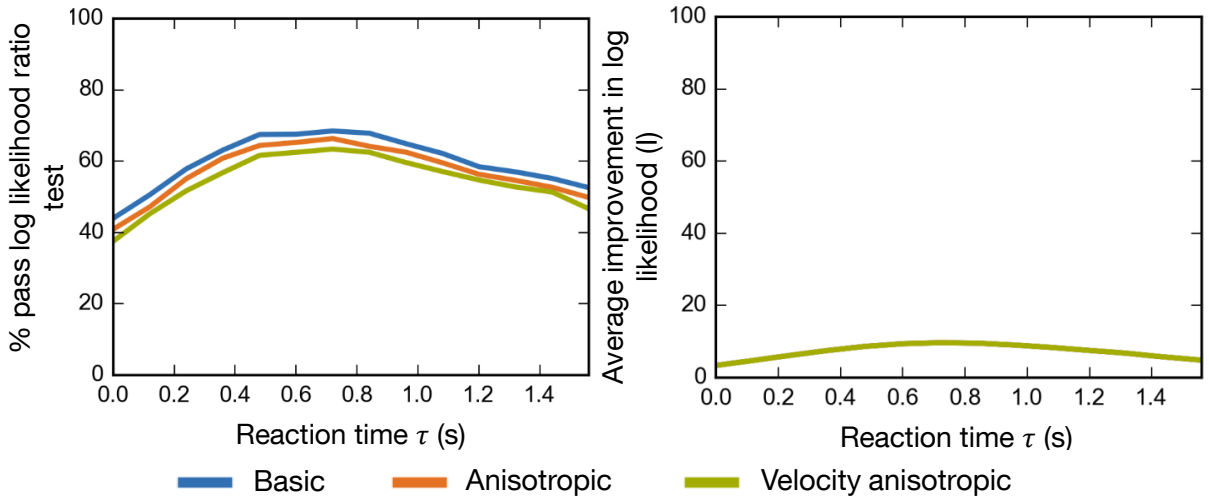



**Fig 5.11** Validation values from the $\Delta\theta_b(t+\tau)$ models for varying reaction times $\tau$

The optimal reaction time $\tau$ lies between 0.4 s and 0.8 s, which is lower than the range identified for the $\Delta V_b(t)$ portion of the model (~1.2 s). Although the perception and processing components of the reaction time are the same for both actions, the different levels of complexity associated with the tasks may account for the difference in overall reaction time. This may indicate that bicyclists can react faster to stimuli by adjusting their direction of travel rather than their speed. A reaction time of $\tau = 0.6\ s$ for change in direction is selected and the calibrated model parameters from this reaction time are presented and evaluated.

The calibrated model parameter distributions are shown in Fig 5.12 (page 126). Based on a qualitative assessment of the distributions, the hypothesis of normal distribution is rejected for the interaction factor $A_{\theta b}$, the radius of interaction $R_{\theta b}$, the anisotropic factor $\eta_{\theta b}$ and the velocity direction factor $\gamma_{\theta b}$. All of these factors show a very small deviation around a centric value. For this reason, the median value of each calibrated parameter is selected as a constant for the entire population of bicyclists. This simplifies the models for the simulation. The means, confidence intervals and standard deviations are nevertheless reported in Tab. 5.13 – Tab. 5.15 if the reader is interested. The relaxation time $T_{\theta b}$ is deemed to be normally distributed and as such a bicycle specific parameter is used.

| N=613 | | Interaction factor $A_{\theta b}$ | Radius of interaction $R_{\theta b}$ (m) | Relaxation time $T_{\theta b}$ |
|---|---|---|---|---|
| Mean [CI] | | 0.48 [0.47, 0.49] | 1.92 [1.91, 1.93] | 1.17 [1.11, 1.23] |
| Median | | 0.50 | 1.99 | 0.99 |
| Std. | | 0.10 | 0.14 | 0.74 |
| Correlation $R_{ab}$ | $A_{\theta b}$ | 1 | 0.36 | -0.39 |
| | $R_{\theta b}$ | - | 1 | -0.33 |
| | $T_{\theta b}$ | - | - | 1 |

**Tab. 5.13** Calibration results for the $\Delta\theta_b(t+\tau)$ basic model (reaction time: $\tau = 0.6\ s$)

| N=613 | | Interaction factor $A_{\theta b}$ | Radius of interaction $R_{\theta b}$ | Relaxation time $T_{\theta b}$ | Anisotropic factor $\eta_{\theta b}$ |
|---|---|---|---|---|---|
| Mean [CI] | | 0.48 [0.47, 0.49] | 1.91 [1.91, 1.92] | 1.15 [1.09, 1.21] | 1.96 [1.95, 1.97] |
| Median | | 0.50 | 1.99 | 0.98 | 1.99 |
| Std. | | 0.10 | 0.14 | 0.72 | 0.10 |
| Correlation $R_{ab}$ | $A_{\theta b}$ | 1 | 0.33 | -0.32 | 0.30 |
| | $R_{\theta b}$ | - | 1 | -0.48 | 0.41 |
| | $T_{\theta b}$ | - | - | 1 | -0.21 |
| | $\eta_b$ | - | - | - | 1 |

**Tab. 5.14** Calibration results for the $\Delta\theta_b(t+\tau)$ anisotropic model (reaction time: $\tau = 0.6\ s$)

| N=613 | | Interaction factor $A_{\theta b}$ | Radius of interaction $R_{\theta b}$ | Relaxation time $T_{\theta b}$ | Anisotropic factor $\eta_{\theta b}$ | Velocity direction factor $\gamma_{\theta b}$ |
|---|---|---|---|---|---|---|
| Mean [CI] | | 0.48 [0.48, 0.49] | 1.92 [1.91, 1.93] | 1.12 [1.06, 1.17] | 1.97 [1.97, 1.98] | 0.97 [0.96, 0.98] |
| Median | | 0.50 | 1.99 | 0.97 | 1.99 | 1.00 |
| Std. | | 0.08 | 0.14 | 0.66 | 0.07 | 0.09 |
| Correlation $R_{ab}$ | $A_{\theta b}$ | 1 | 0.53 | -0.51 | 0.23 | 0.21 |
| | $R_{\theta b}$ | - | 1 | -0.59 | 0.28 | 0.30 |
| | $T_{\theta b}$ | - | - | 1 | -0.30 | -0.42 |
| | $\eta_{\theta b}$ | - | - | - | 1 | 0.72 |
| | $\gamma_{\theta b}$ | - | - | - | - | 1 |

**Tab. 5.15** Calibration results for the $\Delta\theta_b(t+\tau)$ velocity anisotropic model (reaction time: $\tau = 0.6\ s$)

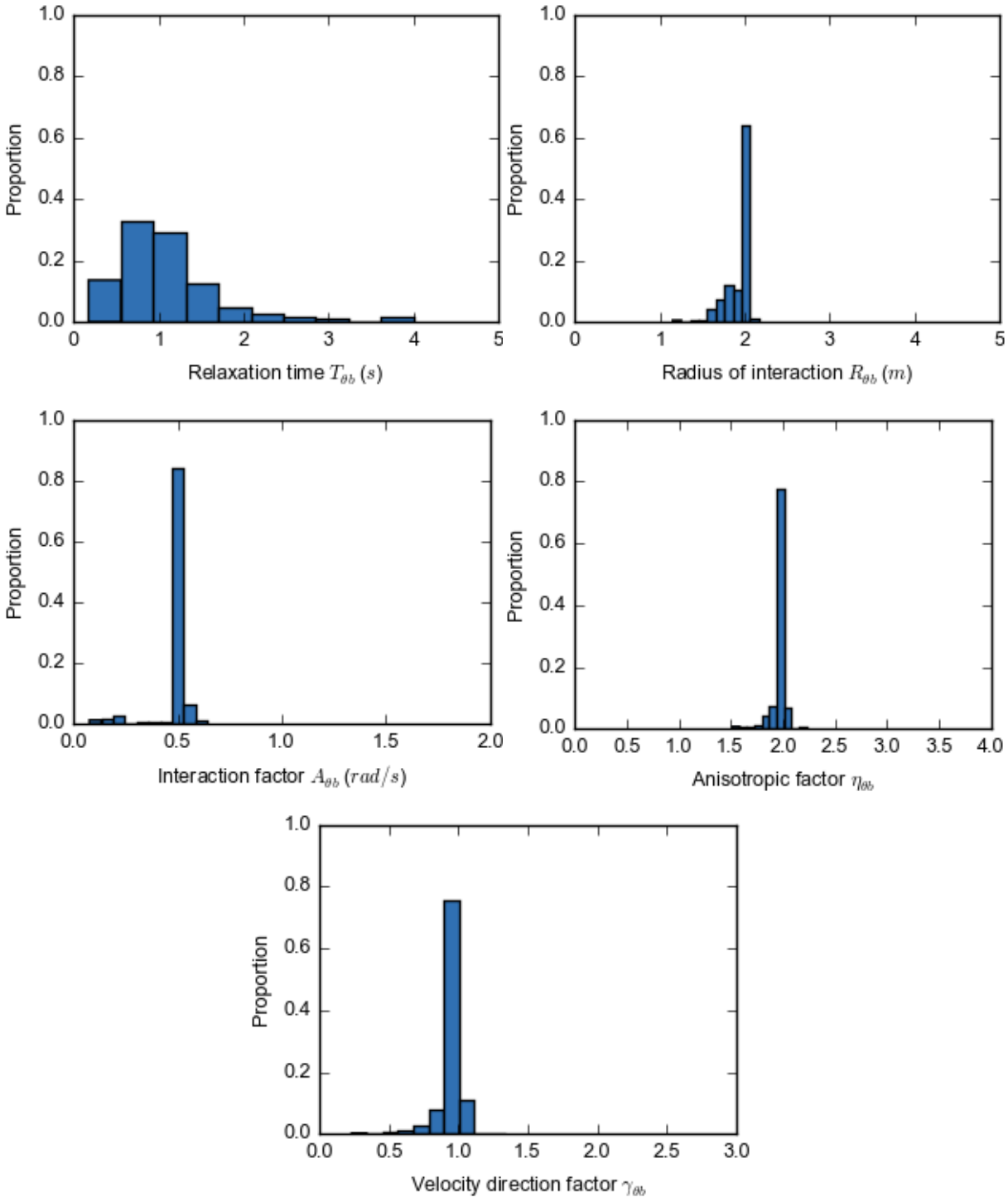



**Fig 5.12** Distributions of calibrated model parameters for the $\Delta\theta_b(t+\tau)$ velocity anisotropic model

The calibrated parameters have the expected magnitude and sign. However, it is difficult to compare the results to the findings of other studies because no studies were identified that examined the response to other road users regarding the change in direction (without change in speed). The mean of the interaction factor $\overline{A_{\theta b}} = 0.48\,rad/s$ (28°/s) signifies the maximum

response to an interacting road user. The calibrated relaxation time for the change in direction models is considerably lower than that in the change in speed portion of the model. This makes logical sense as changes in direction, which tend to be quite small, are realised without significant delay. This value, however, is highly dependent on the desired direction $\theta_b^0(t)$, which as stated before, is difficult to isolate based on observed trajectory data. The anisotropic factor $\eta_{\theta b}$ and the velocity direction factor $\gamma_{\theta b}$ agree very strongly with the calibrated values found for the change in speed portion of the model and appear to be realistic.

The Spearman correlation coefficient $R_{ab}$ between the parameters (assuming normal distribution) are shown in Tab. 5.13 – Tab. 5.15 but are not investigated further as the bicyclist specific parameters $A_{\theta b}$, $R_{\theta b}$, $\eta_{\theta b}$, $\gamma_{\theta b}$ were found to be adequately represented by the population parameters $A_\theta$, $R_\theta$, $\eta_\theta$, $\gamma_\theta$. These population parameters are set as the median value for the observation.

# 6. Implementation and evaluation

In order to evaluate the feasibility of the proposed models as an integrated approach for simulating bicycle traffic, the operational and tactical models presented in Section 5 are implemented in a microscopic traffic simulation environment. Microscopic traffic simulation is a useful tool for the design and evaluation of traffic measures because it recreates the road environment as a complex system in which numerous components interact and influence one another. Although the developed behaviour models have been independently calibrated and validated using K-fold cross validation, the interaction between the proposed models is still to be examined. Furthermore, the interactions between the proposed models and other facets of the simulated road environment, including traffic signals and other road users, are explored in this section.

## 6.1 *SUMO* simulation

*SUMO* is an open source microscopic, multi-modal traffic simulation software that is developed by the Institute of Transportation Research at the German Aerospace Centre (DLR) [KRAJZEWICZ ET AL., 2006]. Development of the tool began in 2000 and the models and software have been continually advanced and improved since then. *SUMO* is mainly a link-based program in which road users progress along lanes of roadway links. The car-following model developed by KRAUß [1998] is used as a default in the simulation, but several other car-following models are directly implemented in the software and can be selected by the user. The default lane changing model was developed by ERDMANN [2015] and is based on a four-layer hierarchy for lane changing decisions. Several improvements in recent years have enabled overtaking within the same lane [SEMRAU ET AL., 2016] as well as travelling in two directions on a lane. These features have improved the potential for realistically simulating pedestrians and bicyclists [KRAJZEWICZ ET AL., 2014].

*SUMO* is used in this dissertation to implement and test the proposed models for the operational and tactical behaviour of bicyclists. The open source code, a well-developed API and the support at DLR made using this software an excellent choice for model development and testing. The operational model is calibrated and validated using data from three of the four research intersections in Section 5.2. The remaining research intersection, Marsstraße and Seidlstraße is simulated in *SUMO* to evaluate the modelling approach (Fig 6.1).

The traffic volumes and turning rates for cars, trucks and bicycles are based on observations from the video data (see Tab. 3.2). Pedestrians are not included in this test of the bicyclist models. All road users are simulated using default models for the given vehicle types in *SUMO*. The sublane feature of *SUMO* is activated, which divides lanes into an adjustable number of narrower sublanes. Simulated road users can flexibly adapt their lateral position

within the lane by selecting the most suitable sublane. The SUMO default model is used for bicycle traffic until they reach the model controlled simulation zone, within which bicyclist movement is controlled using the models developed in this dissertation. Other than the parameters shown in Tab. 6.1, the default *SUMO* model parameters are used. Several additional parameters, including the desired lateral position, the maximum lateral speed and the minimum lateral distance, must specified for each type of road user when using the sublane feature. The intersection is controlled by a fixed time control with a cycle length of 90 s, both in reality and in the simulation.

**Fig 6.1** Simulation of the intersection Marsstraße and Seidlstraße in *SUMO* (background image: Google Earth 2013)

The parameters in the following table are selected based on data measured from the video data or assumptions based on the average characteristics of vehicles in Germany. A distribution of speeds amongst a population of road users is achieved using the speed factor and speed deviation parameters. Setting the speed factor to 1.0 and speed deviation to 0.1 will lead to a situation in which 95% of the road users move at a speed between 80% and 120% of the given maximum speed. No distribution in the other parameters, including acceleration, deceleration and desired headway, are included in the current version of *SUMO*.

| | Bicycles | Passenger cars | Trucks and buses |
|---|---|---|---|
| Maximum speed (m/s) | 5.2 | 13.9 | 11.1 |
| Speed factor | 1.4 | 1.2 | 1.05 |
| Speed deviation (m/s) | 0.1 | 0.1 | 0.1 |
| Acceleration ($m/s^2$) | 1.0 | 1.5 | 0.8 |
| Deceleration ($m/s^2$) | 2.0 | 4.5 | 3.0 |
| Length (m) | 2.0 | 5.0 | 10.0 |
| Desired headway (s) | 1.5 | 1.5 | 1.5 |
| Minimum lateral gap (m) | 0.05 | 0.30 | 0.30 |
| Lateral alignment | compact | center | center |
| Sublane resolution | | 0.4 | |

**Tab. 6.1** Selected parameter values for *SUMO* vehicle behaviour models

## 6.2 Integration of modelling approach with *SUMO*[10]

The models described in Section 5 are to be integrated with the simulation software *SUMO* in order to evaluate feasibility. Model development was carried out using the computer programming language Python. In order to connect these behaviour models written in Python with the simulation software *SUMO*, the interface *TraCI* is used. *TraCI* offers a generic interface for connecting road traffic simulators with a network simulator [WEGENER ET AL., 2008]. The use of this interface in conjunction with *SUMO* enables the extraction and manipulation of many attributes of the simulated road users and road environment. Here, *TraCI* is used to control the position and speed of the simulated bicyclists as well as to extract the position and velocity of the other road users and attributes describing the phase of the traffic signal.

The bicyclists are introduced to the network using the standard *SUMO* method for creating bicycle traffic. A vehicle type “bicycle” is defined with the attributes shown in Tab. 6.1. Each

[10] Research similar to the final work presented in this section was published in the papers:

TWADDLE, H.; GRIGOROPOULOS, G. & BUSCH, F. [2016A]: An approach for simulating bicycle traffic using observed trajectory data. In SUMO Conference 2016 Post Conference Proceedings. Berlin. [TWADDLE ET AL., 2016A]

TWADDLE, H.; GRIGOROPOULOS, G. & BUSCH, F. [2016B]: Integration of an external bicycle model in SUMO. In SUMO 2016 Conference Proceedings. Berlin. [TWADDLE ET AL., 2016B]

simulated bicycle has a different desired speed that is assigned randomly based on the maximum speed, speed factor and speed deviation used to parameterise the vehicle type. The other behaviour parameters are the same for all bicyclists. Bicycle flows are created for each of origin/destination (OD) pair in the simple network based on the observed volume of bicycle traffic at the intersection. Routes are defined to specify the links used by bicyclists belonging to each of the OD pairs.

The proposed modelling approach is applied while the simulated bicyclists are within the model controlled simulation zone, which corresponds roughly to the area shown by the areal image in Fig 6.1. This section is slightly larger than the field of view of the video frame and includes the area in which observed trajectory data are available. In order to simplify the initial integration of the modelling approaches with *SUMO*, only bicyclists who approach and exit the intersection travelling in the correct direction of travel are included in the simulation. Although the modelling approaches are capable of simulating bicyclists travelling against the mandatory direction of travel, *SUMO* is not yet able to route bicycle traffic such that simulated bicyclists travel against the direction of travel on certain segments of their journey through the network. This makes it impossible to test the direction aspect of the tactical behaviour that was examined in Section 5.1 within the framework of a working *SUMO* simulation network. For this reason, this tactical behaviour was not further investigated in this feasibility study.

Within the model controlled simulation zone, the movement of the simulated bicyclists is controlled by the operational model presented in Section 5.2. While travelling through the model controlled simulation zone, the position and speed of the bicyclists are controlled using the *moveToXY* and *setSpeed TraCI* methods in accordance with the output of the operational model. The *keepRoute* option of the *moveToXY* method is set to *2* to allow bicyclists to move outside of the simulated road network. This makes it possible for simulated bicyclists to move on sidewalks and other infrastructure elements that are not explicitly included in the test network. The *SUMO* option *time-to-teleport* is used to remove vehicles from the simulation that cannot complete the route through the network in a predefined time duration. The *SUMO* road users are modelled such that they react to the movement patterns of other *SUMO* simulated road users. However, issues were found to arise when the *moveToXY* function is used to move externally simulated bicyclists to unexpected positions in the road network. For example, bicyclists carrying out an indirect left turn stop within the intersection after completing the first phase of the crossing. Although lanes are not obstructed by the stopped bicyclist, *SUMO* simulated road users moving in parallel lanes sometimes stop and do not pass the waiting bicyclist. Gridlocks are prevented by removing problematic road users that stop for an extended duration within the intersection. Blockages that still occur tend to be brief and do not have a significant impact on the overall simulation. However, this issue should be addressed if the bicycle models were to be integrated permanently into the simulation software.

### 6.2.1 Network alignment

A challenge when integrating an external model with a simulation tool is the alignment of the coordinate systems of two environments. This problem can be eliminated by developing both environments using the same coordinate system. However, in some cases, the externally developed model operates on a uniquely defined coordinate system and adjusting this system to fit the coordinate system of the simulation environment is tedious. Here, the trajectories extracted from the video data use a local coordinate system in which the point *(0 px, 0, px)* is located at the top left corner of the real world image used for video calibration. The movement of the road users, which is originally measured in pixels per frame, is translated to a local coordinate system in meters using a perspective transformation of the coordinates and a measured meter per pixel value. The resulting position and speed data are based on a coordinate system defined by this initial perspective transformation. Unfortunately, it is relatively difficult to construct a road network in *SUMO*-based on this coordinate system.

To address this issue, the same perspective transformation approach that was implemented to translate the video frame to the local coordinate system is used again. Using the method *findHomography* of the open source library *OpenCV*, a perspective transformation matrix $H$ is found between two plans defined by a number of corresponding points. The coordinates of points identified in both the model (original plane) and simulation (target plane) environments, such as start and end points of stop lines at intersections, are input into the method. The resulting perspective transformation matrix $H$ is then used to translate the coordinates of the representative trajectories from each of the identified clusters to those of the simulation environment.

### 6.2.2 Pathway selection

Upon entering the model controlled simulation zone, each bicyclist is assigned a desired pathway based on the planned manoeuvre (straight, right turn or left turn) assigned by *SUMO*. The potential pathways are the exemplars of the trajectory clusters identified in Section 4.1 for each approach of each intersection. Bicyclists turning right or travelling straight across the intersection are assigned a pathway corresponding to the direct option, as shown in Section 4.2. The other pathway options are not possible to carry out without moving against the mandatory direction of travel, which is not tested here. Bicyclists turning left, however, have three possible pathway options that approach and exit the intersection in the expected direction of travel (the fourth pathway option identified in Section 4.2 is excluded due to difficulties simulating road users travelling against the given direction of travel in *SUMO*). The tactical model estimated in Section 5.1 (Tab. 5.6) is used to select the type of pathway across the intersection. The necessary situational data for the choice model, including the number of bicyclists in the approach and the signal phase, are extracted from *SUMO* using *TraCI*.

Relevant characteristics of the intersection, such as the type of bicycle facility and the availability of parking, are stored for each approach.

It is assumed that the ratio of the probability of choosing one option over the probability of choosing another option is independent of the presence of additional alternatives (Independence-from-Irrelevant-Alternatives (IIA) axiom [BEN-AKIVA & LERMAN, 1985]) and the random components of the utility of each option are independent across alternatives. The utility $U_{qi}$ of option $q$ for road user $i$ can then be reduced to the systematic element $V_{qi}$ [HENSHER ET AL., 2005].

$$U_{qi} = V_{qi} = \sum_{k=1}^{K} \beta_{qki} X_{qki} \qquad \text{Eq. 6.1}$$

Where $X_{qki}$ is the value of attribute $k$, $\beta_{qki}$ is the alternative specific weighting parameter for road user $i$ and attribute $k$ and $K$ is the total number of attributes. The probability $P_{qi}$ of road user $i$ choosing alternative $q$ is given using the basic multinomial logit model:

$$P_{qi} = \frac{e^{V_{qi}}}{\sum_{j=1}^{J} e^{V_{qj}}} \qquad \text{Eq. 6.2}$$

The total number of alternatives is denoted by $J$. The probabilities of the potential alternatives for the manoeuvre, which are limited here to three options for bicyclists turning left, are calculated using Eq. 6.2 and ordered based on the magnitude of $P_{qi}$ (highest to lowest). A random number $rn$ between zero and one is drawn and used to assign the preference order for each simulated bicyclist. The alternative corresponding to the random number is appended as the first choice for the bicyclist. The options to the left of the first choice option are appended to the end of the sorted alternative list. An example is shown in Fig 6.2. This is done to ensure that all bicyclists do not continually select the most probable option.

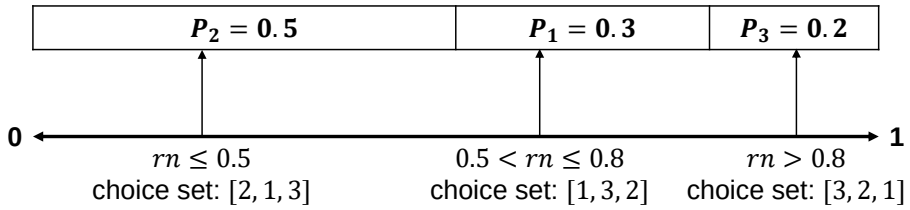



**Fig 6.2** Use of random number to select alternative priorities

The actual pathway to be followed across the intersection is found by augmenting the approach specific pathway into the form of the generic feature vectors. The trained Affinity Propagation matrix is used to categorise each of the approach specific pathways into one of the generic manoeuvre clusters. The generic clusters correspond with the manoeuvre alternatives (i.e. direct and indirect left turns), which are organised in an ordered option set

using the method demonstrated in Fig 6.2. This step builds a crucial bridge between the approach specific pathways followed across the intersection, which vary in form depending on the geometry of the intersection, and the type of manoeuvre used to cross the intersection. The bicyclist is assigned the approach specific pathway corresponding to the selected alternative in the ordered options set. Each approach is characterised by a number of pathways options, which may not represent all of the possible tactical alternatives. This is because not all of the general manoeuvre pathways are observed at all intersection approaches. For example, at one approach specifically, there may not have been any bicyclists observed carrying out a direct left turn. If the first choice is not available at the given approach, the bicyclist is assigned a pathway corresponding to the second alternative in the ordered choice set, and so forth. Conversely, each tactical manoeuvre may be represented by a number of approach specific clusters. If this is the case, one of the clusters is selected randomly for the simulated bicyclist. Finally, certain manoeuvre alternatives can be exempt from the choice set using flags that can be set by the practitioner as necessary.

Once selected, the pathways are extended to the entry and exit points of the model controlled simulation zone. The entry point is the position where the simulated bicyclist first touches the model controlled simulation zone. The exit point is defined by the intersection between the boundary of the model controlled simulation zone and a line representation of the *SUMO* edge to be used after crossing the intersection. The desired pathway is extended to the entry and exit points of the model controlled simulation zone using splines such that a complete pathway is generated through the entire model controlled simulation zone. The resulting pathways are shown in Fig 6.3 below for the test intersection. The actions taken by the simulated bicyclist to follow this pathway while responding to other road users and the traffic signal are determined using the operational model described in Section 5.2.

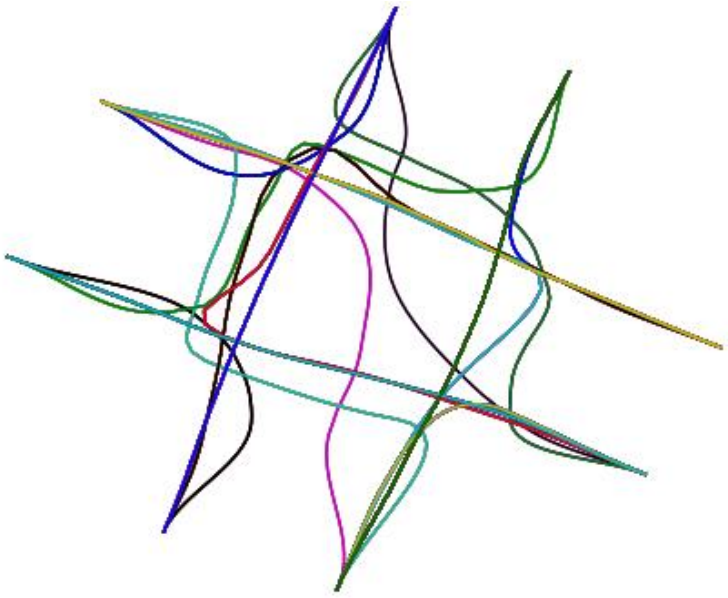


**Fig 6.3** Example of the desired pathways for bicyclists simulated at the example intersection

### 6.2.3 Operational model implementation

The operational model specified, calibrated and validated in Section 5.2 is implemented and tested within the model controlled simulation zone. All three model variations, the basic, anisotropic and velocity anisotropic models, were implemented in the simulation and the results were subjectively assessed. Although the most complex model, the velocity anisotropic model, was not found to perform significantly better in the model validation (Section 5.2.3) than the other two models, it was found to improve the behaviour of bicyclists in the simulation. The varied reaction to other road users based on the relative position and velocity of both interacting road users was found to recreate behaviour that is much more realistic. Particularly following situations in which bicyclist $b$ moved behind another road user travelling in nearly the same direction at nearly the same speed were simulated more realistically using the velocity anisotropic model.

The parameters of the velocity anisotropic model are calibrated using observed trajectory data. A number of the parameters are found to be normally distributed amongst the population of bicyclists. In the simulation, the calibrated distribution of these parameters, including the covariance between the parameters, is used to assign each simulated bicyclist with a realistic combination of attributes. Maximum and minimum values for the parameters assigned using distributions are defined to prevent the occurrence of unrealistic parameters and are given in Tab. 6.2. As discussed in Section 5.2.4, the bicycle specific parameters, $A_{\theta b}$, $R_{\theta b}$, $\eta_{\theta b}$ and $\gamma_{\theta b}$, are substituted with population parameters, $A_\theta$, $R_\theta$, $\eta_\theta$ and $\gamma_\theta$, that are assigned to all bicyclists (median values from the calibration results).

| Parameter | | Mean ($\mu$) | Std. ($\sigma$) | Minimum | Maximum |
|---|---|---|---|---|---|
| Desired speed (m/s) | $V_b^0$ | 5.2 | 1.4 | 3.0 | 10.0 |
| Relaxation time - speed (s) | $T_{vb}$ | 3.8 | 2.1 | 3.0 | 5.0 |
| Radius of interaction (m) | $R_{vb}$ | 3.1 | 0.8 | 1.0 | 5.0 |
| Anisotropic factor | $\eta_{vb}$ | 2.0 | 0.3 | 1.0 | 3.0 |
| Velocity direction factor | $\gamma_{vb}$ | 1.0 | 0.3 | 0.5 | 1.5 |
| Relaxation time – direction (s) | $T_{\theta b}$ | 1.1 | 0.1 | 1.0 | 1.2 |

**Tab. 6.2** Mean, standard deviation, minimum and maximum values for operational model parameter distributions estimated in Section 5.2

Additional parameters are applied to the simulation to ensure realistic bicyclist behaviour and simulation feasibility. The first of which is a designated minimum lateral and longitudinal safety distance. Here, the longitudinal safety distance is modelled using a normal distribution ($\mu = 0.5\,m,\ \sigma = 0.1\,m$) with a maximum and minimum safety distance of $0.7\,m$ and $0.3\,m$, respectively. The lateral safety distance is modelled in the same way with a mean of $\mu = 0.4\,m$, a standard deviation of $\sigma = 0.1\,m$ and a minimum and maximum safety distance of $0.2\,m$ and $0.6\,m$, respectively. These parameters have not yet been calibrated, as it was not possible to

extract this information from the trajectory data. However, once calibrated, these parameters could prove instrumental in simulating bicyclist impatience.

Because $\Delta V_b(t)$ and $\Delta\theta_b(t)$ are calculated separately, there is no link between the change in speed and change in angle in each time step. This was found to lead to unrealistic situations in which bicyclists moving very slowly continue to change direction at the same rate as bicyclists travelling quickly. In order to address this issue, the angle change per simulation time step is controlled based on the current speed of the bicyclist. If the speed drops below 0.5 m/s, the change in angle $\Delta\theta_b(t)$ is multiplied by the speed. This ensures that bicyclists that are not moving, or are moving slowly, do not change direction in an unrealistic manner. The total change of angle in response to all interacting road users is limited to $A_\theta$, which prevents simulated bicyclists from changing direction drastically in one time step. These stipulations were not included in the original model in Section 5.2 and were not calibrated or validated using the trajectory data. This extra control may not be necessary, considering that it is possible for bicyclists to pivot (slowly) without moving forward. In any case, model testing should be carried out to determine if it is necessary to link speed with change in angle and to calibrate and validate of these stipulations if necessary in future work.

The shape of bicyclists and all other road users is recreated in the simulation using a simple rectangle. Simulation of bicyclists using a diamond shape, as proposed by FALKENBERG ET AL. [2003] and implemented in PTV Vissim, was tested but no advantages were found. The diamond shape allowed bicyclists to queue in a more compactly, although this was not found to be more realistic than the rectangle shape using the calibrated parameters.

### 6.2.4 Signal control

The reaction of bicyclists to the signal control proved to be difficult to simulate within the *SUMO* environment. Road users simulated in *SUMO* travel along the network edges until they reach a node (intersection) which is defined by a polygon. The stop line of the approach is the intersection between the edge and the intersection polygon. Once road users receive a green signal and are able to cross the stop line, they enter the intersection polygon and complete their manoeuvre. Here, the polygon representations of the intersection (and signal control) in *SUMO* are extracted and forwarded to the operational behaviour models. The operational reaction of the simulated bicyclists to the signal, including decelerating towards the stop line at a red signal head, is modelled using the same approach as the reaction to other road users. When the signal is red for a bicyclist on a given approach, the polygon of the intersection is activated, forcing the bicyclists to react and stop at the stop line. The polygon is then deactivated when the signal becomes green. Once a bicyclist has entered the intersection polygon, it can no longer react to the presence of the intersection polygon when it is activated. This is not a problem for bicyclists that only react to one signal during the intersection

crossing. However, the reaction to the signal of bicyclists requiring two signal phases to cross the intersection, such as those carrying out indirect left turns, is difficult to simulate.

To address this issue, an approach is proposed to allow for the flexible reaction of simulated bicyclists to different signal heads along their intersection crossing. Each edge approaching the intersection is characterised by angle, shown by the vectors $\vec{N}$, $\vec{E}$, $\vec{S}$ and $\vec{W}$ in Fig 6.4, and a principal signal head. The phase of the signal head associated with the approaching edge currently used by the simulated bicyclist is used to activate or deactivate the intersection polygon. Bicyclists are assigned an edge upon entering the model controlled simulation zone based on the route (e.g. *NE)*. In each simulation step, the difference between the direction of travel of the bicyclist $\theta_b(t)$ and the angle of all approaching edges is calculated. The bicyclist is assigned the edge with the smallest difference. A factor of 1.1 is applied to the angle differences of all approaches not currently used by the bicyclist. This buffer prevents simulated bicyclists from oscillating between signal heads.

The shape of the polygon is modified to allow bicyclists to react to a second signal head once they are already within the intersection polygon (Fig 6.4). This is particularly important for bicyclists performing indirect left turns. The polygon is modified in the time step in which a bicyclist first switches signal heads (approach edge). This is done using a direction change polygon (B in Fig 6.4), which is defined by taking the current position of the bicyclist $p_b(t)$ as a starting point for the vector $\vec{D}$ with an angle equal to that of the newly assigned edge. The length of the vector is inconsequential. The polygon is defined by offsetting vector $\vec{D}$ 5 m to the right hand side and 5 m to the left hand side. Polygon B is defined by the start and end points of the two offset vectors. The intersection of the original intersection polygon and the direction change polygon is the modified intersection polygon. This polygon is used to represent the signal control for the given bicyclist for the remainder of the intersection crossing or until there is another change in the approaching edge (signal head). An example of a signal head change is shown in Fig 6.4.

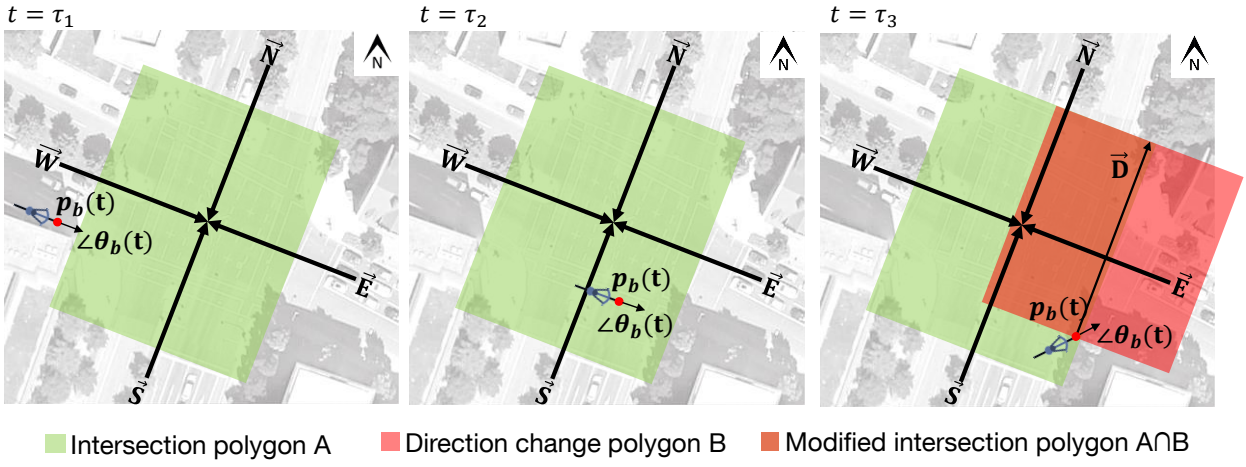



**Fig 6.4** Example of the derivation of the modified intersection polygon during an indirect left turn at the test intersection

In this example, a bicyclist approaches the intersection from the west and carries out an indirect left turn. The signal is green upon approaching the stop line and the bicyclist proceeds across the intersection. When the bicyclist turns to carry out the second portion of the left turn manoeuvre, the south signal head is activated and a new intersection polygon is created.

## 6.3 Simulation evaluation

The main goal of this work is to develop an approach for simulating the flexible behaviour of bicyclists at signalised intersections. Bicyclists, unlike motor vehicles or pedestrians, are able to use different parts of the infrastructure (bicycle facility, roadway and sidewalk) to tactically select their pathway across an intersection. Typical measures used to assess the realism of traffic simulations, such as travel time, speed distribution and traffic flow, do not enable the assessment of flexible pathway selection. In order to compare patterns of movement of bicyclists in reality and the traffic simulation, the area encompassing the intersection is discretized into an array of cells (1.5 m by 1.5 m). Two measures are calculated for each cell in the array, the occupancy, which measures the proportion of time a cell is occupied by at least one bicyclist, and the average speed of bicyclists while in the cell. The results are compared using heat maps, which exhibit the relative occupancy and average speed in each of the cells. Three cases are compared, observed trajectories from the test intersection, the trajectories of bicyclists simulated using the default *SUMO* simulation approach and bicyclists simulated using the proposed operational and tactical behaviour models. Observations for both occupancy and speed are taken every one second along the trajectory of the observed and simulated bicyclists.

In order to account for the variation in simulation runs, several runs for both of the two simulation cases are carried out. The simulation is stochastic due to the behaviour parameter distributions both in the *SUMO* models and in the proposed operational behaviour model. The arrival of the road users, their assigned desired routes across the intersection and the selection of manoeuvres based on the proposed tactical model creates further stochasticity. A period of 60 minutes is assessed for both of the simulation cases. The network is allowed 200 s to fill before the 60-minute simulation time is started. The number of simulation runs is determined based on a predetermined confidence interval, which is given by the equation below [DOWLING ET AL., 2004].

$$CI_{1-\alpha\%} = 2t_{\left(1-\frac{\alpha}{2}\right),N-1}\frac{s}{\sqrt{N}} \qquad \text{Eq. 6.3}$$

Where $t_{\left(1-\frac{\alpha}{2}\right),N-1}$ is Student's t-statistic, $s$ is the standard deviation of the results (variables occupancy and average speed) and $N$ is the number of runs. Due to the exploratory nature of this simulation experiment, a relatively high ratio of the confidence interval divided by the standard deviation of 2.0 is selected. According to DOWLING ET AL. [2004], this ratio with a

confidence of 95% can be reached with a minimum of eight simulation runs. To ensure statistically significant results, ten simulation runs are carried out here. The actual confidence interval from the ten runs is calculated. The ten simulation runs are accepted if $CI < 2\sigma$. Otherwise, two additional runs are carried out and the confidence interval is recalculated. This is repeated until an acceptable confidence interval is reached.

The developed modelling approach is compared with observed data from the intersection Marsstraße and Seidlstraße. The performance of the integrated modelling approaches proposed in this dissertation and the currently available method for simulating bicycle traffic in *SUMO* are compared. This comparison is slightly biased because the simulation of bicycle traffic can be improved to some extent with simulation tricks, such as creating a multitude of links representing the potential pathways for bicyclists. However, implementing simulation tricks can be tedious and makes the simulation prone to errors (e.g. in the definition of conflict points). Moreover, the intention of this work is to create a method for explicitly modelling flexible behaviour. Therefore, a comparison with the default *SUMO* bicycle simulation approach is deemed appropriate. The performance of the developed model in comparison to the default approach for simulating bicyclist traffic in *SUMO* is quantitatively assessed using the root mean square error ($RMSE$).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - x)^2}{n}} \qquad \text{Eq. 6.4}$$

Where $x_i$ is the observed occupancy or average speed and $x$ is the simulated value in each cell of the array. The average speed is set to $0.0$ in array cells where no bicyclists are observed or simulated. This makes it possible to compare all cells in the array but makes the $RMSE$ measure sensitive to the presence of bicyclists in the cell throughout the observation or simulation period. The mean and standard deviation of the $RMSE$ values for the ten simulation runs of the two simulation approaches are shown in Tab. 6.3. The percent improvement of the developed approach in comparison to the *SUMO* approach is shown in the right column.

| | | ***SUMO* default approach** | **Proposed approach** | **Percent improvement** |
|---|---|---|---|---|
| Occupancy | Mean $RMSE$ (std.) | 0.361 (0.001) | 0.318 (0.004) | 12.7 |
| | CI | [0.360, 0.362] | [0.315, 0.321] | |
| Average speed (m/s) | Mean $RMSE$ (std.) | 1.421 (0.005) | 1.341 (0.016) | 5.6 |
| | CI | [1.417, 1.425] | [1.329, 1.353] | |

**Tab. 6.3** $RMSE$ for the *SUMO* default and proposed simulation approach compared to observed trajectories

In addition to this quantitative assessment, the simulation approach is further evaluated based on a qualitative assessment of the heat maps shown in Fig 6.5. Here the occupancy (left column), which is measured as the percentage of time which the cell is occupied by at least one bicyclist, and the average speed in each cell (right column) are shown.

As expected, bicyclists observed in reality (top row in Fig 6.5) cover a large portion of the intersection. Not only the bicycle facilities but also sections of the sidewalks, crosswalks and roadway are used by bicyclists as they cross the intersection. The highest average speeds are reached by bicyclists using the bicycle facility while slower average speeds are observed at the stop lines and on the sidewalks and crosswalks. The cells with the highest occupancy rate are those directly in front of the stop lines of the approaches with the highest volume of bicycle traffic (north and south).

The bicyclists simulated using the default *SUMO* model move in a more consistent manner with very little variation in the speed or pathway used, as shown in the bottom row of Fig 6.5. This consistency in the realised trajectories is expected as simulated bicyclists are limited to predefined edges and pathways through the intersection. However, the average speed of the bicyclists in each of the cells is also much more uniform across space than the average speed observed in reality. Decreases in speed are simulated only at the stop lines and points of intersection/interaction with other flows of traffic. Because bicyclists are limited to predefined pathways, lower speeds when using the crosswalks or sidewalks were not possible to simulate.

The proposed approach shows a higher potential for simulating the dispersed movement of bicyclists as they cross the intersection. In comparison to the *SUMO* default approach, a larger portion of the cells is occupied by at least one bicyclist during the simulation. The distribution of the desired speed parameter $V_b^0$ calibrated using the trajectory data resulted in lower overall average speeds in the cells than is observed in reality. However, the speeds in the cells show more variation in space than those in the default *SUMO* simulation.

Perhaps the most important advantage of the proposed simulation approach in comparison to the *SUMO* approach is the ability to simulate indirect left turns, both with and against the expected direction of travel. Currently, it is only possible to include direct left turns of bicyclists in the *SUMO* simulation [DLR, 2016]. If a simulated bicyclist turning left uses a bicycle facility upon approaching the intersection, the bicyclist unrealistically crosses in front of road users travelling straight on the adjacent roadway lanes. The resulting left turn pathways can be seen in the *SUMO* default simulation heat maps in Fig 6.5. This behaviour causes improbable situations to arise at the intersection, the outcome of which is often congestion that is not observed in reality.

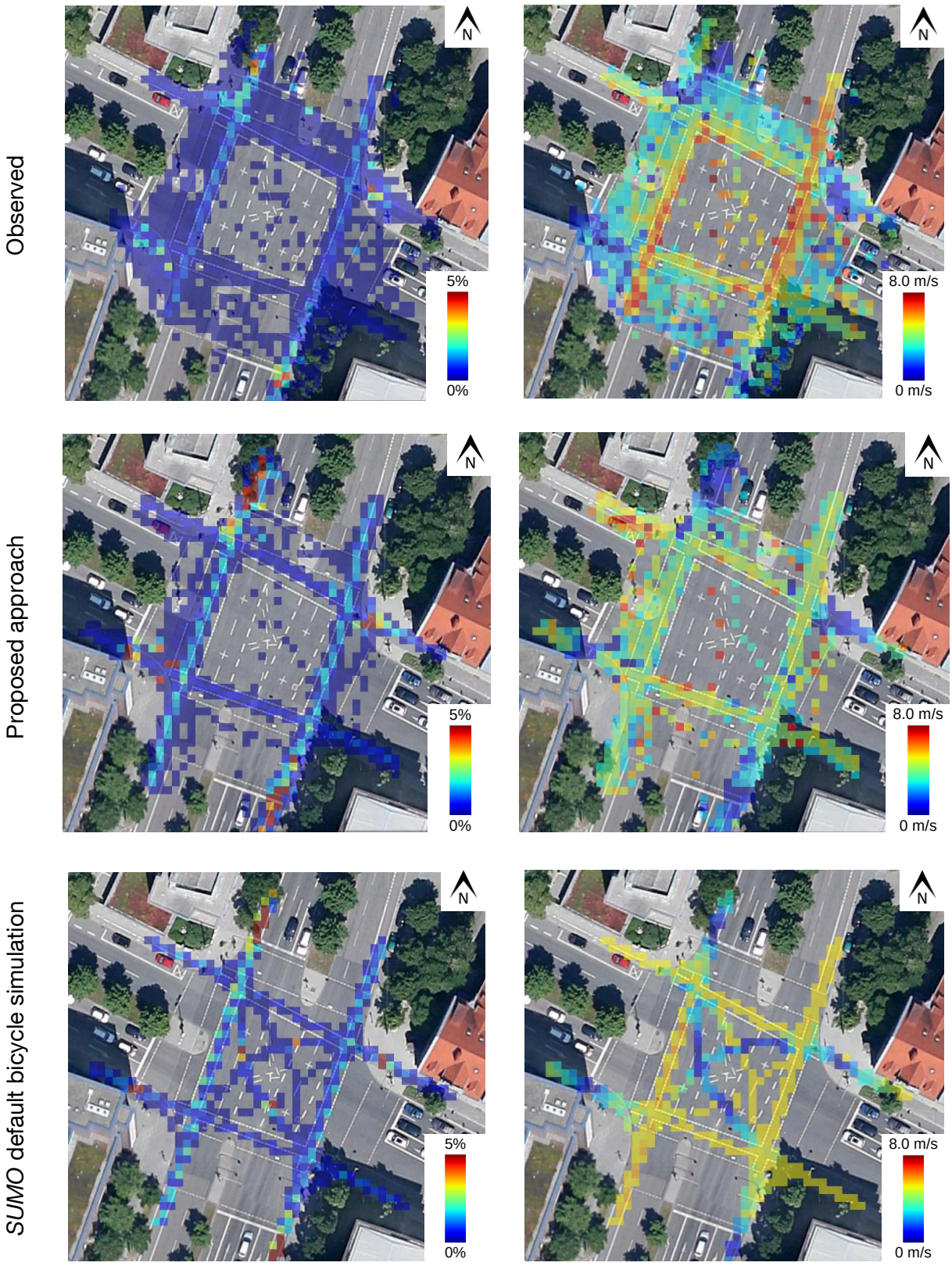



**Fig 6.5** Heat maps of occupancy (left column) and speed (right column) for all manoeuvres

Using the approach outlined in Fig 6.4, the relevant signal head is adapted depending on the orientation of the simulated bicyclist and not only on the route assigned by *SUMO*. This allows bicyclists carrying out left turn manoeuvres, particularly indirect left turns, to respond to the correct signal head at any given moment along their trajectory without this information being explicitly supplied to the simulation. Moreover, the dynamic definition of stop lines in accordance with the orientation and position of the bicyclist within the intersection makes it possible to realistically simulate the waiting position of bicyclists making left-hand turns over two signal phases. Examples of the progression of two simulated bicyclists, one carrying out an indirect left turn in the expected direction of travel (Fig 6.6) and the other an indirect left turn against the given direction of travel (Fig 6.7) are displayed below. The red circle locates the bicyclist throughout the manoeuvre.

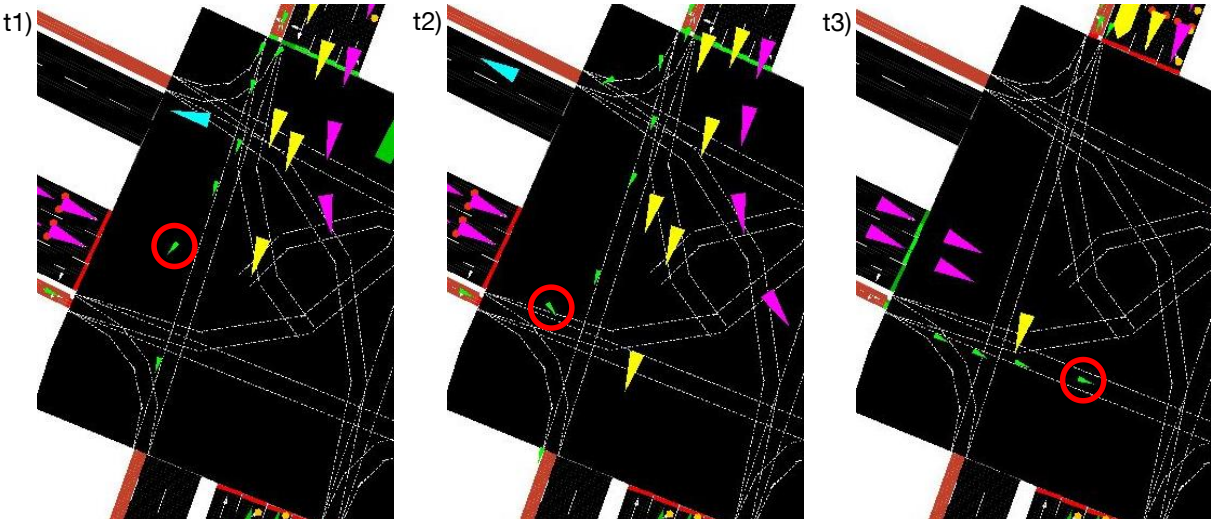



**Fig 6.6** Indirect left turn manoeuvre with the expected direction of travel

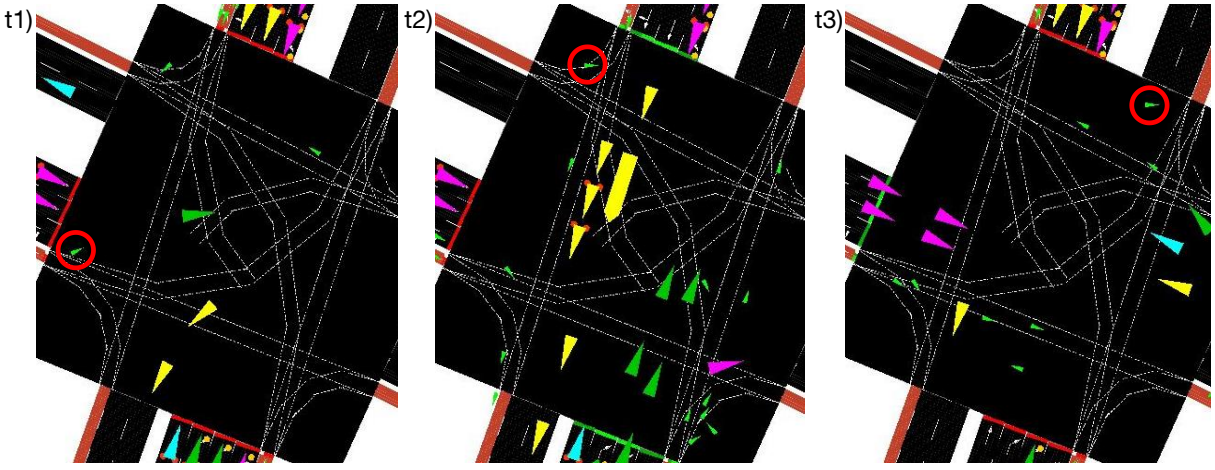



**Fig 6.7** Indirect left turn against the expected direction of travel

In addition to simulating realistic left turning behaviour of bicyclists, the proposed approach also enables the recreation of non-uniform queues at red signals. The default approach for simulating bicycle traffic in *SUMO* includes sublanes, which makes it possible for road users

to move beside one another in the same lane. Nevertheless, the queues formed by bicyclists at the stop line are undeviating and the resulting length of the queue varies linearly with the number of bicyclists in the queue. In reality, the form of bicycle queues varies greatly depending on the behaviour of the particular bicyclists in the queue. To examine the realism of the queues resulting from the proposed approach, the volume of bicycle traffic is doubled at the test intersection. Four examples of queues are shown in the figure below. The first three result from the proposed simulation approach while the last is taken from the *SUMO* default approach for simulating bicycle traffic with sublanes.

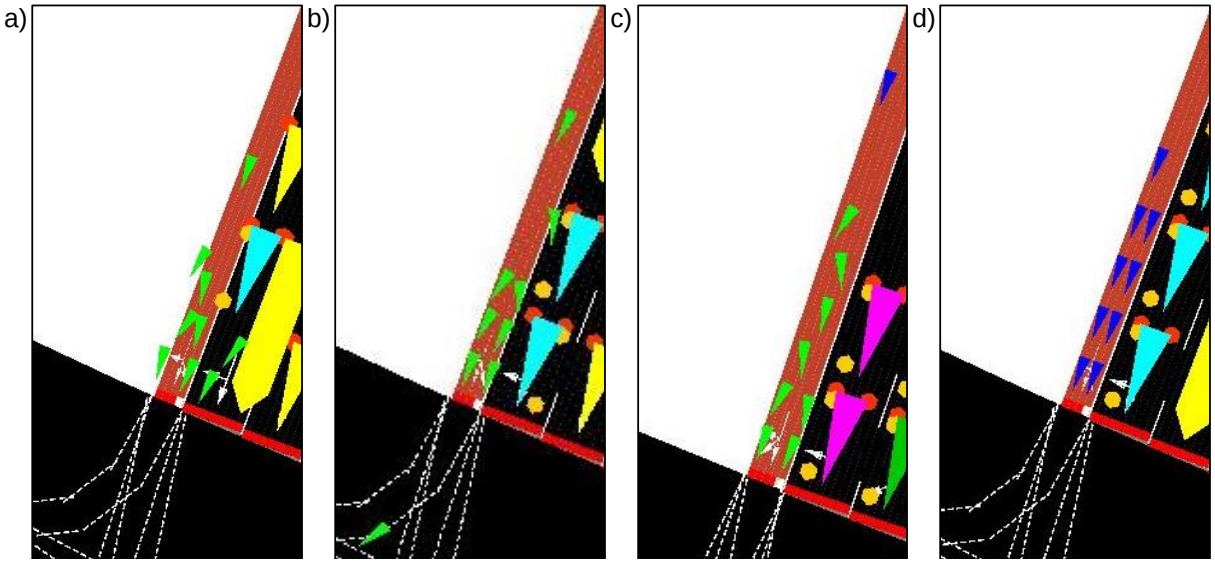



**Fig 6.8** Queues from the proposed approach (a, b and c) and the default *SUMO* approach (d)

Queues observed at the Marsstraße and Seidlstraße intersection are shown in Fig 6.9. The observed variation in the queue length and density, as well as the position of the bicyclists in the queue, varied enormously between signal cycles. The developed approach for modelling the operational behaviour is found to better reproduce this randomness than the current *SUMO* approach.

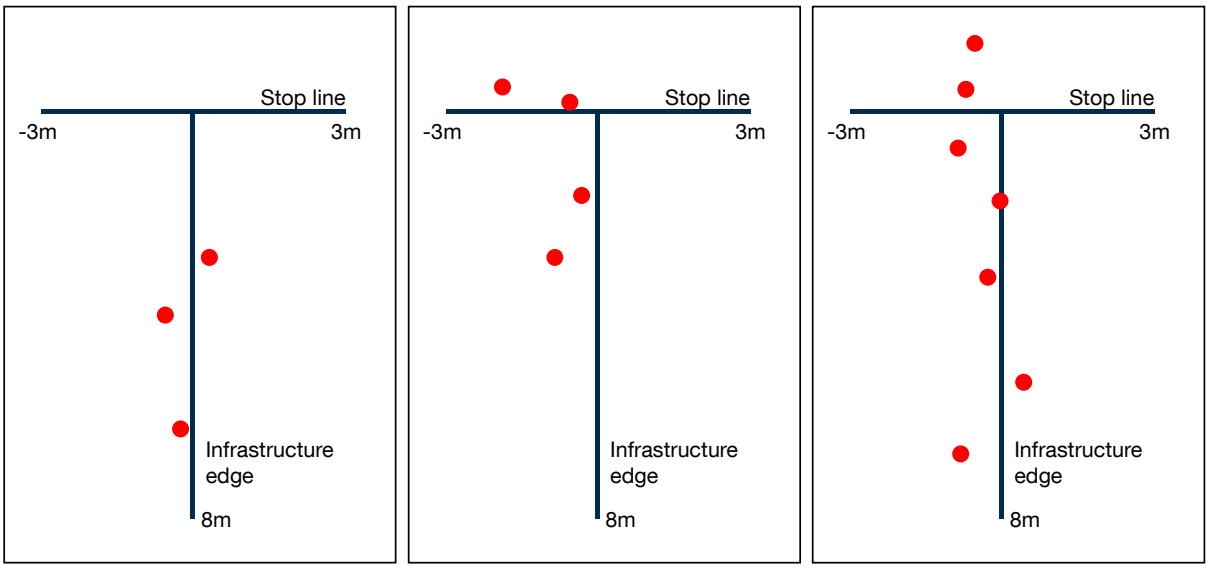



**Fig 6.9** Examples of observed queues at Marsstraße / Seidlstraße (adapted from TAHEDL [2014])

The preference of the bicyclists between the three type of left turn manoeuvre is modelled in the simulation using the simplified logistic regression model introduced in Section 5.1 (Tab. 5.7). An element of randomness is introduced using the method demonstrated in Fig 6.2. The mean, standard deviation (std.) and resulting confidence interval (CI) are shown for the percentage of each type of left turn manoeuvre assigned in the simulation.

| | **Simulated values** | | | **Observed** |
|---|---|---|---|---|
| | Mean (%) | Std. (%) | CI | |
| Direct left turn | 6.6 | 2.4 | [ 3.8, 9.3 ] | 18.8% |
| Indirect left turn | 63.5 | 5.7 | [ 57.0, 70.0 ] | 53.1% |
| Indirect left turn against the mandatory direction of travel | 30.0 | 4.2 | [ 25.1, 34.8 ] | 28.1% |

**Tab. 6.4** Simulated and observed types of left turns

The observed percentage of bicyclists carrying out the three types of left-turn manoeuvres is extracted from the generic clustering resulting from the Marsstraße and Seidlstraße intersection. Just over half of the bicyclists turning left at this intersection carried out an indirect left turn in the expected direction of travel. This falls slightly below the mean percentage of bicyclists carrying out this type of manoeuvre in the simulation. This discrepancy is accounted for in the direct left turns, with fewer bicyclists carrying out direct left turns in the simulation environment than in reality. The proportion of bicyclists carrying out an indirect left turn against the expected direction of travel is very similar in both reality and the simulation environment. To improve the proportion of direct left turns and indirect left turns in the expected direction of travel, it may be worthwhile to implement the full model for the left turn manoeuvre in Appendix 1 in the simulation.

# 7. Conclusions

The research and contributions in this dissertation include the automated extraction of trajectories from video data, the processing of this data, operational and tactical behaviour model specification, calibration and validation as well as the implementation of the developed models within the existing simulation software *SUMO*. A summary of these contributions is given in Section 7.1, followed by the limitation of the work presented in Section 7.2. Finally, an outlook for future work based on the findings of this dissertation is given in Section 7.3.

## 7.1 Summary

In this dissertation, approaches for modelling the behaviour of bicyclists for application in microscopic traffic simulation are presented. High-resolution trajectory data is used to specify and calibrate models of bicyclist behaviour. Although behaviour modelling is the overarching goal of the work, the data collection and analysis methods applied to reach this goal are further contributions to the state-of-the-art. Historically, traffic analysis and evaluation have been based on point observations (e.g. inductive loops), which yield traffic parameters at a given cross section (e.g. traffic volume, average speed, etc.). Alternatively, road user recognition techniques, such as Bluetooth and Automatic Number Plate Recognition (ANPR), provide the travel times between two data collection points (e.g. on a road segment). Trajectory data, in contrast, provides a rich description of the spatial progression of road users and their interactions with the environment and other road users. Previously, the ability of researchers to collect trajectory data has been limited by the extremely high time effort required to create trajectories manually. Recent advances in computer vision have made automated trajectory extraction possible. However, methods for analysing and using trajectory data to specify and calibrate models have remained rudimentary.

The work in this dissertation adds to the budding field of behavioural analysis and model development using trajectory data. Automated methods for extracting trajectory data from video data and processing the trajectories are extended and developed. A method for classifying road users as bicyclists, pedestrians or motor vehicles based on their positions throughout the crossing manoeuvre relative to the road geometry and their maximum speed is developed and implemented. An approach for correcting the distortion in trajectories resulting from a wide-angle lens by applying available image processing tools is presented here, which is found to rectify the trajectories adequately.

Two methods for clustering the extracted trajectory data to identify types of pathways used by bicyclists to cross an intersection are developed and evaluated; a generic pathway clustering method and an approach specific clustering method. The generic clustering method classifies the type of pathways used by bicyclists to travel straight across or carry

out a left or right-hand turn at any intersection with four approaches. This approach can be applied directly to classify trajectories collected at intersections not examined within this dissertation, eliminating the need for time-consuming manual classification in future work. In contrast, the approach specific clustering method developed in this dissertation enables the identification of representative pathways used by bicyclists at a particular approach of a particular intersection. These pathways can be imported directly into a microscopic traffic environment to provide the desired pathways for simulated bicyclists across the intersection.

Behaviour models describing the tactical and operational behaviour of bicyclists are specified in this dissertation. The clustered trajectory data is used to calibrate and validate these models. Binomial and multinomial logistic regression models are estimated for four tactical choices using a dataset of observed choice outcomes; the reaction to a red traffic signal, the choice between using a bicycling facility, the roadway or the sidewalk, the direction of travel and the selection of a pathway across an intersection. There is considerable variation between the predictive power of the four regression models. The models estimated to predict the reaction to a red signal and the type of left turn manoeuvre are capable of predicting each of the choice categories with exceptional accuracy. In contrast, the infrastructure selection models and the direction of travel model have difficulty predicting seldom occurring choice outcomes. A possible explanation for this variation could be that the four tactical behaviours are motivated by different types of factors. The tactical behaviours that are highly explainable using the developed regression models are likely motivated by situational factors that are externally observable. The difficulty in predicting at least one of the outcomes for the remaining behaviours, the direction of travel and infrastructure selection, suggests that the motivating factors for these choices are intrinsic or cannot be observed in the restricted observation area.

The predictive power of the estimated models, as well as the variables that were found to be significant in the logistic regression models, are useful in determining which rule breaking behaviours can be addressed through infrastructure design and traffic signal control measures and which are rooted in non-observable factors. In this dissertation, all the variables that were investigated are externally observable, meaning that they are related to externally controllable factors (infrastructure and signal design). The high predictive power of the models for the reaction to a red signal and the type of left turn manoeuvre suggests that these behaviours can be modified by altering the situation at the intersection. Riding against the given direction of travel and riding on the sidewalk, on the other hand, may be more responsive to softer measures such as safety campaigns and traffic rule enforcement.

An operational behaviour model is specified to reflect the unique characteristics of bicyclists. The NOMAD model, which is a simple force model for pedestrian movement, is adapted to reflect the restrained dynamics of bicycling. This is accomplished by using the norm and angle representation of the velocity and acceleration vectors. Two equations, one for the change in

speed and one for the change in angle, are specified and calibrated using trajectory data. This model allows for the direct control of the maximum change in speed and change in direction at each time step. Furthermore, a delay $\tau$ is integrated in the model that represents the reaction time of the observed bicyclists. Two reaction times are estimated, one for each of the models. The reaction time is estimated for the population of bicyclists by maximising the number of observations passing the log likelihood ratio test over the entire sample and are found to be $\tau = 0.6\,s$ for change in direction and $\tau = 1.2\,s$ for change in speed. The total reaction time is the sum of the time taken to perceive a situation, develop a strategy to cope with the situation and execute this strategy. The duration of time used to execute a strategy is dependent upon the complexity of the task. Here, it is argued that the tasks 'change direction' and 'change speed' have different complexity levels and therefore lead to different reaction times.

In order to evaluate the modelling approaches, the operational and tactical behaviour models are integrated with the microscopic traffic simulation software *SUMO*, which is an open source system developed by the German Aerospace Center (DLR). The position and velocity of the simulated bicyclists are specified using the models presented in this dissertation and are communicated with *SUMO* using functions of the interface *TraCI*. This method was found to be extremely useful for development and evaluation.

Within the simulation environment, the developed modelling approach is found to offer a more realistic simulation of bicycle traffic compared to the model that is available in *SUMO*. The bicyclists are able to move and respond to other road users in a way that better reflects the operational behaviour observed in reality. The flexible pathway selection across the intersection, which results in a wide dispersion of bicyclists, is simulated with success. Although it is possible to generate the dispersed movement in other simulation tools, this is usually done by specifying a multitude of routes and interaction points at an intersection. Using this approach, the flexible behaviour is implicitly modelled. This results in a more realistic simulation of bicyclists, in particular when performing an indirect left turn. The simulated queueing behaviour is also found to reflect the non-uniform shape of bicyclist queues realistically in a way that is not possible with many other simulation software.

There are a number of practical applications for the presented approach for modelling the tactical and operational behaviour of bicyclists in microscopic traffic simulation. For example, signal control measures for bicyclists, including the integration of bicyclists in actuated or adaptive signal control algorithms and the prioritisation of bicyclists at signalised intersections, can be developed more precisely using the methods proposed in this dissertation. The predicted progression (positions and speed) can be taken into account when planning the coordination of signals. Predictions about the tactical behaviour of bicyclists developed in this dissertation, including red light violations and travelling direction, can provide further inputs for signal control algorithms in the future.

Another example is advanced driver assistance systems (ADAS) that are developed and implemented today to alert drivers to the presence and position of bicyclists and intervene by breaking or swerving if a collision is imminent. Microscopic traffic simulation is often used for the initial development and testing of such systems. Clearly, in order to create a realistic simulation environment and extract correct results, the models of bicyclist behaviour in the simulation must be realistic and precise. The operational and tactical models developed in this dissertation and the method for integrating the models with the microscopic simulation *SUMO* offer a means to develop and test ADAS systems. Particularly the highly accurate prediction of the position and speed of bicyclists as they cross the intersection is necessary to evaluate the performance of ADAS. This prediction is greatly improved with the models and methods proposed in this dissertation.

If one looks further into the future, automated vehicles that operate in complex urban environments will require extremely accurate and realistic models of the behaviour of other road users. The behaviour of bicyclists and pedestrians is particularly important to understand and predict because these vulnerable road users may not be equipped with the means to communicate their positions and planned movements electronically, for example with Cooperative Awareness Messages (CAM), as proposed with connected vehicles. If automated vehicles are to adopt strategies that guarantee the safety of vulnerable road users, models that accurately predict the tactical decisions of bicyclists, as well as their movement in the next instant, will be of vital importance. Models that combine situational parameters, personal characteristics of the bicyclist, the type, position and velocity of surrounding road users, the state of the traffic signal and the previous actions of the bicyclist to predict behaviour must be developed. The behaviour models presented in this dissertation offer the first step in this direction. However, it will be necessary to collect significantly more data from a multitude of locations to specify and calibrate models with sufficient accuracy and robustness to meet the needs of automated vehicles.

## 7.2 Limitations

Although the work in this dissertation was completed successfully, the following limitations are noted:

- The method developed for classifying road users as pedestrians, motor vehicles and bicyclists is based on the spatial progression of the road users with respect to the geometry of the intersection and the maximum speed. This basic method of classification works well at intersections with separated bicycle traffic (e.g. bicycle facility). However, in mixed traffic situations, the results were less satisfactory and manual correction was necessary. Additionally, a method was developed for post processing trajectory data to remove distortion caused by collecting data with a wide-angle lens. This method was found to remove distortion adequately. Nevertheless,

rectification of video data before the extraction of trajectory data is recommended. This prevents variation in the effects of the parameters that control the trajectory extraction process across the distorted video frame.

- The clustering methods proposed were found to produce excellent clusters for the four research intersections examined in this dissertation. This sample size is relatively small, however, and better results that are applicable to a wider variety of intersections could be achieved if a larger pool of data were collected.

- Although the calibration and validation of the operational behaviour model were successful, the estimation of the change in direction model was hindered by the lack of knowledge concerning the desired direction of the observed bicyclist in each moment. Here, this value was estimated using the approach specific pathways identified using the clustering approach. However, this is only an estimate based on all observed bicyclists arriving on the given approach and does not necessarily match the desired direction of an individual bicyclist. Furthermore, a stipulation was implemented during the integration of the operational model with *SUMO* that constrains the change in angle of simulated bicyclists at low speeds. This stipulation was not included in the model specification and was not calibrated or validated using the trajectory data.

- The predictive power of two of the tactical choice logistic regression models is limited by the low number of observations describing bicyclists performing one or more of the options. The prediction of bicyclists riding on the roadway or sidewalk when a bicycle facility is available, as well as the prediction of bicyclists riding against the mandatory direction of travel, are particularly problematic due to unbalanced observation sets. The collection of more data, particularly from bicyclists executing seldom-occurring behaviours would enable a better prediction of the choice outcome.

## 7.3 Outlook

This dissertation offers an example of using trajectory data directly to improve the microscopic simulation of bicycle traffic. However, rapid advances in the fields of computer vision and data analytics are likely to forge a more fluid connection between reality and traffic simulation. The potential to distinguish movement patterns and quantify the underlying causation of bicyclist behaviour is continually increasing. The transfer of this information into the simulation environment will greatly increase the realism of traffic simulation and the consequent meaningfulness of simulation results.

To this aim, more advanced methods for the automated extraction of trajectories are needed. Potential improvements that were noted during this work include a more advanced method

for road user classification that takes into account the physical shape and the movement patterns of the different types of road users. The calibration of the software used for automated trajectory extraction is currently tedious and could be improved with a (semi-)automated calibration algorithm. Furthermore, methods to improve the automated extraction and post processing of trajectory data, including merging of trajectories belonging to the same road user (e.g. after a stop) or deleting erroneous trajectories would be helpful in the future.

The automated classification of manoeuvres and pathways is also necessary for the fluid connection of reality and simulated environments. Here, a pathway generic clustering method is proposed that is trained for application at signalised, four arm intersections. However, the method used to create the clustering structure can be applied to trajectory datasets collected at other types of locations, such as road segments or T-intersections. The addition of trajectories from other intersections to the existing dataset would improve the overall quality of the clustering structure, as clustering algorithms tend to perform best and deliver more robust results with more observations.

The work in this dissertation provides a step forward in the creation of realistic and accurate models of bicyclist behaviour. However, there is still much to be done before these models can be applied in many situations (e.g. use in automated vehicles). The methods proposed here for calibrating behaviour models using observed trajectory data will continue to have applications. However, considering the difficulties in isolating the desired direction of travel for the calibration of the change in direction component of the operational model, other methods for observing reactions while controlling the desired direction are necessary. Possibilities for such data collection include the use of a bicycle simulator in which the test subject follows a prescribed path. This could enable the identification of parameter distributions rather than population point estimates, which were used in this dissertation because a number of the parameters in the change in direction model were found to have very little variation.

Furthermore, both the operational and tactical models are calibrated using data representing the complete population of bicyclists in Munich. The models are specified to describe the generic movement of all types of bicyclists and the methods outlined in this dissertation can be used to calibrate the models for specific types of bicyclists. Possible attributes to consider in identifying types of bicyclists include personal attributes (e.g. gender and age) and type of bicycle (e.g. E-bike, pedelecs and cargo bicycles). This would enable a more specific prediction of behaviour depending not only on the situation but also based on the bicyclist himself or herself.

# References

ALLEN, D.; BYGRAVE, S. & HARPER, H. [2005]: Behavior at Cycle Advanced Stop Lines, London.

ALLEN, D.; HUMMER, J.; ROUPHAIL, N. & MILAZZO, J. [1998]: Effect of Bicycles on Capacity of Signalized Intersections. Transportation Research Record: Journal of the Transportation Research Board, 1646(1), pp.87–95.

ALRUTZ, D.; BOHLE, W.; MÜLLER, H.; PRAHLOW, H.; HACKE, U. & LOHMANN, G. [2009]: Unfallrisiko und Regelakzeptanz von Fahrradfahrern. Berichte der Bundesanstalt für Straßenwesen, Unterreihe Verkehrstechnik, 184, pp.1–127.

AMINI, S.; TWADDLE, H. & LEONHARDT, A. [2016]: Modeling of the Tactical Path Selection of Bicyclists at Signalized Intersections. In Transportation Research Board 95th Annual Meeting Proceedings.

AULTMAN-HALL, L. & HALL, F. [1998]: Research Design Insights from a Survey of Urban Bicycle Commuters. Transportation Research Record: Journal of the Transportation Research Board, 1636(1), pp.21–28.

BACCHIERI, G.; BARROS, A.J.D.; DOS SANTOS, J. V & GIGANTE, D.P. [2010]: Cycling to work in Brazil: Users profile, risk behaviors, and traffic accident occurrence. Accident Analysis & Prevention, 42(4), pp.1025–1030.

BARCELÓ, J. [2010]: Fundamentals of Traffic Simulation, New York: Springer-Verlag.

BEN-AKIVA, M.E. & LERMAN, S.R. [1985]: Discrete choice analysis: theory and application to travel demand, MIT press.

BOTMA, H. [1995]: Method to determine level of service for bicycle paths and pedestrian-bicycle paths. Transportation Research Record: Journal of the Transportation Research Board, (1502), pp.38–44.

BOTMA, H. & PAPENDRECHT, H. [1991]: Traffic operation of bicycle traffic. Transportation Research Record: Journal of the Transportation Research Board, (1320), pp.65–72.

BUNDESMINISTERIUM FÜR VERKEHR BAU UND STADTENTWICKLUNG [2013]: Straßenverkehrs-Ordnung (StVO), Germany.

CHENG, S.; YAO, D.; YI, Z.; SU, Y. & XU, W. [2008]: A CA model for intrusion conflicts simulation in vehicles-bicycles laminar traffic flow. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, pp.633–638.

DALLMEYER, J. [2014]: Simulation des Straßenverkehrs in der Großstadt: das Mit-und Gegeneinander verschiedener Verkehrsteilnehmertypen, Springer-Verlag.

DATAFROMSKY [2015]: DataFromSky. Retrieved from: http://datafromsky.com/ [Accessed 9. February 2016].

DEPARTMENT FOR TRANSPORT [2017]: Road traffic statistics. Retrieved from: https://www.gov.uk/government/collections/road-traffic-statistics [Accessed 6. June 2017].

DILL, J. & CARR, T. [2003]: Bicycle commuting and facilities in major US cities: if you build them, commuters will use them. Transportation Research Record: Journal of the Transportation Research Board, (1828), pp.116–123.

DLR [2016]: SUMO Wiki Webpage - Information for "Simulation/Bicycles." Retrieved from: http://sumo.dlr.de/wiki/Simulation/Bicycles [Accessed 24. February 2016].

DOWLING, R.; SKABARDONIS, A. & ALEXIADIS, V. [2004]: Traffic analysis toolbox volume III: guidelines for applying traffic microsimulation modeling software,

DUTHIE, J.; BRADY, J.F.; MILLS, A.F. & MACHEMEHL, R.B. [2011]: Effects of On-Street Bicycle Facility Configuration on Bicyclist and Motorist Behavior. Transportation Research Record: Journal of the Transportation Research Board, 2190(1), pp.37–44.

EBERT, A. [2004]: Cycling towards the nation: the use of the bicycle in Germany and the Netherlands, 1880–1940. European Review of History—Revue européenne d'Histoire, 11(3), pp.347–364.

ERDMANN, J. [2015]: SUMO's Lane-Changing Model BT - Modeling Mobility with Open Data: 2nd SUMO Conference 2014 Berlin, Germany, May 15-16, 2014. In M. Behrisch & M. Weber, eds. Cham: Springer International Publishing, pp. 105–123.

FACULTY OF ENGINEERING LTH [2015]: T-Analyst. Retrieved from: http://www.tft.lth.se/en/research/video-analysis/co-operation/software/t-analyst/ [Accessed 23. March 2016].

FALKENBERG, G.; BLASE, A.; BONFRANCHI, T.; COSSE, L.; DRAEGER, W.; VORTISCH, P.; KAUTZSCH, L.; STAPF, H. & ZIMMERMANN, A. [2003]: Bemessung von Radverkehrsanlagen unter verkehrstechnischen Gesichtspunkten. Berichte der Bundesanstalt fuer Strassenwesen, Unterreihe Verkehrstechnik, 103.

FAWCETT, T. [2006]: An introduction to ROC analysis. Pattern recognition letters, 27(8), pp.861–874.

FELLENDORF, M. & VORTISCH, P. [2010]: Microscopic traffic flow simulator VISSIM. In Fundamentals of traffic simulation. Springer, pp. 63–93.

FERRARA, T.C. [1975]: A Study of Two-Lane Intersections and Crossings Under Combined Motor Vehicle and Bicycle Demands, University of California Davis.

FIGLIOZZI, M.; WHEELER, N. & MONSERE, C. [2013]: A Methodology to Estimate Bicyclists' Acceleration and Speed Distributions at Signalized Intersections. Transportation Research Record: Journal of the Transportation Research Board, 2387, pp.66–75.

FORSCHUNGSGESELLSCHAFT FÜR STRAßEN- UND VERKEHRSWESEN [2010]: Empfehlung für Radverkehrsanlagen (ERA).

FORSCHUNGSGESELLSCHAFT FÜR STRAßEN- UND VERKEHRSWESEN [2015]: Handbuch für die Bemessung von Straßenverkehrsanlagen (HBS).

FREY, B.J. & DUECK, D. [2007]: Clustering by passing messages between data points. Science, 315(5814), pp.972–976.

GERSTENBERGER, M. [2015]: Accidents at Intersections: Basic Analysis of Causation and Approaches for Improvements. Dissertation. Doctoral dissertation Technical University of Munich.

GOULD, G. & KARNER, A. [2010]: Modeling Bicycle Facility Operation. Transportation Research Record: Journal of the Transportation Research Board, 2140, pp.157–164.

GREEN, P. [2007]: Why Driving Performance Measures are Sometimes Not Accurate (and Methods to Check Accuracy). Proceedings of the Fourth International Driving Symposium on Hu man Factors in Driver Assessment, Training and Vehicle Design, pp.394–400.

GUO, H.; WANG, W.; GUO, W. & ZHAO, F. [2013]: Modeling lane-keeping behavior of bicyclists using survival analysis approach. Discrete Dynamics in Nature and Society, 2013.

GUYON, I. & ELISSEEFF, A. [2003]: An introduction to variable and feature selection. Journal of machine learning research, 3, pp.1157–1182.

HARKEY, D. & STEWART, J. [1997]: Evaluation of Shared-Use Facilities for Bicycles and Motor Vehicles. Transportation Research Record: Journal of the Transportation Research Board, 1578(1), pp.111–118.

HELBING, D. & MOLNAR, P. [1995]: Social force model for pedestrian dynamics. Physical Review E, 51(5), p.4282.

HENSHER, D.A.; ROSE, J.M. & GREENE, W.H. [2005]: Applied choice analysis: a primer, Cambridge University Press.

HERSLUND, M.-B. & JØRGENSEN, N.O. [2003]: Looked-but-failed-to-see-errors in traffic. Accident Analysis & Prevention, 35(6), pp.885–891.

HOOGENDOORN, S.P. [2001]: Normative pedestrian flow behavior theory and applications, Delft University of Technology, Faculty Civil Engineering and Geosciences.

HOOGENDOORN, S.P. & DAAMEN, W. [2007]: Microscopic calibration and validation of pedestrian models: Cross-comparison of models using experimental data. In Traffic and Granular Flow'05. Springer, pp. 329–340.

HOSMER, D.W.J.; LEMESHOW, S. & STURDIVANT, R.X. [2013]: Applied logistic regression, John Wiley & Sons.

VAN HOUTEN, R. & SEIDERMAN, C. [2005]: Part 1: Bicycles: How Pavement Markings Influence Bicycle and Motor Vehicle Positioning: Case Study in Cambridge, Massachusetts. Transportation Research Record: Journal of the Transportation Research Board, 1939(1), pp.1–14.

HUNTER, W.; STEWART, J. & STUTTS, J. [1999]: Study of Bicycle Lanes Versus Wide Curb Lanes. Transportation Research Record: Journal of the Transportation Research Board, 1674(1), pp.70–77.

HUNTER, W.W. & STEWART, J.R. [1999]: An Evaluation of Bike Lanes Adjacent To Motor Vehicle Parking an Evaluation of Bike Lanes, University of North Carolina.

JACKSON, S.; MIRANDA-MORENO, L.; ST-AUBIN, P. & SAUNIER, N. [2013]: A Flexible, Mobile Video Camera System and Open Source Video Analysis Software for Road Safety and Behavioural Analysis. Transportation Research Record: Journal of the Transportation Research Board, 2365, pp.90–98.

JAIN, A.K. & DUBES, R.C. [1988]: Algorithms for clustering data, Prentice-Hall, Inc.

JIA, B.; LI, X.G.; JIANG, R. & GAO, Z.Y. [2007]: Multi-value cellular automata model for mixed bicycle flow. European Physical Journal B, 56(3), pp.247–252.

JIANG, R.; JIA, B. & WU, Q.-S. [2004]: Stochastic multi-value cellular automata models for bicycle flow. Journal of Physics A: Mathematical and General, 37(6), pp.2063–2072.

JOHNSON, M.; CHARLTON, J.; OXLEY, J. & NEWSTEAD, S. [2013]: Why do cyclists infringe at red lights? An investigation of Australian cyclists' reasons for red light infringement. Accident Analysis & Prevention, 50, pp.840–847.

JOHNSON, M.; NEWSTEAD, S.; CHARLTON, J. & OXLEY, J. [2011]: Riding through red lights: The rate, characteristics and risk factors of non-compliant urban commuter cyclists. Accident Analysis & Prevention, 43(1), pp.323–328.

JUNGHANS, M.; KOZEMPEL, K. & SAUL, H. [2015]: Chances for the evaluation of the traffic safety risk at intersections by novel methods. Transport Rossijskoj Federacii, 56(1), pp.56–63.

KENNISINSTITUUT VOOR MOBILITEITSBELEID (KIM) [2013]: Mobiliteitsbalans 2013,

KHAN, S. & RAKSUNTORN, W. [2001]: Characteristics of Passing and Meeting Maneuvers on Exclusive Bicycle Paths. Transportation Research Record: Journal of the Transportation Research Board, 1776(1), pp.220–228.

KITANI, K.; ZIEBART, B.D.; BAGNELL, J.A. & HEBERT, M. [2012]: Activity Forecasting. In European Conference on Computer Vision.

KRAJZEWICZ, D.; BONERT, M. & WAGNER, P. [2006]: The open source traffic simulation package SUMO. RoboCup 2006 Infrastructure Simulation Competition, pp.1–5.

KRAJZEWICZ, D.; ERDMANN, J.; HÄRRI, J. & SPYROPOULOS, T. [2014]: Including Pedestrian and Bicycle Traffic in the Traffic Simulation SUMO. In 10th ITS European Congress. Helsinki. pp. 1–10.

KRAJZEWICZ, D. & HERTKORN, G. [2002]: SUMO (Simulation of Urban MObility) An open-source traffic simulation. Symposium on Simulation, pp.63–68.

KRAMER, U. [2008]: Kraftfahrzeugführung: Modelle – Simulation – Regelung, Carl Hanser Verlag GmbH & Co. KG.

KRAUß, S. [1998]: Microscopic modeling of traffic flow: Investigation of collision free vehicle dynamics. Dissertation.

KROLL, B.J. & RAMEY, M.R. [1977]: Effects of bike lanes on driver and bicyclist behavior. Transportation engineering journal of the American Society of Civil Engineers, 103(2), pp.243–256.

KUDERER, M.; KRETZSCHMAR, H. & BURGARD, W. [2013]: Teaching mobile robots to cooperatively navigate in populated environments. In 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, pp. 3138–3143.

KUHN, M. [2008]: Building Predictive Models in R Using the caret Package. Journal of Statistical Software, 28(5).

KUHN, M. [2016]: The caret Package. Retrieved from: http://topepo.github.io/caret/index.html [Accessed 18. July 2016].

KULLER, E.C.; GERSEMANN, D. & RUWENSTROTH, G. [1986]: Regelabweichendes Verhalten von Fahrradfahrern. Forschungsberichte der Bundesanstalt für Straßenwesen, Bereich Unfallforschung, (142).

LAROSE, D.T. & LAROSE, C.D. [2015]: Data mining and predictive analytics, John Wiley & Sons.

LAWSON, S.D. [1991]: Red-light running: accidents and surveillance cameras.

LEICH, A.; JUNGHANS, M.; KOZEMPEL, K. & SAUL, H. [2015]: Road user tracker based on robust regression with GNC and preconditioning. In IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics.

LI, M.; SHI, F. & CHEN, D. [2011]: Analyze bicycle-car mixed flow by social force model for collision risk evaluation. 3rd International Conference on Road Safety and Simulation, pp.1–22.

LIANG, X.; MAO, B. & XU, Q. [2012]: Psychological-Physical Force Model for Bicycle Dynamics. Journal of Transportation Systems Engineering and Information Technology, 12(2), pp.91–97.

LING, H. & WU, J. [2004]: A study on cyclist behavior at signalized intersections. IEEE Transactions on Intelligent Transportation Systems, 5(4), pp.293–299.

LING, H. & WU, J. [2009]: Cyclists' path planning behavioral model at unsignalized mixed traffic intersections in China. Intelligent Transportation Systems Magazine, IEEE, 1(2), pp.13–19.

LOCE, R. & SABER, E. [2015]: Video Surveillance and Transportation Imaging Applications 2015, Spie Press.

LONG, J.S. [1997]: Regression Models for Categorical and Limited Dependent Variables, Taylor & Francis.

LUO, D. & MA, X. [2016]: Modeling of Cyclist Acceleration Behavior Using Naturalistic GPS Data. In TRB 95th Annual Meeting Compendium of Papers. Washington, DC, pp. 1–14.

LUSK, A.C.; FURTH, P.G.; MORENCY, P.; MIRANDA-MORENO, L.F.; WILLETT, W.C. & DENNERLEIN, J.T. [2011]: Risk of injury for bicycling on cycle tracks versus in the street. Injury prevention, 17(2), pp.131–135.

MALLIKARJUNA, C. & RAMACHANDRA RAO, K. [2008]: Cellular Automata Model for Heterogeneous Traffic. Journal of Advanced Transportation, 43(3), pp.321–345.

McHenry, S.R. & Wallace, M.J. [1985]: Evaluation of Wide Curb Lanes as Shared Lane Bicycle Facilities, Baltimore.

Michon, J.A. [1985]: A critical view of driver behavior models: what do we know, what should we do? Human behavior and traffic safety, pp.485–520.

Miovision [2016]: miovision rethink traffic. Retrieved from: https://miovision.com/ [Accessed 9. February 2016].

Mohamed, M. & Saunier, N. [2013]: Motion prediction methods for surrogate safety analysis. Transportation Research Record: Journal of the Transportation Research Board, 2386, pp.168–178.

van der Molen, H.H. & Bötticher, A.M.T. [1988]: A hierarchical risk model for traffic participants. Ergonomics, 31(4), pp.537–555.

Moritz, W. [1998]: Adult bicyclists in the United States: characteristics and riding experience in 1996. Transportation Research Record: Journal of the Transportation Research Board, 1636, pp.1–7.

Näätänen, R. & Summala, H. [1974]: A model for the role of motivational factors in drivers' decision-making. Accident Analysis & Prevention, 6(3), pp.243–261.

Näätänen, R. & Summala, H. [1975]: A simple method for simulating danger-related aspects of behavior in hazardous activities. Accident Analysis & Prevention, 7(1), pp.63–70.

Nagel, K. & Schreckenberg, M. [1992]: A cellular automaton model for freeway traffic. Journal de Physique I, 2(12), pp.2221–2229.

National Research Council [2000]: Highway Capacity Manual 2000, Washington, DC.

National Research Council [2010]: Highway Capacity Manual 2010, Washington, DC.

Oldenziel, R. & Albert de la Bruhèze, A. [2011]: Contested spaces: bicycle lanes in urban Europe, 1900–1995. Transfers, 1(2), pp.29–49.

OpenCV [2015]: OpenCV. Retrieved from: http://opencv.org/ [Accessed 14. December 2016].

Opiela, K.S.; Khasnabis, S. & Datta, T.K. [1980]: Determination of characteristics of bicycle traffic at urban intersections. Transportation Research Record: Journal of the Transportation Research Board, 743, pp.30–38.

Pai, C.-W. & Jou, R.-C. [2014]: Cyclists' red-light running behaviours: An examination of risk-taking, opportunistic, and law-obeying behaviours. Accident Analysis & Prevention, 62, pp.191–198.

Pang-Ning, T.; Steinbach, M. & Kumar, V. [2006]: Cluster Analysis: Basic Concepts and Algorithms. In Introduction to Data Mining. Pearson Addison-Wesley, pp. 487–568.

Parkin, J. & Rotheram, J. [2010]: Design speeds and acceleration characteristics of bicycle traffic for use in planning, design and appraisal. Transport Policy, 17(5), pp.335–341.

Pein, W. [1997]: Bicyclist Performance on a Multiuse Trails. Transportation Research Record:

Journal of the Transportation Research Board, 1578(1), pp.127–131.

PIN, C.; SAYED, T. & ZAKI, M. [2015]: Assessing safety improvements to pedestrian crossings using automated conflict analysis. Transportation Research Record: Journal of the Transportation Research Board, 2514, pp.58–67.

VAN DER PLOEG, T.; AUSTIN, P.C. & STEYERBERG, E.W. [2014]: Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. BMC Medical Research Methodology, 14(1), pp.1–13.

POWELL, M.J.D. [1994]: A direct search optimization method that models the objective and constraint functions by linear interpolation. In Advances in optimization and numerical analysis. Springer, pp. 51–67.

PROBABILISTIC AND STATISTICAL INFERENCE GROUP [2009]: Affinity Propagation. Retrieved from: http://www.psi.toronto.edu/affinitypropagation/ [Accessed 30. September 2016].

PTV PLANUNG TRANSPORT VERKEHR AG [2015]: VISSIM 8 User Manual.

PUCHER, J. & BUEHLER, R. [2008]: Making Cycling Irresistible: Lessons from The Netherlands, Denmark and Germany. Transport Reviews, 28(4), pp.495–528.

RADKE, S. [2015]: Verkehr in Zahlen 2015/2016, Berlin.

RAKSUNTORN, W. [2002]: A study to examine bicyclist behavior and to develop a microsimulation for mixed traffic at signalized intersections, Denver: University of Colorado at Denver.

RÄSÄNEN, M. & SUMMALA, H. [1998]: Attention and expectation problems in bicycle-car collisions: An in-depth study. Accident Analysis & Prevention, 30(5), pp.657–666.

REYNOLDS, C.C.O.; HARRIS, M.A.; TESCHKE, K.; CRIPTON, P. A & WINTERS, M. [2009]: The impact of transportation infrastructure on bicycling injuries and crashes: a review of the literature. Environmental Health, 8(47).

RICHARDSON, M. & CAULFIELD, B. [2015]: Investigating traffic light violations by cyclists in Dublin City Centre. Accident Analysis & Prevention, 84, pp.65–73.

RICKERT, M.; NAGEL, K.; SCHRECKENBERG, M. & LATOUR, A. [1996]: Two lane traffic simulations using cellular automata. Physica A: Statistical Mechanics and its Applications, 231(4), pp.534–550.

RIVARA, F.P.; THOMPSON, D.C. & THOMPSON, R.S. [1997]: Epidemiology of bicycle injuries and risk factors for serious injury. Injury prevention, 3, pp.110–114.

RODGERS, G.B. [1995]: Bicyclist deaths and fatality risk patterns. Accident Analysis & Prevention, 27(2), pp.215–223.

ROUSSEEUW, P.J. & KAUFMAN, L. [1990]: Finding Groups in Data, Wiley Online Library.

RUBINS, D. & HANDY, S. [2005]: Times of Bicycle Crossings: Case Study of Davis, California. Transportation Research Record: Journal of the Transportation Research Board, 1939(1), pp.22–27.

SAITZ, H.H. [1968]: Zeitlückenuntersuchungen für den Fahrradverkehr an ungesteuerten Straßenknoten. Die Straße, 8(3).

SAUNIER, N. [2016]: Traffic Intelligence Project. Retrieved from: https://bitbucket.org/Nicolas/trafficintelligence/wiki/Home [Accessed 4. February 2016].

SAUNIER, N. & SAYED, T. [2006]: A feature-based tracking algorithm for vehicles in intersections. In Third Canadian Conference on Computer and Robot Vision. IEEE.

SAVITZKY, A. & GOLAY, M.J. [1964]: Smoothing and differentiation of data by simplified least squares procedures. Analytical chemistry, 8, pp.1627–1639.

SAYED, T.; SACCHI, E. & DE LEUR, P. [2016]: Evaluating the Safety Benefits of the Insurance Corporation of British Columbia Road Improvement Program Using Full Bayes Approach. In Transportation Research Board 95th Annual Meeting.

SCHLEINITZ, K.; PETZOLDT, T.; FRANKE-BARTHOLDT, L.; KREMS, J. & GEHLERT, T. [2016]: The German Naturalistic Cycling Study – Comparing cycling speed of riders of different e-bikes and conventional bicycles. Safety Science.

SCHÖNAUER, R. & SCHROM-FEIERTAG, H. [2010]: Mikrosimulation von Mischverkehr – Konzept MiMiSim und Ausblick auf MixME. In 15th International Conference on Urban Planning and Regional Development in the Information Society. pp. 1157–1161.

SCHÖNAUER, R.; STUBENSCHROTT, M.; HUANG, W.; RUDLOFF, C. & FELLENDORF, M. [2012]: Modeling concepts for mixed traffic: Steps towards a microscopic simulation tool for shared space zones. Transportation Research Board 91st Annual Meeting, pp.1–16.

SCHRAMM, A.J.; RAKOTONIRAINY, A. & HAWORTH, N.L. [2008]: How much does disregard of road rules contribute to bicycle-vehicle collisions? In Proceedings High risk road users - motiv ating behaviour change: what works and what doesn't work? National Conference of the Australasian College of Road Safety and the Travelsafe Committe e of the Queensland Parliament. Brisbane.

SCIKIT-LEARN DEVELOPERS [2014]: sci learn Clustering. Retrieved from: http://scikit-learn.org/stable/modules/clustering.

SEMRAU, M.; ERDMANN, J.; FRIEDRICH, B. & WALDMANN, R. [2016]: Simulation framework for testing ADAS in Chinese traffic situations. In SUMO 2016 Conference Proceedings. Berlin, pp. 103–113.

SHLADOVER, S.E.; KIM, Z.; CAO, M.; SHARAFSALEH, A. & JOHNSTON, S. [2011]: Intersection Crossing Times of Bicyclists. Transportation Research Record: Journal of the Transportation Research Board, 2247, pp.91–98.

SIVARAMAN, S. & TRIVEDI, M.M. [2013]: Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis. IEEE Transactions on Intelligent Transportation Systems, 14(4), pp.1773–1795.

SMITH, D.T. [1972]: Bicycle Circulation and Safety Study, San Francisco.

SONKA, M.; HLAVAC, V. & BOYLE, R. [2014]: Image processing, analysis, and machine vision, Cengage Learning.

VON STACKELBERG, H. [1934]: Marktform und gleichgewicht, J. springer.

STATISTISCHES BUNDESAMT [2016]: Destatis. Retrieved from: https://www-genesis.destatis.de/genesis/online/data [Accessed 2. June 2017].

STATISTISCHES BUNDESAMT [2014]: Verkehrsunfälle: Zweiradunfälle im Straßenverkehr, Wiesbaden.

SUMMALA, H.; PASANEN, E.; RÄSÄNEN, M. & SIEVÄNEN, J. [1996]: Bicycle accidents and drivers' visual search at left and right turns. Accident Analysis & Prevention, 28(2), pp.147–153.

SWOV [2016]: Real number of deaths and seriously injured by mode of transport and age. Data - casualties, drivers and crashes. Retrieved from: https://www.swov.nl/UK/Research/Cijfers/Cijfers_ongevallen-UK.htm [Accessed 2. June 2016].

TAHEDL, J. [2014]: Messung des operativen Verhaltens von Radfahrern an Haltelinien von lichtsignalisierten Knotenpunkten. Dissertation. Technical University of Munich.

TAYLOR, D. & DAVIS, W. [1999]: Review of Basic Research in Bicycle Traffic Science, Traffic Operations, and Facility Design. Transportation Research Record: Journal of the Transportation Research Board, 1674(1), pp.102–110.

TAYLOR, D.B. [1993]: Analysis of Traffic Signal Clearance Interal Requirements for Bicycle-Automobile Mixed Traffic. Transportation Research Record: Journal of the Transportation Research Board, 1405, pp.13–20.

TAYLOR, D.H. [1964]: Drivers' galvanic skin response and the risk of accident. Ergonomics, 7(4), pp.439–451.

TESCHKE, K.; HARRIS, M.A.; REYNOLDS, C.; WINTERS, M.; BABUL, S.; CHIPMAN, M.; CUSIMANO, M.; BRUBACHER, J.; HUNTE, G. & FRIEDMAN, S. [2012]: Route infrastructure and the risk of injuries to bicyclists: a case-crossover study. American journal of public health, 102(12), pp.2336–2343.

THE R FOUNDATION [2016]: The R Project for Statistical Computer. Retrieved from: https://www.r-project.org/ [Accessed 18. July 2016].

THE SCIPY COMMUNITY [2016]: SciPy v0.18.0 Reference Guide. Retrieved from: http://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html#scipy.optimize.minimize [Accessed 5. September 2016].

THOMPSON, D.C.; REBOLLEDO, V.; THOMPSON, R.S.; KAUFMAN, A. & RIVARA, F.P. [1997]: Bike speed measurements in a recreational population: validity of self reported speed. Injury Prevention, 3(1), pp.43–45.

TWADDLE, H. [2017]: Analysis and Modelling of the Operational and Tactical Behaviour of Bicyclists. In K. Bengler, S. Hoffmann, D. Manstetten, A. Neukum, & J. Drüke, eds. UR:BAN Human Factors in Traffic. Springer.

TWADDLE, H.; GRIGOROPOULOS, G. & BUSCH, F. [2016A]: An approach for simulating bicycle traffic using observed trajectory data. In SUMO Conference 2016 Post Conference Proceedings. Berlin.

TWADDLE, H.; GRIGOROPOULOS, G. & BUSCH, F. [2016B]: Integration of an external bicycle model in SUMO. In SUMO 2016 Conference Proceedings. Berlin.

TWADDLE, H. & HOFFMANN, S. [2016]: Forschungsprojekt UR:BAN - Mensch im Verkehr Schlussbericht, Germany.

TWADDLE, H.; SCHENDZIELORZ, T. & FAKLER, O. [2014]: Bicycles in Urban Areas: Review of Existing Methods for Modeling Behavior. Transportation Research Record: Journal of the Transportation Research Board, 2434, pp.140–146.

TWADDLE, H.; SCHENDZIELORZ, T.; FAKLER, O. & AMINI, S. [2014]: Use of automated video analysis for the evaluation of bicycle movement and interaction. In Proc. SPIE 9026, Video Surveillance and Transportation Imaging Applications 2014. San Francisco.

UR:BAN PROJECT PARTNERS [2016]: UR:BAN (Urbaner Raum: Betuzergerechte Assistenzsysteme und Netzmanagement) - Ergebnisse, Germany.

VASIĆ, J. & RUSKIN, H. [2012A]: A CA-Based Model for City Traffic Including Bicycles. Urban Development, pp.79–92.

VASIĆ, J. & RUSKIN, H. [2011A]: A Discrete Flow Simulation Model for Urban Road Networks, with Application to Combined Car and Single-File Bicycle Traffic. International Conference on Computational Science and Its Applications, pp.602–614.

VASIĆ, J. & RUSKIN, H. [2012B]: Cellular automata simulation of traffic including cars and bicycles. Physica A: Statistical Mechanics and its Applications, 391(8), pp.2720–2729.

VASIĆ, J. & RUSKIN, H. [2011B]: Throughput and delay in a discrete simulation model for traffic including bicycles on urban networks. In Proceedings of the ITRN2011. University College Cork.

WACHTEL, A. & LEWISTON, D. [1994]: Risk factors for bicycle-motor vehicle collisions at intersections. ITE Journal, 64(9), pp.30–35.

WANG, H.; SU, H.; ZHENG, K.; SADIQ, S. & ZHOU, X. [2013]: An effectiveness study on trajectory similarity measures. In Proceedings of the Twenty-Fourth Australasian Database Conference-Volume 137. Australian Computer Society, Inc., pp. 13–22.

WEGENER, A.; PIORKOWSKI, M.; MAXIM, R.; HELLBRÜCK, H.; FISCHER, S. & HUBAUX, J.-P. [2008]: TraCI: An Interface for Coupling Road Traffic and Network Simulators. In Proceedings of the 11th communications and networking simulation symposium. pp. 155–163.

WHEELER, N.; CONRAD, R. & FIGLIOZZI, M. [2010]: A Statistical Analysis of Bicycle Rider Performance: The impact of gender on riders' performance at signalized intersections. Transportation Research Board 89th Annual Meeting, pp.1–21.

WILDE, G.J.S. [1988]: Risk homeostasis theory and traffic accidents: propositions, deductions and discussion of dissension in recent reactions. Ergonomics, 31(4), pp.441–468.

WILDE, G.J.S. [1982]: The theory of risk homeostasis: implications for safety and health. Risk analysis, 2(4), pp.209–225.

WU, J.; LING, H. & ZHAO, J. [2004]: The Behavior of Cyclists and Pedestrians at Signalized

Intersections in Beijing. Communication and Transportation Systems Engineering and Information, 2, p.23.

YAO, D.; ZHANG, Y.; LI, L.; SU, Y.; CHENG, S. & XU, W. [2009]: Behavior modeling and simulation for conflicts in vehicles-bicycles mixed flow. IEEE Intelligent Transportation Systems Magazine, 1(2), pp.25–30.

ZANGENEHPOUR, S.; MIRANDA-MORENO, L.F. & SAUNIER, N. [2013]: Impact of bicycle boxes on safety of cyclists: a case study in Montreal. In Transportation Research Board 92nd Annual Meeting.

ZHAO, D.; WANG, W.; LI, C.; LI, Z.; FU, P. & HU, X. [2013]: Modeling of passing events in mixed bicycle traffic with cellular automata. Transportation Research Record: Journal of the Transportation Research Board, (2387), pp.26–34.

# Variable list

Section 3 – Experimental design, data collection and processing

| | |
|---|---|
| $N$ | Number of observations |
| $IV$ | Number of independent variables |
| $T$ | Duration of the observation |
| $q$ | Average bicycle traffic flow |
| $P_{min}$ | Estimated proportion of bicyclists faced with the least frequent tactical choice |
| $i$ | Road user |
| $S_i$ | Trajectory of road user $i$ |
| $t$ | Time step |
| $(x_i, y_i)_t$ | Position coordinate of road user $i$ at time step $t$ |
| K | Camera matrix |
| $d$ | Distortion coefficients matrix |

Section 4 – Trajectory clustering

| | |
|---|---|
| $f_{ij}^*$ | Feature $j$ for road user $i$ |
| $F_i^*$ | Feature vector for road user $i$ |
| $\mathcal{A}^*$ | Pattern matrix of all feature vectors |
| $D_i$ | Distance travelled by road user $i$ |
| $r(i,k)$ | "Responsibility" for observation $i$ and exemplar $k$ (Affinity Propagation) |
| $a(i,k)$ | "Ability" for observation $i$ and exemplar $k$ (Affinity Propagation) |
| $\lambda$ | Damping parameter (Affinity Propagation) |
| $n$ | Number of points in a trajectory (clustering) |
| $s_i$ | Silhouette score for observation $i$ |
| $a_i$ | Proximity of an observation $i$ to its assigned exemplar $k_i$ |
| $b_i$ | Proximity of an observation $i$ to the next nearest exemplar |
| $\bar{s}$ | Mean silhouette score |

Section 5 – Modelling bicycle behaviour

| | |
|---|---|
| $E(Y\|x)$ | Expected value of $Y$ given $x$ (linear regression model) |
| $\pi(x)$ | Conditional mean of $Y$ given $x$ (logistic regression model) |
| $x$ | Vector of explanatory variables |
| $\beta_0$ | Constant coefficient associated with the expected value of $Y$ |
| $\beta_1$ | Vector of weighting parameters for $x$ estimated using observed data |

| | |
|---|---|
| $b$ | Modelled bicyclist |
| $\Delta V_b(t)$ | Change in speed of bicyclist $b$ at time $t$ |
| $V_b^0$ | Desired speed of bicyclist $b$ |
| $v_b(t)$ | Velocity vector of bicyclist $b$ at time $t$ |
| $V_b(t)$ | Speed of bicyclist $b$ at time $t$ |
| $T_{vb}$ | Relaxation parameter unique to bicyclist $b$ (speed) |
| $R_{vb}$ | Radius of interaction for bicycle $b$ (speed) |
| $A_{vb}$ | Stopping parameter for bicyclist $b$ calculated from $V_b^0$, $V_b(t)$ and $T_{vb}$ |
| $\Delta\theta_b(t)$ | Change in direction of bicyclist $b$ at time $t$ |
| $\theta_b^0(t)$ | Desired direction of bicyclist $b$ at time $t$ |
| $\theta_b(t)$ | Current direction of bicyclist $b$ at time $t$ |
| $T_{\theta b}$ | Relaxation parameter unique to bicyclist $b$ (direction) |
| $A_{\theta b}$ | Weighting parameter for reaction to interacting road users |
| $R_{\theta b}$ | Radius of interaction for bicyclist $b$ (direction) |
| $U_{bi}(t)$ | Relative position value |
| $d_{bi}(t)$ | Distance between road user $i$ and bicyclist $b$ (vector) |
| $D_{bi}(t)$ | Distance between road user $i$ and bicyclist $b$ (scalar) |
| $\phi$ | Angle between the vectors $v_b(t)$ and $d_{bi}(t)$ |
| $w_b(t)$ | Vector perpendicular to $v_b(t)$ oriented in the direction of road user $i$ |
| $p_b(t)$ | Position of bicyclist $b$ at time $t$ |
| $p_i(t)$ | Position of interacting road user $i$ at time $t$ |
| $D_{bi}^*(t)$ | Adjusted distance between road user $i$ and bicyclist $b$ (Anisotropic model) |
| $D_{bi}^{**}(t)$ | Adjusted distance between road user $i$ and bicyclist $b$ (Velocity anisotropic model) |
| $\eta_b$ | Relative position weighting parameter (bicyclist $b$ and road user $i$) |
| $\gamma_b$ | Relative velocity weighting parameter (bicyclist $b$ and road user $i$) |
| $\mathcal{L}(\beta)$ | Log likelihood of a model for observed data given parameter set $\beta$ |
| $\mathcal{L}(\beta;\hat{\sigma}^2)$ | Log likelihood of a model for observed data given parameter set $\beta$ and standard deviation $\hat{\sigma}^2$ |
| $\hat{\beta}$ | Parameter set leading to the maximum $\mathcal{L}(\beta;\hat{\sigma}^2)$ |
| $\hat{\beta}_b$ | Parameter set leading to the maximum $\mathcal{L}(\beta;\hat{\sigma}^2)$ for bicyclist $b$ |
| $I$ | Average improvement in log likelihood |
| $D$ | Test statistic for the log likelihood ratio test |
| $R_{ab}$ | Correlation between the parameters $a$ and $b$ |

Section 6 – Implementation and evaluation

| | |
|---|---|
| $q$ | Tactical choice option |
| $k$ | Attribute of an option |
| $i$ | Road user making tactical choice |
| $U_{qi}$ | Utility of option $q$ for road user $i$ |
| $\beta_{qi}$ | Alternative specific weighting parameter for road user $i$ and option $q$ |
| $X_{qki}$ | Value of attribute $k$ |
| $P_{qi}$ | Probability of road user $i$ choosing alternative $q$ |
| $J$ | Total number of alternatives |
| $RMSE$ | Root mean square error |

# Glossary

| | |
|---|---|
| Cargo bicycle | A type of bicycle that is equipped with large compartments for carrying cargo. |
| Direct left turn | A left turn carried out in the manner of vehicle traffic. The bicyclist queues with traffic in the left most lane of the roadway and crosses the intersection during one signal phase. |
| E-bike | A type of bicycle that offers electric motor support to the bicyclist until they reach a speed of 45 km/h. |
| Features (computer vision) | Distinctive attributes or patches of pixels with strong gradients in a video or picture frame, such as object corners and edges. |
| Indirect left turn | A turn carried out over two signal phases. Within the first phase, the bicyclist crosses the intersection as if riding straight across the intersection. The bicyclist then stops once across the intersection and waits for the next signal phase to continue the second part of the manoeuvre. |
| Indirect left turn against the mandatory direction of travel | A turn carried out over two signal phases. Within the first phase, the bicyclist crosses the approaching road. The bicyclist then stops once across the roadway and waits for the next signal phase to continue across the intersection and complete the second part of the manoeuvre. |
| Instantaneous speed/acceleration | A speed/acceleration value at a particular point in time. |
| Manoeuvre | Movement carried out by a bicyclist or other road user to realise a route planned at the strategic behavioural level. Examples of manoeuvres include right turns, left turns and riding straight across an intersection. |
| Mean acceleration/speed | The average speed/acceleration of a speed/acceleration profile. |
| On-road bicycle facility | A bicycle facility within the roadway that is marked using a solid or dashed line. |

| | |
|---|---|
| Operational behaviour | Behaviour at this level includes subconscious action patterns that are carried out to realise a manoeuvre while reacting to other road users and obstacles in the environment. Behaviour at this level includes acceleration, deceleration and changing direction. |
| Path | Movement through space to realise a given manoeuvre. Examples of paths include direct and indirect left turns. |
| Pedelec | A type of bicycle that offers electric motor support to the bicyclist until they reach a speed of 25 km/h. |
| Separated bicycle facility | A marked bicycle facility that is physically separated from the roadway by a curb, a green strip, parked vehicles or another type of barrier. |
| Shared facility | A facility that is shared by bicycle traffic and motor vehicles and/or pedestrians. |
| Speed/acceleration profile | A sequence of speed/acceleration values over time. |
| Tactical behaviour | Behaviour at this level includes conscious decisions made by a road user to achieve strategic goals while coping with the current situation. Examples of behaviour at this level include path planning and infrastructure selection (roadway, sidewalk or bicycle facility). |
| Trajectory | Observed movement through space, which is characterised by vectors of position coordinates. Velocity and acceleration coordinates can be derived from the position coordinates. |

# List of figures

# List of tables

# Appendix 1

## Full model parameters binomial and multinomial regression models

| **N = 451**<br>**Sidewalk use = 0, Roadway use = 1** | $\beta$ | **Odds Ratio** | $p$ |
|---|---|---|---|
| Intercept | 3.75 | 42.35 | 0.002 |
| Cars in approach | -0.56 | 0.57 | 0.001 |
| Pedestrians in approach | 0.15 | 1.16 | 0.126 |
| Manoeuvre (right turn) | -0.88 | 0.41 | 0.211 |
| Trucks in approach | -0.81 | 0.45 | 0.273 |
| Driving lanes (same direction) | 0.26 | 1.30 | 0.713 |
| Manoeuvre (left turn) * Driving lanes (same direction) | -1.00 | 0.37 | 0.000 |
| | | | |
| Classification threshold: | | 0.94 | |
| AUC | | 0.83 | |
| Accuracy | | 0.80 | |
| Sensitivity | | 0.80 | |
| Specificity | | 0.78 | |
| Positive predictive value | | 0.99 | |
| Negative predictive value | | 0.17 | |

**Tab 1** Full binomial logistic regression model with evaluation for infrastructure selection without bicycle facility

| **N = 3727**<br>**No bicycle facility use = 0, Bicycle facility use = 1** | $\beta$ | **Odds Ratio** | $p$ |
|---|---|---|---|
| Intercept | -38.82 | 0.00 | 0.000 |
| Manoeuvre (right turn) | -2.11 | 0.12 | 0.000 |
| Bicyclist volume – approach (bicycle/h) | 0.10 | 1.10 | 0.000 |
| Bicycle facility width (m) | 27.02 | 5.43e11 | 0.000 |
| Bicycle facility type (separated) | -6.99 | 0.00 | 0.000 |
| Driving lanes (same direction) | 1.26 | 3.51 | 0.000 |
| Sidewalk width (m) | -1.96 | 0.14 | 0.000 |
| Centre island | 2.67 | 14.48 | 0.000 |
| Parking | -1.74 | 0.17 | 0.001 |
| Right lane occupied | -0.68 | 0.51 | 0.047 |
| Pedestrians in approach | 0.87 | 2.38 | 0.059 |
| Bicyclists in approach | -0.03 | 0.97 | 0.544 |
| Cars in approach | -0.02 | 0.98 | 0.725 |
| Manoeuvre (left turn) | 0.43 | 1.54 | 0.776 |
| Bicycle facility width (m) * Bicyclist volume – approach (bicycle/h) | -0.06 | 0.94 | 0.000 |
| Bicycle facility type (separated) * Bicyclists in approach | -0.27 | 0.77 | 0.001 |
| Bicycle facility type (separated) * Sidewalk width (m) | 1.11 | 3.03 | 0.006 |
| Cars in approach * Pedestrians in approach | 0.08 | 1.08 | 0.007 |
| Bicycle facility width (m) * Pedestrians in approach | -0.58 | 0.56 | 0.016 |
| Right lane occupied * Parking | 0.95 | 2.59 | 0.026 |
| Manoeuvre (left turn) * Bicycle facility width (m) | -1.28 | 0.28 | 0.116 |
| Right lane occupied * Manoeuvre (right turn) | -0.41 | 0.66 | 0.309 |
| | | | |
| Classification threshold: | | 0.95 | |
| AUC | | 0.76 | |
| Accuracy | | 0.73 | |
| Sensitivity | | 0.73 | |
| Specificity | | 0.78 | |
| Positive predictive value | | 0.98 | |

**Tab 2** Full binomial logistic regression model with evaluation for infrastructure selection with bicycle facility

| **N = 1935**<br>**Stop = 0, Violate = 1** | $\beta$ | **Odds Ratio** | $p$ |
|---|---|---|---|
| Intercept | -1.61 | 0.20 | 0.000 |
| Manoeuvre (right turn) | 4.87 | 129.80 | 0.000 |
| Time since signal change (s) | -0.03 | 0.97 | 0.000 |
| Manoeuvre (left turn) | 2.45 | 11.53 | 0.000 |
| Roadway width – opposite (m) | -0.35 | 0.71 | 0.000 |
| Infrastructure selection (sidewalk) | 0.76 | 2.13 | 0.025 |
| Bicycle facility | 0.90 | 2.46 | 0.039 |
| Cars in approach | -0.52 | 0.60 | 0.047 |
| Pedestrians in approach | 0.16 | 1.17 | 0.056 |
| Bicyclists in approach | -0.14 | 0.87 | 0.233 |
| Manoeuvre (left turn) * Roadway width – opposite (m) | 0.32 | 1.38 | 0.004 |
| Bicycle facility * Pedestrians in approach | -0.07 | 0.93 | 0.460 |
| | | | |
| Classification threshold: | | 0.36 | |
| AUC | | 0.93 | |
| Accuracy | | 0.91 | |
| Sensitivity | | 0.85 | |
| Specificity | | 0.92 | |
| Positive predictive value | | 0.73 | |
| Negative predictive value | | 0.96 | |

**Tab 3** Full binomial logistic regression model with evaluation for response to red signal

| **N =4710**<br>**With direction = 0, Against direction = 1** | $\beta$ | **Odds Ratio** | $p$ |
|---|---|---|---|
| Intercept | -5.24 | 0.01 | 0.000 |
| Manoeuvre (left turn) | 2.19 | 8.92 | 0.000 |
| Left turn lane | -2.59 | 0.08 | 0.000 |
| Bicycle facility type (separated) | -4.25 | 0.01 | 0.011 |
| Sidewalk width (m) | 0.55 | 1.73 | 0.043 |
| Cars in approach | 0.24 | 1.27 | 0.161 |
| Roadway width – opposite (m) | -0.10 | 0.90 | 0.364 |
| Manoeuvre (right turn) | 0.79 | 2.20 | 0.505 |
| Parking | -0.05 | 0.95 | 0.965 |
| Pedestrians in approach | 0.01 | 1.01 | 0.971 |
| Parking * Bicycle facility type (separated) | 4.84 | 126.34 | 0.008 |
| Roadway width – opposite (m) * Cars in approach | -0.08 | 0.92 | 0.043 |
| Bicycle facility type (separated) * Cars in approach | 0.27 | 1.31 | 0.214 |
| Parking * Pedestrians in approach | 0.17 | 1.19 | 0.266 |
| Roadway width – opposite (m) * Pedestrians in approach | 0.05 | 1.05 | 0.291 |
| Parking * Manoeuvre (right turn) | -1.17 | 0.31 | 0.401 |
| | | | |
| Classification threshold: | | 0.02 | |
| AUC | | 0.84 | |
| Accuracy | | 0.82 | |
| Sensitivity | | 0.80 | |
| Specificity | | 0.82 | |
| Positive predictive value | | 0.06 | |
| Negative predictive value | | 0.99 | |

**Tab 4** Full binomial logistic regression model with evaluation for direction of travel

| N = 426<br>Base category = Direct left turn | | $\beta$ | Odds Ratio | $p$ |
|---|---|---|---|---|
| Indirect left turn | Intercept | -0.42 | 0.66 | 0.650 |
| | Infrastructure selection (roadway) | -4.46 | 0.01 | 0.000 |
| | Bicycle facility type (separated) | 1.59 | 4.89 | 0.004 |
| | Parking | 0.91 | 2.48 | 0.089 |
| | Signal phase (green) | 1.31 | 3.70 | 0.017 |
| | Right lane occupied | 0.85 | 2.34 | 0.056 |
| | Bicyclists in approach | 0.49 | 1.63 | 0.023 |
| | Pedestrians in approach | -0.43 | 0.65 | 0.144 |
| | Trucks in approach | -0.34 | 0.71 | 0.680 |
| | Bicycle facility width (m) | -0.41 | 0.67 | 0.361 |
| | Sidewalk width (m) | -0.08 | 0.93 | 0.625 |
| | Time since last phase change (s) | 0.03 | 1.03 | 0.057 |
| | Pedestrian volume (peds/h) | 0.00 | 1.00 | 0.173 |
| | Parking * Bicycle facility type (separated) | -0.84 | 0.43 | 0.201 |
| | Signal phase (green) * Bicyclists in approach | -0.68 | 0.51 | 0.010 |
| | Signal phase (green) * Trucks in approach | 0.52 | 1.69 | 0.659 |
| | Pedestrians in approach * Sidewalk width (m) | 0.12 | 1.13 | 0.205 |

**Tab 5** Full multinomial regression model with evaluation for left turn manoeuvre

| **N = 426**<br>**Base category = Direct left turn** | | $\beta$ | **Odds Ratio** | $p$ |
|---|---|---|---|---|
| **Indirect left turn (wrong way)** | Intercept | 0.07 | 1.07 | 0.935 |
| | Infrastructure selection (roadway) | -2.96 | 0.05 | 0.000 |
| | Bicycle facility type (separated) | 3.44 | 31.19 | 0.000 |
| | Parking | 2.71 | 15.07 | 0.000 |
| | Signal phase (green) | -2.31 | 0.10 | 0.000 |
| | Right lane occupied | 0.23 | 1.25 | 0.608 |
| | Bicyclists in approach | 0.33 | 1.40 | 0.108 |
| | Pedestrians in approach | -0.34 | 0.71 | 0.167 |
| | Trucks in approach | 0.31 | 1.36 | 0.672 |
| | Bicycle facility width (m) | 0.09 | 1.09 | 0.820 |
| | Sidewalk width (m) | -0.34 | 0.71 | 0.030 |
| | Time since last phase change (s) | -0.02 | 0.99 | 0.338 |
| | Pedestrian volume (peds/h) | 0.00 | 1.00 | 0.001 |
| | Parking * Bicycle facility type (separated) | -2.72 | 0.07 | 0.000 |
| | Signal phase (green) * Bicyclists in approach | -1.19 | 0.31 | 0.016 |
| | Signal phase (green) * Trucks in approach | -0.49 | 0.61 | 0.748 |
| | Pedestrians in approach * Sidewalk width (m) | 0.21 | 1.23 | 0.021 |
| Accuracy | | | 0.76 | |
| Mean Sensitivity | | | 0.73 | |
| Mean Specificity | | | 0.87 | |
| Mean Positive predictive value | | | 0.75 | |
| Mean Negative predictive value | | | 0.87 | |

**Tab 6** Full multinomial regression model with evaluation for left turn manoeuvre (cont.)