EINDHOVEN UNIVERSITY OF TECHNOLOGY

Eindhoven University of Technology

MASTER

Data-Driven Estimation of Equilibrium Distribution Functions for Pedestrian Dynamics

Cleven, Frenk J.P.L.

Award date:
2021

Link to publication

Eindhoven University of Technology

Department of Applied Physics
Fluids and Flows

Data-Driven Estimation of Equilibrium Distribution Functions for Pedestrian Dynamics

Author:
F.J.P.L. Cleven

Supervisors:
dr. Alessandro Corbetta
prof. dr. Federico Toschi

A thesis submitted for the degree of

MSc Applied Physics

16th July 2021

R-2063-A

Abstract

The dynamics of humans walking in crowds are highly complex due to continuous interactions between pedestrians and the surrounding environment. This complexity is particularly true for urban environments, where pedestrians often experience high-density conditions. Recent technological advancements have enabled real-life high-accuracy measurements of pedestrian trajectory data through the usage of overhead depth-sensing cameras. This approach allows for the anonymous, large-scale acquisition of pedestrian trajectories without compromising on quality or privacy. In this research, a mesoscopic Boltzmann-like analysis method is proposed for the analysis of pedestrian trajectory data. Using a Python processing pipeline, an estimation of the equilibrium distribution functions in phase space is made. Additionally, a microscopic pedestrian simulation model has been created, which allows for the generation of artificial trajectories based on the measured equilibrium distributions. Pedestrian trajectories are generated using a Monte Carlo approach, in which pedestrian velocities are conditionally sampled from the measured phase space distributions. A trajectory measurement setup has been installed in Amsterdam to acquire a large trajectory dataset, consisting of approximately 50.000 measured pedestrian trajectories. Measurements took place during the first seven evenings of September 2019, before the outbreak of the COVID-19 pandemic. A strongly bidirectional pedestrian flow is observed in the dataset, and it is shown that the analysis model can memorise the statistics of the captured pedestrian motion efficiently. A thorough analysis of the measured trajectories is presented, along with an in-depth comparison with the artificial trajectories generated by the simulation model. The simulated microscopic pedestrian behaviour shows good agreement with the measured behaviour. As the proposed analysis approach features many similarities with the Lattice Boltzmann Method (LBM), future research could create a genuine data-driven mesoscopic pedestrian simulation approach.

Acknowledgements

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Walking is an essential and ubiquitous mode of transportation for many forms of life on earth. This observation holds just as well for humans, for which walking is the primary and one of the healthiest forms of transportation. The dynamics of human crowds are highly complex due to the continuous interactions with the environment and other surrounding pedestrians. This complexity is particularly true for urban environments, where pedestrians often experience high-density conditions. Since pedestrian motion is crucial for the design of comfortable and safe walkways, it is imperative to have a complete understanding of pedestrian dynamics. Over the last decades, the dynamics of human crowds have seen significant interest among different scientific branches, including engineering, physics, psychology and social sciences.

In the context of physics, pedestrian dynamics is modelled in multiple ways. Although an extensive literature review is presented in this work, a taxonomical classification can be made using the modelling scale. Firstly, a straightforward approach is the microscopic modelling of crowd dynamics, in which pedestrians are modelled as individual particles that can propagate themselves and interact with each other. Secondly, a macroscopic approach has been proven to correctly describe pedestrian motion, in which pedestrian flows are modelled as a continuous fluid. The third category considers the mesoscopic description, which does not track individual particles. Instead, it tracks distributions of representative collections of pedestrians. These concepts are further elaborated in the remainder of this work, and in particular in Chapter 5, where a novel mesoscopic analysis method is proposed.

In recent years, technological and scientific advancements have enabled real-life high-accuracy measurements of pedestrian trajectory data. The usage of overhead depth-sensing cameras allows for the anonymous, large-scale acquisition of pedestrian trajectories without compromising quality or privacy. Before the arrival of such data, only qualitative models of pedestrian behaviour were available, but these developments have enabled research on quantitive models. Large-scale trajectory data opens up new possibilities for research on statistical descriptions of pedestrian ensembles, but many other applications have already been published [1][2][3] as well.

In this thesis, a framework for a set of novel simulation methods of pedestrian dynamics is proposed. The framework is based on the concept of probability distribution functions. A data-driven statistical description of the pedestrian motion in phase space is created using large-scale pedestrian trajectory measurements. Additionally, a microscopic simulation model is developed, allowing for the generation of artificial trajectories based on this statistical description. The simulation model shows promising results, and the proposed analysis method enables further research on data-driven mesoscopic approaches.

The content of this thesis is organised into three parts. In Part I, a concise literature review is given on the modelling and simulation of pedestrian dynamics. Specifically, Chapter 2 gives a brief introduction to the fundamentals of crowds dynamics, and Chapter 3 provides an overview of pedestrian modelling approaches. A review on the state of the art of the recording of pedestrian trajectories is given as well in Chapter 4. In Part II, a data-driven simulation method based on measured probability distribution functions is provided (Chapter 5). In Part III, the simulation method is applied and validated against a real-life pedestrian trajectory dataset. This dataset was acquired in Amsterdam, where a large-scale trajectory measurement setup has been installed. Details on the measurement setup and the statistics of the measured trajectories are given in Chapter 6. An overview of the measured equilibrium distribution functions is given in Chapter 7. A complete validation of the proposed simulation method is given in Chapter 8, which also provides a comparison between the measured and artificially generated trajectories. A final discussion about the generalisability of this approach in Chapter 9 closes this thesis.

Part I

Literature Review

Chapter 2

Fundamentals of Pedestrian Dynamics

In this chapter, a general introduction to the concept of pedestrian dynamics is provided. A historic overview of pedestrian research is provided, along with a summary of relevant definitions and research methods.

2.1 Background on Pedestrian Research

Since the late eighteenth century, theories on human walking and pedestrian movement have been developed from many scientific perspectives [4]. Ongoing research has created widespread and diverse knowledge on this subject, branching research into many different specialisms of pedestrian research. Over the years, multiple literature review papers [5][6] have become available which aid to create taxonomy in the available literature on pedestrian dynamics.

In 1895, Gustave Le Bon stated in [7] that the conscious personality of the individual in a crowd is submerged and that the collective crowd mind dominates; crowd behaviour is unanimous, emotional, and intellectually weak. In the second half of the twentieth century, research was focused on social behaviour in crowded situations, by studying e.g. emergency evacuations and the relation to the corresponding domain layout. In the 1970s, analytical formulas for crowd phenomena were derived from empirical data. The following decade, a split in the research activities occurred: experimental work was joined by studies aided by technology (e.g., computer vision) and computational simulations for graphic applications. In this era, simulations evolved from providing basic numerical data outputs to complex three-dimensional virtual environments.

In the current age, scientific works range from the understanding of single pedestrian behaviour to dynamic crowd interactions. The COVID-19 pandemic has proven that human movements are of extreme relevance for modern society as well [8]. Data collection methods mainly include real-life field observations, controlled experiments, survey-based methods and pedestrian simulation approaches. Recent work shows enormous potential for data collection methods, but restrictions are present as well. For example, field observations are limited by privacy-related issues and controlled experiments often fail to realistically represent real-life scenarios.

Research on pedestrian dynamics is characterised by a very large heterogeneity in published works. This is caused by a large range in the branches of science that research pedestrian dynamics, including computer science, engineering, mathematics, physics, psychology and social science. Additionally, different works aim to study different phenomena, e.g., emergence of crowd self-organisation, vibrations in bridges caused by walking crowds or emergency evacuations. Lastly, the different scales of simulation representation (macroscopic, mesoscopic, microscopic) and the different pedestrian group sizes (individuals, groups, crowds) are to blame for this heterogeneity as well.

2.2 Fundamentals of Pedestrian Dynamics

In order to streamline further discussions on pedestrian dynamics, it is beneficial to introduce some definitions on topics presented in this thesis. Currently, there is some disagreement on literature definitions due to a great heterogeneity in published works. In 2019, the Consortium for the Physics and Psychology of Human Crowd Dynamics, constituted a glossary of terms related to crowd research [9]. Their work is not presented as an absolute truth on formal definitions but reflects current views and

used interpretations of crowd-related terminology. These definitions will be used as a guideline in this thesis as well. In the glossary, a pedestrian is defined as a person moving on foot in a publicly accessible area. Further refinement of different pedestrian types is possible by including their motivation, such as pedestrian-commuter, pedestrian-shopper, or pedestrian-traveller. There is no clear agreement whether motionless persons should be still considered pedestrians, but nonmoving persons are still considered pedestrians in this thesis.

Literature freely uses and interchanges words such as groups and crowds for every situation in which more than two individuals are interacting. In the glossary, a crowd is used to indicate more than two individuals being present at roughly the same place at the same time: an aggregation of people in a physical or non-physical space. Crowds can further be refined into physical crowds, indicating individuals in the same space at the same time, and psychological crowds, referring to individuals that share a common social identity. Smaller aggregated pedestrian collections are often referred to as groups: a number of people which are connected for some time by some measure. This term is often used to refer to a subset of pedestrians in a crowd, without specifying the nature of their relation. Their grouping can follow from purely physical features (e.g. being spatially close), or follow from social relations (friends, family) or interactions (pedestrians having a conversation while walking together).

The dynamics of human mobility are generally non-trivial to study. The first argument arises from the fundamentals of human movement, as movements are not confined to lanes out tracks. The results in a larger freedom of motion compared to other forms of transportation e.g., bikes, cars, or trains, as humans can move in any direction. Secondly, the enormous diversity in walking grounds such as roads, sidewalks, nature or stairs, leads to different behaviour as well. Lastly, psychological aspects and cultural standard are proven to cause variations in measurements [10].

As mentioned, the different scales of representation also lead to very different analytic approaches. Macroscopic experiments and simulations focus on crowd flow quantities, looking into relationships between pedestrian speed, densities and flow rate. Macroscopic approaches do not give insight into the dynamics of individual pedestrians. Typically, pedestrian dynamics are studied using the following quantities:

- **Macroscopic Quantities**
  The most straightforward analysis of pedestrian motion is performed by analysing macroscopic quantities, such as walking speeds and densities. From empirical evidence follows that walking speed is normally distributed and that the mean and variance are strongly dependent on the domain characteristics and group dynamics. Likewise, the pedestrian density is a classic quantity used in analyses. Walking pedestrians require more individual space compared to motionless pedestrians.

- **Fundamental Diagram**
  Macroscopic quantities are typically described by fundamental diagrams as well, which show the relationship between these quantities such as speed, flow and density [11]. These diagrams are used as a primary tool to assess infrastructural capacity and facility designs. Generally, the diagram describes the relationship between the pedestrian current J (pedestrians crossing a surface per time) as a function of the pedestrian density ρ. A typical fundamental diagram is shown in Figure 2.1. Fundamental diagrams remain to be a focus of ongoing research, especially in the limit of dense conditions and emergency evacuations.

- **Level of Service (Engineering)**
  The level of service is a method to describe comfort level for pedestrian movements, shown in Table 2.1. It was introduced by Fruin in 1971 and is widely used as safety criterion in public places. The method consists of six comfort levels ranging from free flow (LOS A) and minor obstructions (LOS B) to critically high densities and extreme difficulty walking movements (LOS F). The comfort levels are defined as a function of flow rate (ped min^{-1} m^{-1}) and density (ped m^{-1}). The provided comfort level criteria are valid for regular walking conditions but unsuitable for other applications such as evacuations.

- **Trajectory Datasets**
  Lastly, it is of researchers' interest to collect trajectory datasets. A trajectory is defined as the path a person follows through space as a function of time. Using modern data collection techniques, researchers are able to create datasets of pedestrian paths. Acquiring such datasets is non-trivial, but has shown enormous potential for pedestrian research. For one, this allows pedestrian sim-

ulations to based on valid empirical evidence instead of approximated assumptions such as the fundamental diagram. This is further detailed in Chapter 4.

Table 2.1: Fruin's Level of Service (LOS) standard. Six comfort levels (A through F) are defined as function of multiple macroscopic flow descriptors.

Level of Service	Flow Rate (ped min^{-1} m^{-1})	Density (ped m^{-1})
A	≤ 7	≤ 0.08
B	7 - 23	0.08 - 0.27
C	23 - 33	0.27 - 0.45
D	33 - 49	0.45 - 0.69
E	49 - 82	0.69 - 1.66
F	≥ 82	≥ 1.66
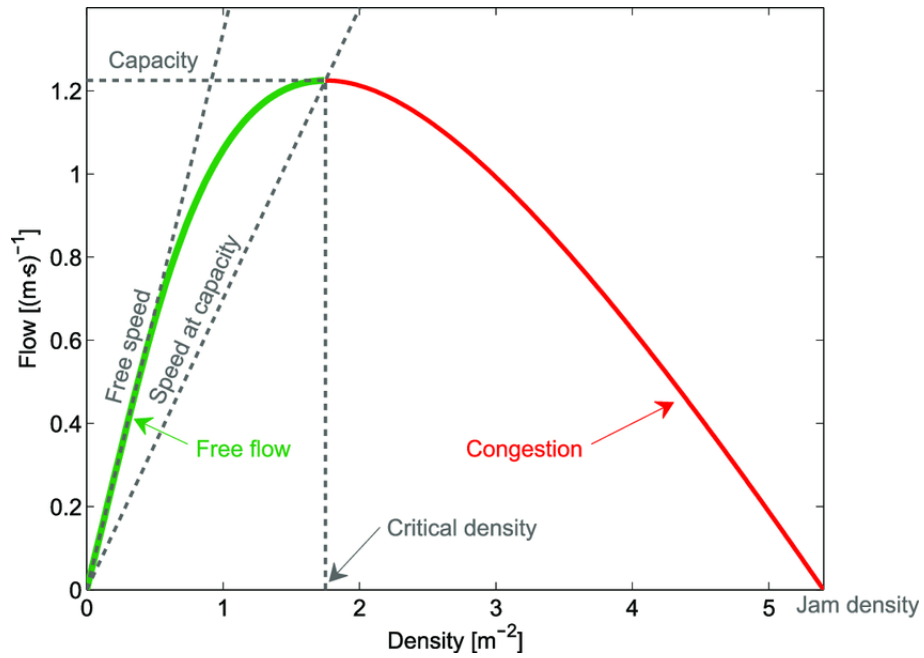
Figure 2.1: A typical fundamental diagram, reproduced from Feliciani et al. [12].

Chapter 3

Pedestrian Dynamics Modelling

Before presenting the proposed pedestrian simulation model, a review on the current state of art concerning pedestrian simulation methods is given. In recent years Duives et al. [6] and Martinez-Gil et al. [5] have performed literature reviews on the modelling, simulation and validation of pedestrian dynamics. In the next section an overview of their findings is given, closely following their work. This is followed by Sections 3.2 and 3.3, giving an in-depth look on simulations of pedestrian dynamics using cellular automata models and data-driven models respectively. To conclude, a research gap present in current literature is introduced in Section ??.

3.1 Overview of Pedestrian Dynamics Modelling

Many simulation models have been proposed by researchers to date, and since research on pedestrian dynamics features many angles of approach, a large heterogeneity in published works is present. Moreover, different researches are motivated by different goals: where some works are focused on the emergence of self-organising behaviours, other wish to reduce evacuation times in pedestrian evacuations. Furthermore, a simulation approach is always created to model a specific situation. Since the layouts, movements and motivations strongly differ between crowded situations, many different modelling approaches exist. For example, considering a festival site, pedestrians roam freely from one stage to another through open spaces. Contrarily, in a subway exit tunnel, people move in a single direction towards the exit through a narrow passage. In both situations, dangerous events could occur at high pedestrian densities, but the dynamics are greatly different. Modelling both scenarios is possible, but such attempts require different modelling approaches. Over the years, many taxonomic criteria have been proposed to categorise pedestrian models. In the literature review by Martinez-Gil et al. [5], multiple discriminating modelling criteria are introduced.

The most straightforward distinction is made by the scale of representation: simulations of a few pedestrians require a radically different approach than modelling of groups, crowds or large crowds. Similarly, a distinction can be made on models using a discretised space-time lattice, where other models may use a continuous approach. Another important discriminating criterion is the simulation scale, of which three different types are concerned: macroscopic, mesoscopic and microscopic approaches.

In the macroscopic approach, there is a strong emphasis on the modelling on dynamics of groups or crowds, omitting characteristics and interactions of individual pedestrians. In other words, pedestrians have on autonomy to change their kinematics or control interactions. Research is often modelled in terms of mean velocities, pedestrian densities, domain capacity and characteristic speeds. The opposite is true for the microscopic approach, in which every pedestrian is modelled as an independent entity: every individual controls their own dynamics and can produce local interactions such as collisions or overtakings. Due to the computational overhead, microscopic approaches are not suitable for simulations of large groups and crowds. The final approach is the mesoscopic approach, which is considered to be a hybrid of the aforementioned methods. Mesoscopic modelling considers individual pedestrians, but does not take individual interactions into account. The focus of this approach is on the modelling of group dynamics, but allows for more analysis on individual pedestrian behaviour. This population representation criterion is not perfect, as large differences on simulation techniques and model purposes exist within the same population representation.

Based on these criteria and other differentiating factors, Martinez-Gil et al. have categorised pedestrian modelling methods into five different groups. These categories are visualised in Figure 3.1.
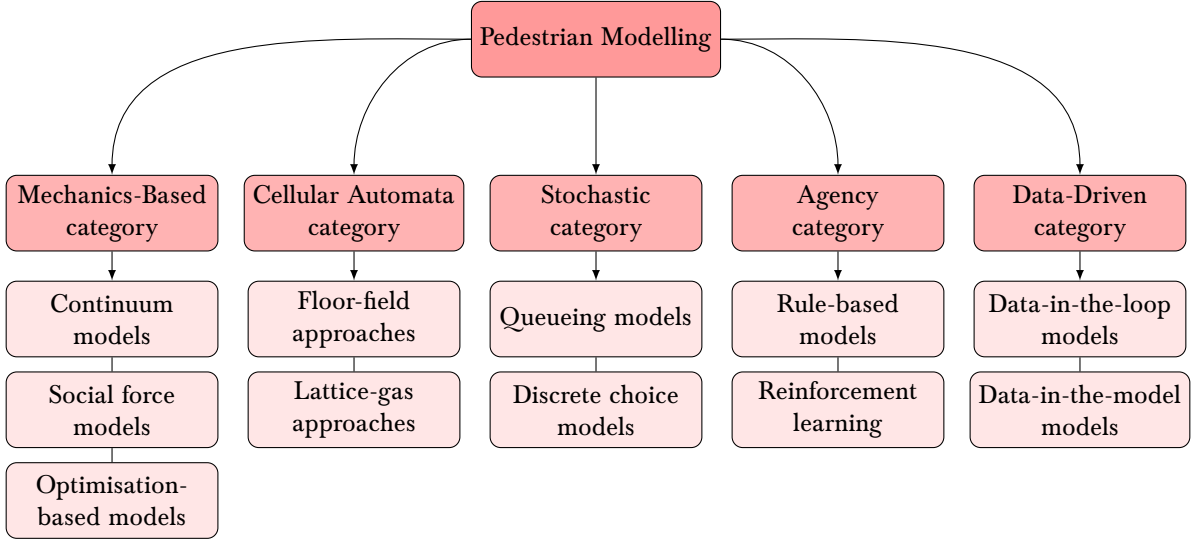
Figure 3.1: Overview of pedestrian modelling categories as created by Martinez-Gil et al. [5].

This categorisation is realised as follows:

- **Mechanics Based Models**
  Mechanics-based models are built upon models describing continuum mechanics or force models. Such models are often formulated by means of differential equations. Examples of mechanics-based models include continuum models, social force models and optimisation-based models.

- **Cellular Automata Models**
  Cellular automata are an idealisation of a physical system in which space and time are discrete, and physical quantities are limited to a finite set of values. Over the years, cellular automata have evolved from a computational paradigm to a sophisticated natural process modelling approach [13]. Cellular automata have proven to be very relevant for the modelling of pedestrian dynamics and the different approaches such as floor-field approaches and lattice-gas models are detailed in the following section.

- **Stochastic Models**
  The third category considers simulation models based on the theories of stochastic processes and random utility theory. Such models are not widely used in the terms of physics research, but very relevant for research on queueing processes and discrete-choice models. Cellular automata can be considered to be a special type of a stochastic model, but are a separate category due to its widespread use in pedestrian and traffic modelling.

- **Agency Models**
  In agency-based models, pedestrians are considered to be agents who are able to sense their environment and make decisions given their current state and navigation goal. Such models include rule-based and reinforcement learning-based approaches.

- **Data-Driven Models**
  The last category features all simulation models which are based on real pedestrian data. Data-driven approaches extract behaviour information from real-life data to build a model and steer individuals in simulations. Two variants exist, namely in-the-loop and in-the-model approaches, which are both detailed in Section 3.3

The following sections further detail the cellular automata and data-driven approaches, as these will be relevant for Chapter 5, where the novel simulation approach is introduced.

3.2   Cellular Automata Modelling

Cellular automata (CA) belong to the family of discretised modelling approaches. The model consists of a discrete spacetime lattice, along with computational capabilities that govern the evolution of the

model through space and time. CA approaches often feature a finite amount of physical states per lattice site, but this is not a requirement. CA models are characterised by two main features: locality, ensuring that interactions can only take place between a given set of neighbouring cells, and modularity, which requires every lattice cell to be an independent process. The latter renders CA approaches very suitable for parallelised computing.

In the context of pedestrian dynamics, CA models discretise the pedestrian domain into a grid of cells, where every cell hols information and the presence and walking direction of pedestrians. Cells can also be flagged to be not accessible, to model boundary conditions in the form of objects and obstacles. The model should also have a set of transition rules, governing pedestrian movements between different cells. Such rules are often defined by probabilities and stochastic choice models, hence the close connection with the stochastic modelling category.

Cellular automata were first applied successfully in the context of pedestrian dynamics by Blue and Adler in 1998 [14], simulating one-dimensional pedestrian traffic, which was later extended to two-dimensional traffic flows [15]. CA models have also been applied successfully in the context of evacuation problems [16][17][18] and in junction with with other modelling categories [19][20][21].

Much of the criticism towards CA-based approaches follows from the method's discrete nature. Since the space-time lattices are often very symmetric, the lattices are considered to be too symmetric for realistic movements. Moreover, the finite number of states and rules per lattice cell cause non-natural homogeneous behaviour, as demonstrated by Bierlaire et al. [22]. Approaches to overcome these limitations have been proposed by Lubas et al. [23], in which the authors created a non-homogeneous and asynchronous CA model with cell-dependent transition rules. Still, the CA model remains a popular platform for studying pedestrian dynamics following its computational simplicity.

The overview of Martinez-Gil et al. [5] creates a clear distinction in CA-based approaches, distinguishing between floor field approaches and lattice gas models. In the floor field approaches of Burstedde et al. [24] and Schadschneider et al. [25], every lattice cell stores the respective probabilities to transition from the considered cell to a neighbouring cell. These probability values can be calculated in many ways, e.g. to minimise the distance to a given destination or the expected travel time given by the Eikonal equation [26]. The floor field approach was further improved by Kirchner et al., introducing friction and clogging effects by distinguishing between competitive and cooperative movements [27][28].

The simplest floor field models consist of one or more static floor fields: the probabilities in the cells do not change for the duration of the simulation [29]. Other works propose more sophisticated superpositions of different floor field layers [30]. However, static fields often yield unrealistic effects. For example, pedestrians are unaware of congestion areas in the domain until they are very close. In dynamic floor field models, the transition probabilities can be modified during the simulation. In the work by Schadschneider et al. [31], the transition probabilities are a superposition of a static and dynamic part. The static probabilities are calculated by the domain layout and static objects present, whereas the dynamic probabilities are a function of other pedestrians' movements. For all floor field models, the chosen tessellation (space and/or time discretisation) is of great importance. To compensate for the orthogonal anisotropy present in square lattices, Leng et al. introduced a hexagonal lattice at the cost of computational simplicity [32].

Lattice gas approaches are considered to be a different category of cellular automata models. In such models, pedestrians are treated as biased random walkers. Again, the transition probabilities are defined between neighbouring cells, but these are modified by a drift parameter D, implementing an intensity of bias towards a preferred direction. This parameter depends on the topological configuration at a given time. Without this drift parameter, lattice gas models are very similar to the floor field approach. The first lattice gas models were introduced by Muramatsu et al. to study pedestrian counterflow simulations, but were used to simulate evacuations as well [33][34].

3.3 Data-Driven Modelling

The data-driven category distinguishes itself by a strong dependence on real-life measured pedestrian behaviour. In the literature review, two different approaches are concerned, namely data-in-the-loop approaches and data-in-the-model approaches.

In the data-in-the-loop models, real pedestrian data(consisting of group behaviour or individual trajectories) are assembled into a collection, which is then used to perform simulations. In Lerner et al. (2007), pedestrian trajectories are captured from video recordings, which are used to generate natural pedestrian behaviour in a virtual environment [35]. In the work by Porzycki (2014), a pedestrian simulation is coupled with a measurement setup, asdetected pedestrians are initialised as embodied

agents in the simulation [36]. In 2010, Ju et al. introduced a crowd generation approach, in which crowd formations and individual trajectories were taken from video recordings [37]. These measurements were then used to create virtual interpolated crowds of different densities. All data-in-the-loop approaches suffer from interpolation artefacts causing non-realistic behaviour, especially in the limit of high densities.

Data-in-the-model are similar to the aforementioned methods, but have one key difference in their workings: the parameters of an existing simulation model are adjusted based on real pedestrian measurements. This category has much more works reported in the literature review, for all modelling categories considered, such as mechanical, cellular automata and stochastic models [38][39][40]. The most travails are encountered in the area of data extraction: it is time-consuming work to capture high-quality pedestrian measurements, moreover for large crowds.

Chapter 4

Trajectory Acquisition Methods

In recent years, technological advancements have lead to automated and highly accurate methods for the acquisition of pedestrian trajectory data. In this chapter, an overview of trajectory acquisition methods is given. Section 4.1 provides an overview of the three-step process (recording, localisation and tracking) used in trajectory acquisition methods, and 4.2 gives a more extensive look at an acquisition approach used in this thesis, the height-augmented histogram of gradients (HA-HOG) method.

4.1 Overview of Trajectory Acquisition Methods

In general, all automated trajectory acquisition methods share the same three-step process of the recording, localisation and tracking of pedestrians. In the first step, human behaviour is captured using a recording device. This is followed by the localisation step in which the pedestrians' location – or additional features, such as orientation [41] – are detected on a frame-by-frame basis. Lastly, the movements of the pedestrians are tracked over multiple frames by a tracking algorithm, linking detections across time into tracks. This sequence of steps yields complete pedestrian trajectories through space and time. Hybrid trajectory acquisition methods exist as well, in which the steps of localisation and tracking are not completely separated. For example, in [42] a tracking framework is proposed, in which the localisation step is strengthened by a data association scheme based on predicted trajectories. In the following subsections, a detailed look into every acquisition step is provided.

4.1.1 Recording

To start the trajectory acquisition, pedestrian movements need to be captured on a recording device. The most obvious recording method is the usage of RGB video cameras, which was considered to be the standard for recording in early studies of pedestrian dynamics. Video based measurements easily allow human interpretation in terms of analysis of domain layout, pedestrian characteristics (age, gender) or other circumstances like lighting or weather conditions [43][44]. Since video cameras are ubiquitous in modern society, video based recordings are easy to acquire. The processing is of video recordings is however expensive both in terms of human time and processing time. Moreover, in light of recent privacy advancements, this method may also cause data protection and privacy regulation issues. This is not an issue for controlled pedestrian experiments, in which pedestrians are fitted with bright, recognisable attributes such as hats. Such experiments come at the cost of limited sample sizes and biased pedestrian behaviour, as subjects are often aware of being observed. Attempts have been made to use infrared sensors for pedestrian behaviour recording as well. Analysis of trajectory data by Kerridge et al. features recordings using a 16 x 16 array of pyroelectric ceramic detectors [45][46]. In this work, a detection accuracy of approximately 3 cm is achieved. At the time, infrared approaches used to have the main benefit of real-time processing, but advancements in computer processing power have enabled real-time processing for video recordings as well.

In recent years, good results have been achieved by overhead trajectory recordings using consumer-grade depth cameras. Pioneering work by Seer et al. [47] has shown excellent pedestrian recordings using relatively cheap Microsoft Kinect systems. These systems feature two cameras: one camera records a regular video stream, while the other camera records 3-dimensional depth image information. In their work, the authors developed an algorithm to record 3-dimensional information in the form of depth images, enabling automatic, high-quality and privacy-friendly trajectory measurements. Around the

same time, Brščič et al. [48] developed a similar method to track pedestrian movements, including position, orientation and height information. In their work multiple 3D range sensors are considered, including Microsoft Kinect systems.

This Kinect-based method has been adopted and further extended by Corbetta et al. at various experiment locations in The Netherlands to obtain large scale real-world pedestrian movement data [1][2][49][3]. In more recent work, depth image recordings have been made using an Orbbec Persee recording system [50]. The Orbbec Persee is a combination of a video camera, height camera and a computer for processing and storing of recordings. Where the Kinect approach purely focuses on the recording method, the Orbit Persee serves as an all-in-one solution for processing of measured recordings. Moreover, the Orbbec system has better sensors with higher recording frame rate and increased tracking height range. The included recording computer greatly increases the system's portability as well. The case study detailed in Chapter 6 features depth recordings made using the Orbbec Persee system as well.

4.1.2 Localisation

As recordings of pedestrian dynamics are obtained, it is non-trivial to detect the locations of pedestrians and track their position over time. In the early stages of pedestrian research, manual annotation of video camera recordings was the gold standard in pedestrian dynamics research. Manual annotation would also allow connotation of other pedestrian attributes, such as age or gender [47]. Manual annotation was later followed up by semi-automated annotation of video recordings, demonstrated in various works in the early 2000's [44][43].

Within the recent past, research has been focused on developing automatic, anonymous pedestrian detection algorithms [51]. Corbetta et al. have demonstrated a pedestrian detection algorithm based on a convolutional neural network, which is trained using synthetic training data [49]. Their work allows for good detection results, but is not suitable for situations with high pedestrian density. Moreover, it is non-trivial to generate realistic high-density depth images to train the network. In the Kinect-based recording methods introduced by Seer et al., and Corbetta et al., an automated localisation method based on random cloud clustering is used. The details of the detection method can be found in the respective papers, but the main workings are as follows.

Let $f^n = f^n(\vec{z})$ be the depth map recorded by Kinect at time instant $n \geq 0$ and at spatial position $\vec{z} = (x, y)$. The following successive steps are taken:

Depth-based significative foreground segmentation
Operation in which a common background B and objects smaller than the pedestrian height threshold limit h are removed from the depth maps. This results in a foreground depth map F^n.

Foreground random sampling
N random points are sampled from the foreground map, resulting in a sparse cloud representation of the foreground map, F_s^n.

Foreground random cloud clustering
The samples in the sparse foreground map F_s^n are agglomerated in clusters using a clustering algorithm. Given some characteristic human body scale size S as a clustering cutoff parameter, the measured points are clustered to form pedestrians.

Head identification
For every pedestrian, the exact position of the head is determined by identifying the set of points closest the camera. To make this distinction, this group of points is chosen such that they are closer than the 10th percentile of points corresponding to the pedestrian.

The clustering approach has shown to work well, accurately determining pedestrian's locations up to an accuracy of 1 cm [1][2][47]. In very densely populated crowds however, clustering-based approaches show substantial performance degradation and require large computational resources. Moreover, these approaches have no notion of pedestrian shape and cannot distinguish (multiple) pedestrians and/or objects under dense conditions. In recent years, a new localisation approach, based on a Histogram of Oriented Gradients (HOG) method, was introduced by Kroneman et al. [52]. This method allows for better detection performance under dense conditions and requires less computational resources. The details of this detection algorithm are discussed in Section 4.2.

4.1.3 Tracking

As the position of every pedestrian is known on a frame-by-frame basis, a tracking algorithm needs to link subsequent positions together to form trajectories. In this way, pedestrians are tracked as they propagate through the measurement domain. A straightforward approach is introduced by Seer et al. [47] in the form of nearest neighbour matching. Through linear extrapolation of previous positions, the tracking algorithm searches for nearest detection within the spatial and temporal threshold. A more elaborate tracking mechanism is introduced by Corbetta et al. [3]. In their work the open source OpenPTV library [53] is used, which is a tracking approach based on Particle Tracking Velocimetry (PTV), widely used in experimental fluid mechanics [54].

In recent years, promising results have been achieved using the Trackpy Python tracking library [55]. Trackpy is a package for tracking blob-like features in video images, following them through time, and analyzing their trajectories. It started from a Python implementation of the widely-used CrockerGrier algorithm [56] and is currently in transition towards a general-purpose Python tracking library. The measurement case study detailed in Chapter 6 makes use of the Trackpy tracking library as well.

4.2 Pedestrian Localisation using Height-Augmented HOG

An improved pedestrian detection algorithm was proposed by Kroneman et al. in 2018 [52], further improving the clustering algorithm introduced by Seer et al. In this work, a height-augmented Histogram of Oriented Gradients (HA-HOG) method is introduced, which enables more accurate tracking at high pedestrian densities at a lower computational cost. This localisation algorithm has been used in the Amsterdam case study, discussed in Chapter 6. In this section the details of capturing pedestrian motion using this HA-HOG approach are discussed.

4.2.1 Localisation Method

The HA-HOG method localisation algorithm consists of five steps, which are visualised in Figure 4.2. An extensive description of this algorithm is given in the work by Kroneman et al. [52]. These steps are as follows:
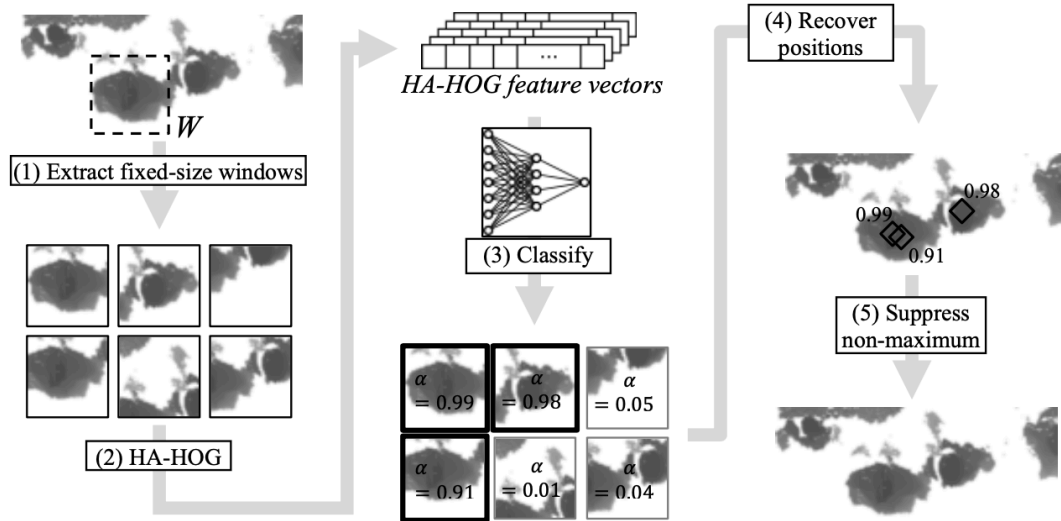
Figure 4.1: The five steps of the HA-HOG localisation algorithm: extract windows (1), calculate histograms (2), classify samples (3), recover positions (4) and suppress non-maxima (5). This figure has been reproduced from the paper by Kroneman et al. [52].

Extract Windows

To start, a fixed-size window is slid over some depth image I. In this way, samples W are generated, essentially being crops of the original image. The window size is taken slightly larger than the typical area covered by a single pedestrian. The window is displaced approximately five or six pixels per sliding movement, causing adjacent samples to have overlap.

Calculate HA-HOG Samples

The calculation of a HA-HOG histogram is a form of an image descriptor. In such cases, a summarised representation of an image is returned, focussing on information relevant for the use case. Usage of image descriptors greatly simplifies content classification as it requires less complex classification models. For every window W taken from I in the previous step, the HA-HOG descriptor is calculated:

1. Compute the gradient $\vec{G}$ of W using a second order finite difference method: $\vec{G} = \nabla W$;

2. Compute a polar representation of $\vec{G}$, such that:

$$\vec{G} = (G_x, G_y) = G_r \cdot \begin{pmatrix} \cos G_\phi \\ \sin G_\phi \end{pmatrix}; \tag{4.1}$$

3. Partition $\vec{G}$ into a grid of n small, non-overlapping cells. The cell size is equal to a factor of the window size, such that the remainder of the division is zero. For every cell, make a histogram of the gradient vectors, binned using G_ϕ and weighted by G_r, and perform L2-normalisation:

$$HOG_{cell} = \frac{Hist(G_\phi, G_r)}{|Hist(G_\phi, G_r)|}; \tag{4.2}$$

4. Concatenate all HOG cell descriptors to yield the HOG descriptor of the window W:

$$HOG(W) = [HOG_{cell}(c_1), HOG_{cell}(c_2), ..., HOG_{cell}(c_n)]; \tag{4.3}$$

5. Include a histogram of the depth values over W to the cell's HOG descriptor, allowing the classifier to avoid misclassification of objects with an overhead shape similar to a pedestrian:

$$HA - HOG(W) = [HOG(W), Hist(W_{heights})]; \tag{4.4}$$

Classify HA-HOG Samples

A neural network performs classification over all accumulated HA-HOG descriptors, assigning a score α, with $0 \leq \alpha \leq 1$, whether the sample is centred on a pedestrian.

Recover Positions from Samples

After this classification, a threshold ($\alpha \geq 0.9$) is applied to create a selection of pedestrian location candidates. For all location candidates, the center is considered to be the pedestrian's position, since the window size is approximately equal to a pedestrian's overhead area.

Suppress Non-Maximum Positions

Due to the small window stride distance, adjacent windows can lead to double detections, resulting in multiple positive samples for the same pedestrian. These duplicate values are removed using non-maximum suppression: if two detected pedestrian locations are closer to each other than a given threshold distance, the location with the lowest value of α is discarded. This process is repeated until all mutual pedestrian distances are sufficiently large. At last, the set of pedestrian positions localised within the measured depth image is returned.

4.2.2 Training and Performance

The HA-HOG method has been trained and benchmarked on two different measurement setups in Eindhoven and Leiden in The Netherlands. Training was performed through back-propagation to return a value for α, and was further strengthened by image augmentation through rotations and noise insertions. Additional training in the form of bootstrapping and human labeling of invalid predictions was performed as well. Parallel processing was enabled through the use of TensorFlow [57] and Keras [58].

In Figure 4.2, it can be seen that the HA-HOG outperforms both the HOG and clustering method in Eindhoven (a) and Leiden (b). This especially holds for dense conditions in which $\rho_{nn} > 2$. In this equation $\rho_{nn} = \frac{1}{\pi * r_{nn}^2}$, where r_{nn} is the distance between each manual annotation and its nearest neighbour.

In Chapter 6, the HA-HOG localisation method is applied to an unseen set of pedestrian depth images, without retraining for different lighting and recording conditions.
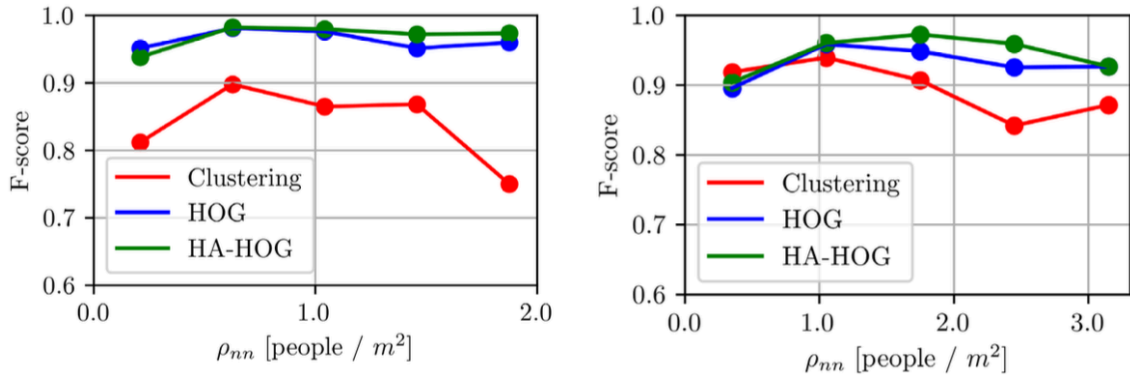
Figure 4.2: Localisation accuracy of different localisation algorithms. Performance measured in F-score values for the clustering algorithm (red), HOG algorithm (blue) and HA-HOG algorithm (red), for the measurement setups in Eindhoven (a) and Leiden (b). This figure has been reproduced from the paper by Kroneman et al. [52].

Part II

Lattice-Based Description of Pedestrian Dynamics

Chapter 5

Data-Driven Equilibrium Distribution Estimation

In this chapter a mesoscopic analysis framework for pedestrian motion is introduced, formulating an understanding from a Boltzmann-like perspective which is inspired by kinetic gas theory. Using large-scale pedestrian trajectory measurements, a data-driven description of the pedestrian motion in phase space is created. In Section 5.1, an introduction to concepts of kinetic theory and phase space is given and in Section 5.2 a brief derivation of the Boltzmann equation is provided. This is followed by Section 5.3, which gives an explanation how trajectory measurements are used create an estimation of the equilibrium probability distribution function in phase space. In Section 5.4, a double discretisation is applied to the phase space, creating a conditional description of the measured behaviour. To conclude, in Section 5.6, a microscopic simulation model is given, which uses the measured distribution functions of pedestrian motion. A perspective on data-driven mesoscopic simulation models is given as well.

5.1   Mesoscopic Modelling and Phase Space

As indicated in Chapter 3, various modelling approaches are used in the context of pedestrian dynamics. From a physical perspective, a group of pedestrians travelling a domain is quite similar to the flow of a gas or a fluid. Therefore, the theory of fluid dynamics can often be applied in the context of pedestrian dynamics as well. In general, three different modelling approaches exist within the context of fluid dynamics [59].

The first approach is the macroscopic approach, which ignores the fact that matter is made up of individual molecules. Instead, following the continuum theory, fluids and gasses are described as continuous blobs of matter. By conservation of mass, momentum and energy, the continuity, Navier-Stokes and energy equations are derived. The macroscopic description allows for usage of tangible quantities, such as density, velocity and temperature. The second approach is the microscopic modelling approach, which considers the behaviour and evolution of every particle present in the fluid or gas. By considering the every particle in the flow, the macroscopic properties such as density, pressure and temperature can be calculated. However, in the limit of a large number of particles, it is impossible to model every particle independently. This problem is solved by the third approach: the mesoscopic kinetic theory.

In the mesoscopic kinetic theory, the behaviour of a fluid or gas is described by the evolution of collections of similar particles. In such a collection, particles which have a similar position $\vec{x}$ and velocity $\vec{\xi}$ are grouped together. In this way, the dynamics of the flow can be described without taking every individual particle's position and velocity into account, but by only using the amount of particles belonging to a small volume element in coordinate and velocity space. The aforementioned collection is known as the particle distribution function $f(\vec{x}, \vec{\xi}, t)$, which can be seen as a generalisation of the particle density ρ while also taking the particle velocity $\vec{\xi}$ into account. In this way, the distribution function represents the density of mass in both the coordinate space and the velocity space. This high-dimensional space is known as the phase space: the space which is spanned by all state variables of a system. Natural systems consist of infinitely many states, but for the point particles considered in this approach, its state is completely defined by its position and velocity. This state is represented by the vector $\vec{q} = (\vec{x}, \vec{\xi})$.

An example illustrating the phase space concept is a point particle moving in a one-dimensional space. For this system, every possible state of the particle is defined using only two variables, the position x and velocity ξ. To illustrate how a particle progresses through phase space as a function of time, it is possible to draw an arc in the phase space diagram. A typical phase space diagram for a one-dimensional point particle is shown in Figure 5.1.
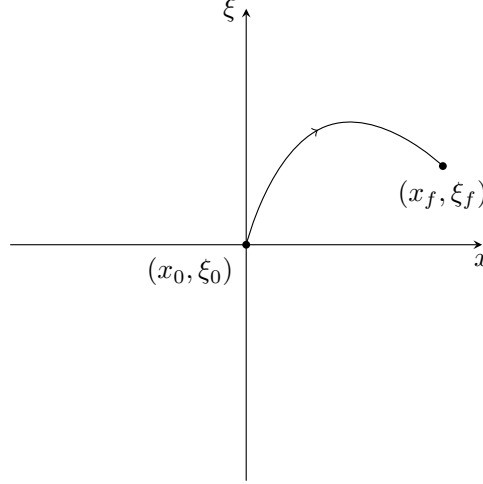
Figure 5.1: Example of a phase space diagram of a one-dimensional particle moving through space. The phase-space trajectory for a one-dimensional particle moving from x_0 to x_f is displayed. The horizontal axis shows the particle's position, while the vertical axis displays the particle's velocity while performing the movement, ranging from ξ_0 to ξ_f. By following the curve from the initial point to the final point, the time-dependent behaviour is visualised.

The generalisation towards a point particle moving in three dimensions is straightforward. If a particle is located at position $\vec{x} = (x, y, z)$ with velocity $\vec{\xi} = (\xi_x, \xi_y, \xi_z)$, its location in the phase space is represented by the vector $\vec{q} = (\vec{x}, \vec{\xi})$. Hence, the phase space is now a six-dimensional space, spanned by the unit vectors $(\hat{x}, \hat{y}, \hat{z}, \hat{\xi}_x, \hat{\xi}_y, \hat{\xi}_z)$. Again, the particle can be parametrised by the time variable t, allowing the particle to change its location in the phase space as it propagates through time. Other properties, such as a possible particle charge Q, can be included in the phase space vector $\vec{q}$ as well, further increasing the dimensionality of the phase space.

As a mesoscopic approach, kinetic theory describes the evolution of collections of particles that are close to each other in phase space. For a system consisting of N point particles with identical mass m, a differential volume element in phase space is considered:

$$\mathrm{d}\vec{q} = \mathrm{d}^3\vec{x}\,\mathrm{d}^3\vec{\xi}. \tag{5.1}$$

As it is guaranteed that every particle resides somewhere in phase space, the amount of particles present in the volume element is given by the particle distribution function:

$$\mathrm{d}N = f(\vec{x}, \vec{\xi}, t)\,\mathrm{d}^3\vec{x}\,\mathrm{d}^3\vec{\xi}. \tag{5.2}$$

The total amount of particles is recovered through integration over the entire phase space:

$$N = \int \mathrm{d}N = \int f(x, y, z, \xi_x, \xi_y, \xi_z, t)\,\mathrm{d}^3\vec{x}\,\mathrm{d}^3\vec{\xi}. \tag{5.3}$$

This naturally raises the question what such a distribution of $f(\vec{x}, \vec{\xi}, t)$ looks like, and how it evolves through space and time. The evolution of $f(\vec{x}, \vec{\xi}, t)$ is governed by the Boltzmann equation, and is derived in the following section. A brief derivation is given, but a more thorough analysis can be found in [59].

5.2   The Boltzmann Equation

By considering all particles present in the flow, it is possible to distinguish random thermal motion of the particles from the bulk motion of the flow. This is done defining the relative velocity $\vec{v}$:

$$\vec{v}(\vec{x}, t) = \vec{\xi}(\vec{x}, t) - \vec{u}(\vec{x}, t), \tag{5.4}$$

where $\vec{\xi}(\vec{x}, t)$ is a single particle's velocity and $\vec{u}(\vec{x}, t)$ is the bulk velocity of the flow. When the flow has been left alone for a sufficient of time, an equilibrium particle distribution $f^{eq}(\vec{x}, \vec{v}, t)$ will be reached, which is isotropic in velocity space around $\vec{\xi} = \vec{u}$. Assuming a separable form solution (see [59]), it can be found that this equilibrium distribution is given by:

$$f^{eq}(\vec{x}, \vec{v}, t) = \rho \left(\frac{1}{2\pi RT} \right)^{3/2} e^{-|\vec{v}|^2/(2RT)}. \tag{5.5}$$

This equilibrium distribution can be used to understand how the particle distribution function evolves in the following way. As the distribution function f is a function of space, velocity and time, the total derivative follows from the chain rule:

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \left(\frac{\partial f}{\partial t} \right) \frac{\mathrm{d}t}{\mathrm{d}t} + \left(\frac{\partial f}{\partial x_\beta} \right) \frac{\mathrm{d}x_\beta}{\mathrm{d}t} + \left(\frac{\partial f}{\partial v_\beta} \right) \frac{\mathrm{d}v_\beta}{\mathrm{d}t}. \tag{5.6}$$

By using the notation $\Omega(f)$ for the total derivative $\frac{\mathrm{d}f}{\mathrm{d}t}$ and Newton's second law, this equation can be written as:

$$\frac{\partial f}{\partial t} + \xi_\beta \frac{\partial f}{\partial x_\beta} + \frac{F_\beta}{\rho} \frac{\partial f}{\partial \xi_\beta} = \Omega(f), \tag{5.7}$$

which is the Boltzmann equation. In this equation, the Greek index is used to denote an arbitrary component of a vector. Summation over repeated indicies is implied as well. The first two terms represent the particle distributions being advected with the velocity ξ of its particles, and the term the represents the effects of forces acting on the particles. The $\Omega(f)$-term on the right hand side is known as the source term, which makes for a local redistribution of f due to collisions. Therefore, this term is known as the collision operator. In particle physics, the BGK collision approximation is generally used, which is given by:

$$\Omega(f) = -\frac{1}{\tau} \left(f - f^{eq} \right). \tag{5.8}$$

In this approximation, it is assumed that collisions are responsible for the relaxation of any distribution f towards the equilibrium distribution f^{eq}. The speed of this relaxation is given by the relaxation time τ. Through conservation requirements and a first-order approximation, it can be proven that Boltzmann equation is able to describe the macroscopic behaviour of a flow, governed by the corresponding Euler and Navier-Stokes equation.

The previously introduced concepts can be applied in the context of pedestrian dynamics as well. As pedestrians on a flat surface show no variation in the z-dimension, the phase space is reduced in its dimensionality. Since a pedestrian has a certain domain position $\vec{x} = (x, y)$ and velocity $\vec{v} = (u, v)$ at a given time t, the four-dimensional phase space vector is given by $\vec{q} = (x, y, u, v)$. Similar to the point particle progressing through phase space as a function of time, the same would hold for a pedestrian moving through a two-dimensional domain.

5.3   Data-Driven Estimation of the Equilibrium Distribution

Following the similarity between particle physics and pedestrian dynamics, the mesoscopic approach will now be applied to an arbitrary pedestrian flow. At a given moment of time, a pedestrian flow can be in any random state. The amount of pedestrians could be high or low, and pedestrians could be at every location with any given velocity. However, as illustrated in the derivation of the Boltzmann equation, after waiting for a sufficiently long time, an equilibrium distribution will be reached. Deviations from this equilibrium are possible, but these are expected to diminish through relaxation in the BGK approximation. Following this assumption, is therefore expected that on average, the distribution of pedestrians in phase space is equal to the equilibrium pedestrian distribution function.

For example, consider a one-dimensional flow of pedestrians on the domain D. The domain D has a length L, ranging from $x = 0$ to $x = 2$. All pedestrians walk with a mean velocity of $u = 1$ in the $\hat{x}$-direction, although a small variance of velocities is present. Therefore, in phase space, most pedestrians will be located at $u = 1$, being uniformly distributed in coordinate space. The resulting average pedestrian density function in phase space is visualised in Figure 5.2. Following the aforementioned argument, it is assumed that the measured average distribution of positions and velocities (as illustrated in Figure 5.2) is equal to the equilibrium pedestrian distribution $f^{eq}(x, u)$. Since the distribution of f is normalised, the value of f at the point (x, u) can be interpreted as the probability to find a pedestrian with velocity v at position x.
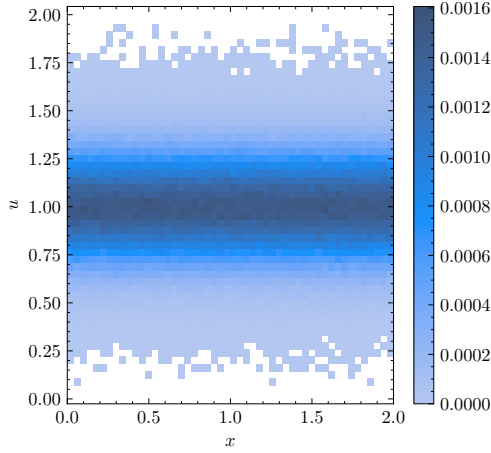
Figure 5.2: Pedestrian density function in phase space for a one-dimensional flow. The domain ranges from $x = 0$ to $x = 2$, and the pedestrian walking velocity is normally distributed with $\mu = 1$ and $\sigma = 0.2$. The colour bar indicates the local pedestrian density in phase space. The total volume under the density surface is normalised to unity.

In this thesis, this expectation is put into practice in the following way. Using a pedestrian tracking approach as discussed in Chapter 4, a real-life large-scale measurement of a pedestrian flow is performed. This measurement approach results in a set of pedestrian trajectories, in which the positions of pedestrians are tracked over time as they move through the measurement domain. By considering the difference between subsequent position measurements, the pedestrian velocity is calculated as well. Using this combination of position and velocity data, a description of the measured pedestrian flow in phase space can be generated. By averaging over all measured positions and velocities, the equilibrium pedestrian density distribution in phase space $f^{eq}(\vec{x}, \vec{u})$, can be calculated.

5.4   Spacetime Discretisation

In order to perform the data-driven estimation of the pedestrian density function, a discretisation is applied to the set of measured trajectories. Using this discretisation, the measured pedestrian positions (in pixels, metres or any other spatial unit) or mapped to a discretised spacetime grid using a spatial sampling resolution $\Delta\vec{x}$ and temporal sampling rate $f_s = \frac{1}{\Delta t}$. In this way, the discretisation gives control over the desired accuracy and amount of computational resources required for the analysis. Furthermore, this enables the usage of lattice Boltzmann-based simulation methods (LBM), which will be addressed later on. Using this discretisation, all pedestrian locations are defined on a two-dimensional integer grid, sized (DX,DY). Time is discretised as well, resulting in an integer time grid ranging from 0 to t_{max}. Hence, every physical position $\vec{x}$ is mapped to a lattice cell (x_i, y_i), and every point in time t is mapped to a lattice time t_i. The discretisation of velocity space is not a free parameter, but follows from the chosen spacetime discretisation. Since every trajectory on the spacetime consists of subsequent positions on the (integer) spacetime grid, every possible velocity is always an integer multiple of steps on the spacetime grid. Therefore, the smallest possible velocity on the spacetime grid is given by:

$$\Delta v = \frac{\Delta x}{\Delta t}, \tag{5.9}$$

and larger velocities are integer multiples of Δv. From here, this discretisation will be referred to as the primary discretisation: the discretised spacetime lattice to which all pedestrian movements are mapped.

The primary discretisation is followed by a second discretisation of the coordinate space, which is necessary for the conditional histogram binning performed in the next section. This secondary discretisation is performed using a coarser sampling resolution Δx_c, which is an integer multiple of the primary discretisation resolution Δx:

$$\Delta x_c = n\Delta x, \text{ where } n \geq 2. \tag{5.10}$$

Hence, a secondary coordinate grid is created, which is coarser than the primary discretisation. Using this system, the primary coordinate grid consists of cells indexed by (x_i, y_i) and the second coordinate grid is indexed by pairs of (x_c, y_c) values. This double space discretisation is illustrated in Figure 5.3.

26
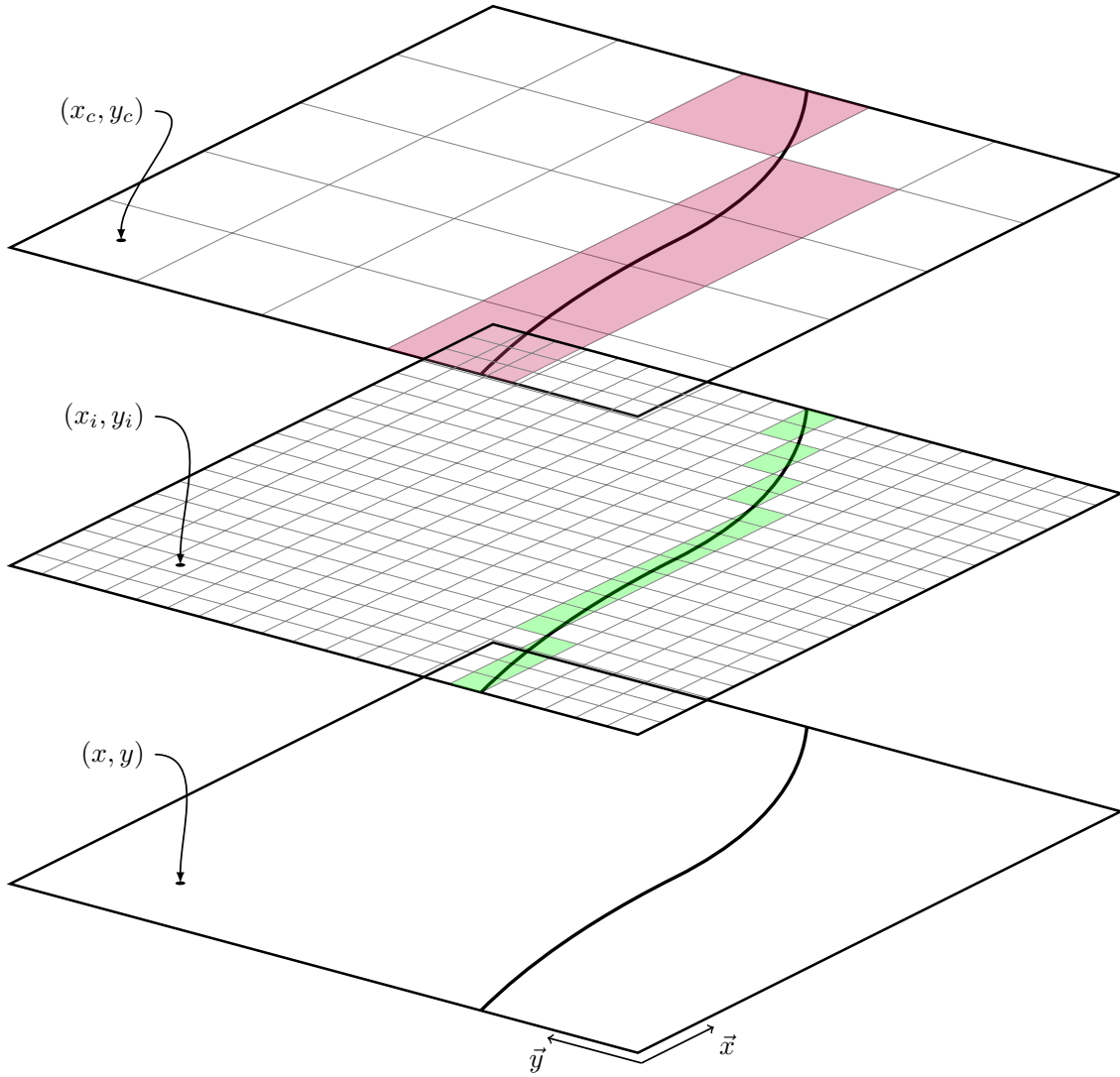
Figure 5.3: Illustration of the double discretisation of the coordinate space. The bottom layer illustrates a single trajectory measurement traversing through a continuous space as a function of time. The middle layer shows the trajectory in the primary discretised form. The trajectory is no longer continuous, but exists as a sequence of grid positions illustrated by the green cells. In this primary grid, positions are discretised as cells, which are indexed using a combination of two integers (x_i, y_i). The top layer illustrates the secondary discretisation, in which by using a larger sampling resolution Δx_c, a coarser grid is created. In this grid, individual cells are indexed using a different set of integers (x_c, y_c).

Using this secondary discretisation, groups of adjacent positions are collected together to form a coarse position grid. The pedestrian velocity grid is discretised twice as well, creating groups of similar velocities. This double velocity discretisation is shown in Figure 5.4.

The double discretisation of the position and velocity space serves the following purpose. By considering a specific coarse grid cell (x_c, y_c), it is possible to examine pedestrian behaviour in a specific part of the domain, as the average behaviour can significantly vary between different parts. Similarly, the double velocity discretisation allows for the distinction between pedestrian moving in different directions, e.g. upward versus downward. As the primary grid is retained, a finer analysis of positions and velocities is still possible. The double discretisation of position and velocity space is extensively used in the next section as well.
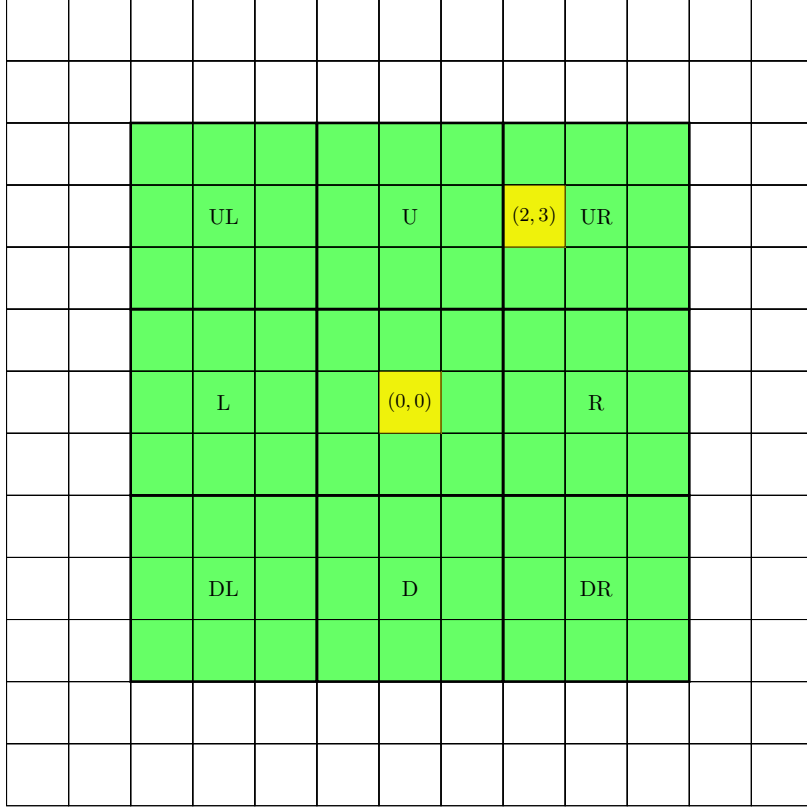
Figure 5.4: Illustration showing the double discretisation of the velocity space. Every grid cell is a possible velocity a pedestrian can have in the discretised space. The center element $(0,0)$ displays the standstill direction in which a pedestrian does not move. The cells indicated in green, ranging from $(-4,-4)$ to $(+4,+4)$, indicate the set of velocities a pedestrian is allowed to have in the discretised space. The second discretisation is used to group similar velocities together, creating a coarse velocity description system. In this figure, sets of 9 velocities are assembled in the coarse discretisation. In this way, the sets of velocity groups v_c (UL, U, ..., D, DR) are formed. For example, all velocities $\vec{v} = (\geq 2, \geq 2)$ are considered to be upward right (UR) velocities.

5.5   Conditional Estimation of Equilibrium Distributions

As the measured pedestrian data is now discretised in phase space, every trajectory measurement now consists of a sequence of observations in phase space. Every observation corresponds with a point (x_i, y_i) on the primary spacetime lattice and a primary velocity (u_i, v_i) on the velocity lattice. Additionally for every observation, a coarser grid position (x_c, y_c) and a coarser velocity (u_c, v_c) are known as well.

In general, it is to be expected that a measurement of a pedestrian flow consists of a superposition of multiple smaller flows. For example, in a bidirectional pedestrian flow, a distinction between pedestrians travelling in either direction can be made. Hence, when considering the equilibrium distribution function for a bidirectional flow, two distinct peaks in velocity space (corresponding with the two different flows) are to be expected.

In this section, using the secondary discretisation lattice, a conditional estimation of the estimation of equilibrium distributions is introduced. By considering pedestrian observations that belong to a specific cell on the secondary lattice, a conditional estimation of the equilibrium distribution can be made. This is possible in three different ways:

1. By considering different parts of the domain in coordinate space, the pedestrian density distribution can be estimated on a small part of the domain. By selecting one specific coarse grid position (x_c, y_c), the corresponding pedestrian observations on the primary lattice can be returned, providing a local estimation of the pedestrian density function. In this way, it can be assessed whether the equilibrium distribution shows spatial variation over the measurement domain.

2. A conditional estimation can be applied to velocity space as well. If a pedestrian is measured to have a certain velocity (u_i, v_i), than under the assumption of a small time difference Δt, the next velocity measurement is very likely to be found in the direct vicinity of the current velocity as well. By considering only observations belonging to a specific coarse velocity (u_c, v_c) at at previous moment in time, it is possible to calculate the density equilibrium distribution for the group of pedestrians moving in a certain coarse direction.

3. A third conditional possibility is a combination of the previous two approaches. By only considering observations corresponding to a specific part of the domain (x_c, y_c) and belonging to a certain coarse velocity in a previous instant of time (u_c, v_c), it allows for the calculation of equilibrium density distributions under these two conditions.

In this way, the primary discretisation allows for measurements to be discretised at a high resolution, preserving measurement accuracy, while the secondary discretisation allows for a conditional grouping of observations. Using this approach, the calculation of equilibrium density distributions can be performed for specific parts of the domain, for pedestrians moving in a specific coarse direction, or a combination of the two.

5.6 Trajectory Simulation using Estimated Probability Distributions

A Monte Carlo-based microscopic simulation model has been developed in the form of a biased random walker model. The details of this simulation model are discussed in this section. By using the same space discretisation used in the measurement processing framework $(\vec{x}_i, \vec{x}_c, \vec{u}_c, \vec{x}_c)$, a virtual pedestrian domain is generated. Again, all possible positions and velocities are are accessed by integer numbers on a two-dimensional spacetime grid. Using the inverse discretisation $(\vec{x} = \vec{x}_i / \Delta \vec{x})$ is mapped to a lattice cell (x_i, y_i)), all discretised units can be converted back to physical units. Pedestrian movements are generated through a random sampling approach. Every timestep, a random velocity is generated, after which the pedestrian takes on the sampled velocity to move to a new location. The velocities are sampled from the measured equilibrium density functions, which are approximated by fitting a Gaussian distribution.

A virtual pedestrian is able to move from one grid cell to another by sampling a random velocities. Consider that a pedestrian is present in a certain point in phase space: $\vec{q}_i = (x_i, y_i, u_i, v_i)$. The next position in space is trivial, as the new position is accessed by the sum of the previous position and velocity: $(x_{i+1}, y_{i+1}) = (x_i + u_i, y_i + v_i)$. After arriving at this new grid position, a new velocity is determined as well. This is done by taking a random sample from the measured probability distribution function in phase space. Using the coarse grid position (x_d, y_d) corresponding to (x_i, y_i) and the coarse arrival velocity VG corresponding to (u_i, v_i), the measured probability distribution function can be looked up from the dataset. Next, the exit velocity is randomly sampled from this distribution. After this, the pedestrian moves to its new position in space and the process is repeated again. If a pedestrian reaches an edge of the domain, the trajectory is terminated, not discarded. An overview of the used simulation algorithm is provided in Algorithm 1.

Currently, the model only allows for simulation of individual pedestrians, without taking their interactions into account. This implies that multiple pedestrians can occupy the same grid cell and that their movements are not dependent on the presence and movements of other pedestrians. There are numerous possibilities for improvements and extensions in this part of the modelling approach. However, future work could fully utilise the similarities between the lattice-based measurement processing and the Lattice Boltzmann simulation Method (LBM), allowing more sophisticated usage of the measured distributions.

Algorithm 1: Microscopic simulation of artificial pedestrians

---

Input: PEDS, STEPS
Output: Dataset of simulated trajectories
Data: Domain dimensions, measured velocity PDFs
/* Initialise (x,y,u,v) array: */
1 data = empty_array(PEDS, STEPS, 4)
2

/* Initialise simulation   */
3 $x(t=0)$ = x_{meas}
4 $v(t=0)$ = v_{meas}
5

/* Perform iteration   */
6 **for** *i in range(0,PEDS)* **do**
7   **for** *t in range (1,STEPS)* **do**
8   Stream pedestrians: $x_t = x_{t-1} + v_{t-1}$
9   **if** x_t *violates boundary conditions* **then**
10   **break**
11   **else**
12   Sample new velocity:
13   v_t =randomvel(x_t, v_{t-1})

/* Finalisation   */
14 Write trajectory dataset to disk
15 Perform comparison with measured trajectories
16

Part III

Case Study: Amsterdam Red Light District

Chapter 6

Measurements

This chapter describes how a sizeable real-life dataset of pedestrian trajectories, as used in the previous chapter, has been obtained for a measurement site in Amsterdam, the Netherlands. In a narrow alley of the popular red-light district De Wallen, pedestrian movements were anonymously recorded using an overhead depth image camera. The details of this measurement location and the experimental setup are discussed in Section 6.1. In Section 6.2, a statistical analysis of the measured trajectory dataset is given. In Chapter 7, the equilibrium distribution functions of the measured trajectories are provided, closely following the proposed method in Chapter 5. A review of the performance of the microscopic simulation model is given in Chapter 8.

6.1   Experimental Setup

All pedestrian behaviour is gathered from a single measurement location in Amsterdam. In cooperation with the municipality of Amsterdam, this specific measurement location was chosen as the municipality was interested in the data as well, since the De Wallen district has been subject to extreme crowdedness in recent years [60]. An overhead measurement setup was installed in the middle of the Stoofsteeg alley, at the approximate location of Stoofsteeg 6A, 1012 EC Amsterdam, GPS (52.37252, 4.89735). The measurement location and the recorded ground area are illustrated in Figure 6.1. At the measurement site, the alley has a width of 3 metres. The overhead height camera was installed at an approximate height of 4 metres, and the major recording axis was chosen orthogonal with respect to the alley axis. Many stores and bars are located in the alley, with their windows and entrances located at the edge of the walking area. Measurements of pedestrian behaviour took place in the evenings (18:00 until 00:00) of the first seven days of September 2019.

As introduced in Chapter 4, the recordings of pedestrian movements were performed using an Orbbec Persee recording system [50], which is a combination of a video camera, height camera and a computer for processing and storing of recordings. This way, the system serves as an all-in-one portable solution for the recording and processing of trajectories. Height images were recorded using at a resolution of 320 x 240 px, at a frame rate of 15 fps. In combination with the installed height installation, this resulted in a measured walkable ground area of 2.20 m x 2.20 m where trajectories could be recorded. In total, this results in a measurement area of 4.84 m^2. In Figure 6.1, the Orbbec Persee is visible as a grey box directly above the center of the alley. The recorded ground area is illustrated in red.

From the recorded height images, the localisation of pedestrians was performed using the height-augmented HOG (HA-HOG) algorithm and the pedestrians were tracked over multiple frames using the OpenPTV library. The localisation algorithm was trained using measurement data from a different measurement setup [3], without retraining for this specific measurement location. Using this tracking approach, every pedestrian was assigned a unique pedestrian identifier (PID).

A typical recorded height image is shown in Figure 6.2. In this image, a single recorded frame is visible, wherein measured pedestrians (gray) are set off against a neutral white background. In this frame, four pedestrians are present are being tracked by the processing algorithm. The current estimated pedestrian location is indicated by a small white square and the sequence of previous locations are indicated by the black trail behind the pedestrian. In the selected frame, two measurement errors are visible on the right side as well. Two particles are being tracked but from the shape of the image it can be seen that no pedestrian is present at this location and this point in time.

Figure 6.1: Measurement setup at Stoofsteeg, Amsterdam. The Orbbec Persee camera system is visible in the picture as a grey box. From this box, a green pyramid-like beam is emitted, illustrating the measurement range of the camera system. The red area on the ground illustrates the measured surface area.

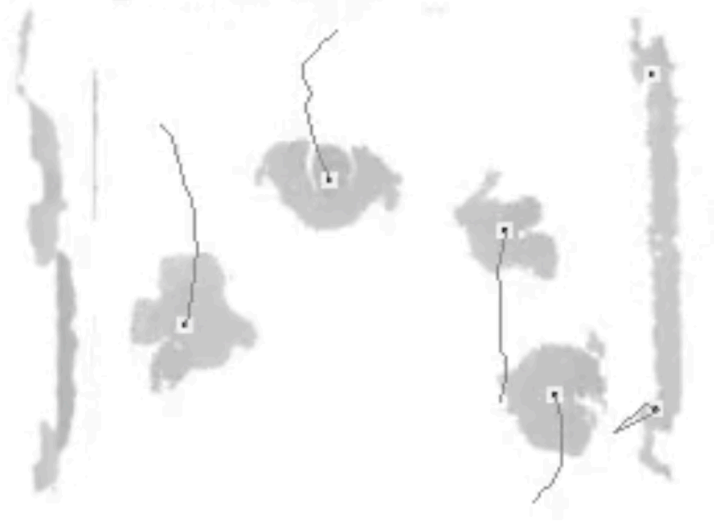**Figure 6.2**: Height image recorded by the Orbbec Persee camera system. Pedestrians, shown in grey, are visible against a white background. The estimated pedestrian locations are indicated by the black dots, and the sequence of previous locations are indicated by the black trail behind the pedestrians. Two measurement errors are visible on the right, and the measurement date and time is shown in the top left corner.

6.2 Measured Trajectories

6.2.1 Dataset Quality

By the localisation and tracking of pedestrians over the entire sequence of depth images, a large trajectory dataset was created. In total, 120,856 trajectories were recorded at the measurement location, consisting of 8,506,881 separate pedestrian location measurements. After inspection, it became apparent that the dataset suffered from a substantial number of measurement errors. This was due to reflections from the windows at two closed edges of the domain, along with the fact that the model was not retrained for this specific use case.

In order to remove these measurement errors from the dataset, a data cleaning model was set up and trained for this specific purpose. Firstly, for every trajectory, the instantaneous velocity and acceleration per measurement was calculated using a Savitzky-Golay filter. Using this information, a descriptive summary for every trajectory was calculated, including the total trajectory path length (m) and total time present (s). Moreover, the distribution of position, velocity and acceleration measurements was taken into account as well. Using this information, a classifier would be able to distinguish valid measurements (a pedestrian crossing the domain at a typical walking velocity) from obvious measurement errors (little to no movement, very long presence in domain).

A binary random forest classification model was set up in Python using Scikit-Learn to perform this task. The classifier was given 500 labeled instances of valid measurements and obvious measurement errors. These labelled examples were manually labeled by a human worker. The model performance was assessed using a five-fold cross-validation approach, leading to a predictive accuracy of 87%. Using this approach, 77.34% of all location data was to be discarded, and a reduction of 59.03% measured trajectories was realised. The details of the dataset reduction is given in Table 6.1. The complete cross-validation score of the trajectory classifier is given in Table 6.2.

Table 6.1: Removal of measurement errors leading to a strong reduction of the dataset.

	Original Dataset	Reduced Dataset
Measurements	8,506,881	1,927,795
Trajectories	120,856	49,514
Fraction Upward	44.52%	39.34%
Fraction Downward	55.48%	60.66%

Table 6.2: Performance of the random forest binary classification model developed which was used to discard invalid measurements. The model performance was assessed using five-fold cross-validation in Scikit-Learn.

Accuracy	Precision	Recall	F1
0.87	0.88	0.90	0.88

Additionally, an active learning framework was implemented to allow future work to easily label new trajectories for the classification model. This framework was developed using modAL [61], a modular active learning framework for Python.

6.2.2 Description of Measured Behaviour

After the data cleansing filtering process, the measured pedestrian behaviour was assessed. As expected from the measurement location, a strongly bidirectional flow was observed. A random sample of 500 measured trajectories is displayed in Figure 6.3. In the figure and throughout the remainder of this work, a Cartesian coordinate system is applied. The horizontal direction ($\hat{x}$) is defined as the width of the alley, while the vertical direction ($\hat{y}$) is aligned with the major axis of the alley, connecting both entrances. As expected, all pedestrians were found to primarily move in the vertical direction, traversing through the alley. In total, 63% of the trajectories were downward trajectories and 37% were upward trajectories). In Figure 6.3, upward trajectories are indicated in green, while downward trajectories are coloured blue. This figure clearly shows the measured bidirectional pedestrian flow, following the local (Dutch) convention to walk on the right-hand side of the road. Moreover, it can be seen that total pedestrian

flow is composed of a superposition of multiple flows. On the left part of the domain, the flow consists mostly of pedestrians walking downward, while on the right side pedestrians walk upward. In the center of the domain a transition area is present, where an equal number of downward and upward movements has been recorded.
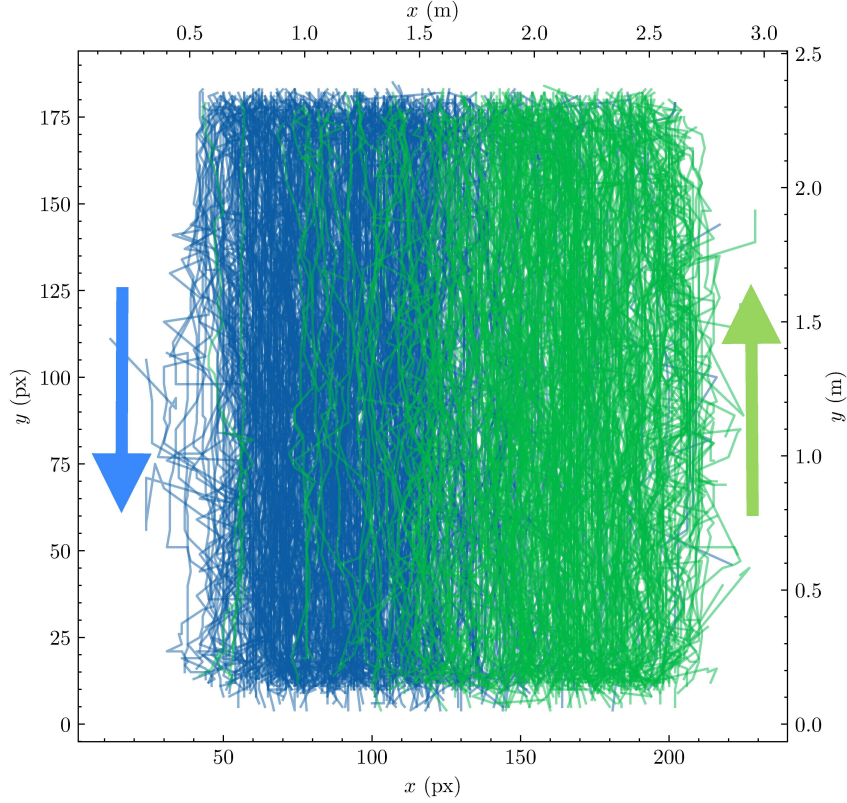
Figure 6.3: Random sample of 500 measured trajectories. Pedestrians travelling upward in the $\hat{y}$-direction are illustrated in green and pedestrians travelling downward in the $-\hat{y}$-direction are coloured blue.

An analysis of the measured velocities of the pedestrians walking through the alley is given in Figure 6.4. On a logarithmic scale, the measured velocities are binned in a histogram. The vertical velocity distribution clearly show two distinct peaks, approximately $-0.85\,\mathrm{m\,s^{-1}}$ and $0.75\,\mathrm{m\,s^{-1}}$, corresponding with the measured upward and downward flows. As the average pedestrian walking speed is generally modelled at $1.2\,\mathrm{m\,s^{-1}}$ [62], the measured velocities are rather low. This can be explained by the characteristics of the measurement domain, as the narrow alley in combination with the bidirectional flow suppresses the pedestrian walking speed. Moreover, the shops and bars present in the alley lead to strolling and gazing pedestrians. The horizontal velocity distribution shows a single peak around $0\,\mathrm{m\,s^{-1}}$. Moreover, it can be seen that variance of horizontal velocities is 9i000 ooooplplopllplower than the variance in the vertical direction. Again, this is probably due to the fact that the alley is rather narrow, prohibiting strong horizontal deviations from the bidirectional flow.

A description the measured pedestrian densities is given in Figure 6.5, in which two different density features are illustrated. In Figure 6.5a it can be seen that the measured area is quite busy, as approximately 75% of all measured trajectories were recorded when another pedestrian was also present in the domain. As the measured domain area is equal to $4.84\,\mathrm{m^2}$, it is seen that the number of pedestrians per area often exceeds $1\,\mathrm{pedestrian/m^2}$. In Figure 6.5b, the temporal distribution of the number of measured pedestrians is given. As can be seen, there is a large variance in the measured densities present. The highest pedestrian densities were measured between 22:00 and 23:00.
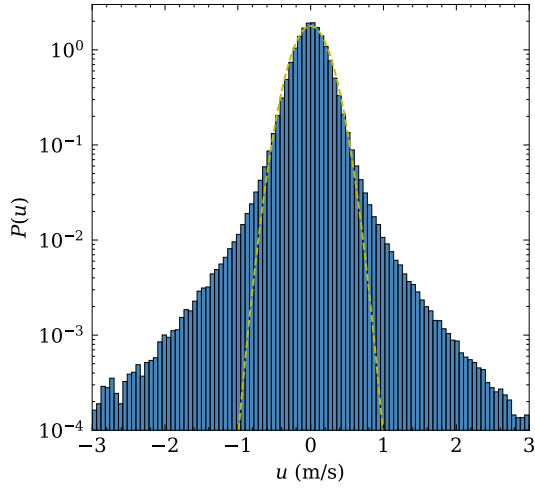
(a) Measured probability density function of horizontal pedestrian velocities, illustrated in blue. The histogram is normalised to unity. In green, a Gaussian distribution ($\mu = 0$, $\sigma = 0.22$) is visualised for comparative purposes.
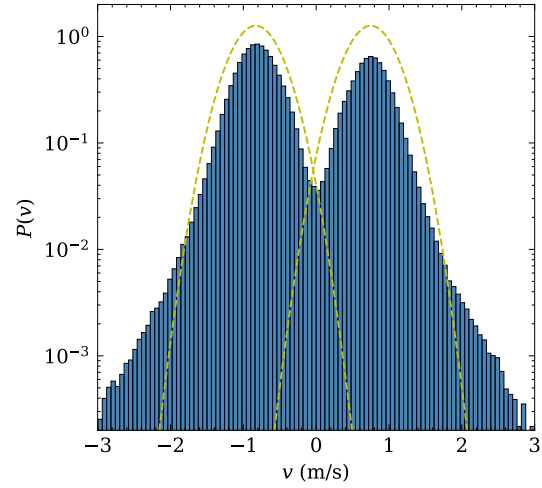
(b) Measured probability density function of vertical pedestrian velocities, illustrated in blue. The histogram is normalised to unity. In green, two Gaussian distributions ($\mu_1 = -0.83$, $\sigma_1 = 0.32$, $\mu_2 = 0.75$, $\sigma_2 = 0.32$) are visualised for comparative purposes.

Figure 6.4: Measured distributions of horizontal and vertical velocity in the Amsterdam measurement data. While the horizontal distribution is strongly peaked around $u = 0\,\mathrm{m\,s^{-1}}$, two distinct peaks are visible in the vertical distribution: around $-0.85\,\mathrm{m\,s^{-1}}$ and $0.75\,\mathrm{m\,s^{-1}}$. These peaks correspond with the two distinct populations of pedestrians that are present in the measurements: the persons walking upward and downward. Both distributions show good similarities with the fitted Gaussian distributions around the corresponding means. The deviations in the tails of the horizontal distribution are most probably measurement errors, as it is very unlikely to have such velocities in a domain that is $2.20\,\mathrm{m}$ wide. The deviations in the vertical distribution can be measurement errors as well, or these can be caused by other forms of transportation (running, skating, biking).
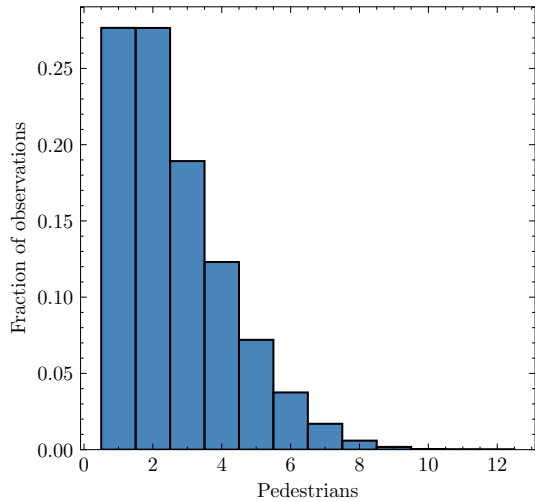
(a) Histogram of the non-zero number of pedestrians present in the measurement domain. On the vertical axis the fraction of the total number of observations is given.
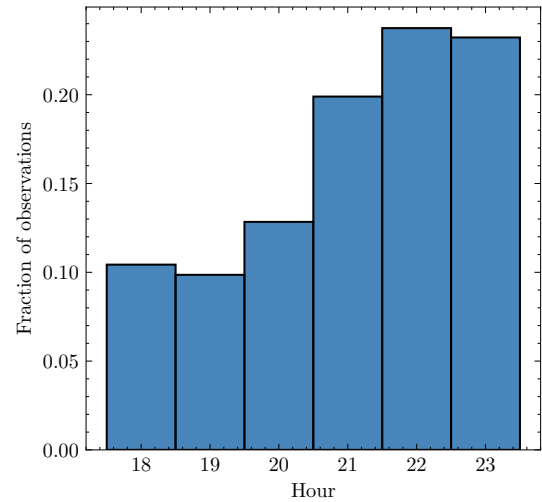
(b) Histogram of the measured pedestrian density as a function of time in the evening. On the vertical axis the fraction of the total number of observations is given.

Figure 6.5: Overview of the measured pedestrian density in the alley in Amsterdam, the Netherlands. The domain can be considered to be rather busy, as more than 70% of the trajectories are recorded when another pedestrian is present in the domain. Moreover, the measured pedestrian density strongly varies with time.

Chapter 7

Measured Phase Space Probability Distributions

In this chapter, the data-driven estimation of the equilibrium distribution functions is applied to the Amsterdam trajectory dataset (Chapter 6). The first section details the used spacetime discretisation to which all trajectories are mapped. Following Chapter 5, a second discretisation is applied as well. In Section 7.2, visualisations of the measured equilibrium distribution functions are provided. A conditional approach is used, in which the distributions are calculated for different parts of the pedestrian phase space.

7.1   Spacetime Discretisation

As detailed in Chapter 5, all measured pedestrian positions are mapped to a discrete lattice. For the measurements in Amsterdam, a spatial lattice of 234 by 186 cells was chosen. This lattice features the same aspect ratio as the recorded height images, resulting in square lattice cells. The temporal sampling resolution was taken to be equal to the sampling rate of the measurement setup. This resulted in the following primary discretisation:

- $\Delta x = \Delta y = 0.0130\,\mathrm{m}$

- $\Delta t = 1/f_s = 1/15\,\mathrm{Hz} = 0.0667\,\mathrm{s}$

- $\Delta v = 0.1956\,\mathrm{m\,s^{-1}}$

- $v_{max} = 12$, resulting in a 25 x 25 transition lattice

Using the secondary discretisation, adjacent cells in position and coordinated space were grouped together. The coordinate space was coarsely discretised using 3 rows and 7 columns, while a 5x5 discretisation was used for the coarse velocity space. Therefore, the total discretisation is as follows:

Table 7.1: Overview of the double spacetime discretisation applied to the dataset of pedestrian trajectories.

	Primary	Secondary
x	[0,1,...,234]	[0,1,...,6]
y	[0,1,...,186]	[0,1,...,2]
u	[-12,-11,...,12]	[-2,-1,...,2]
v	[-12,-11,...,12]	[-2,-1,...,2]

7.2   Measured Equilibrium Distribution Functions

The measured equilibrium distribution functions for the pedestrian positions and velocities are shown in Figure 7.1. Additionally, the secondary discretisation is shown as an overlay in Figure 7.2.
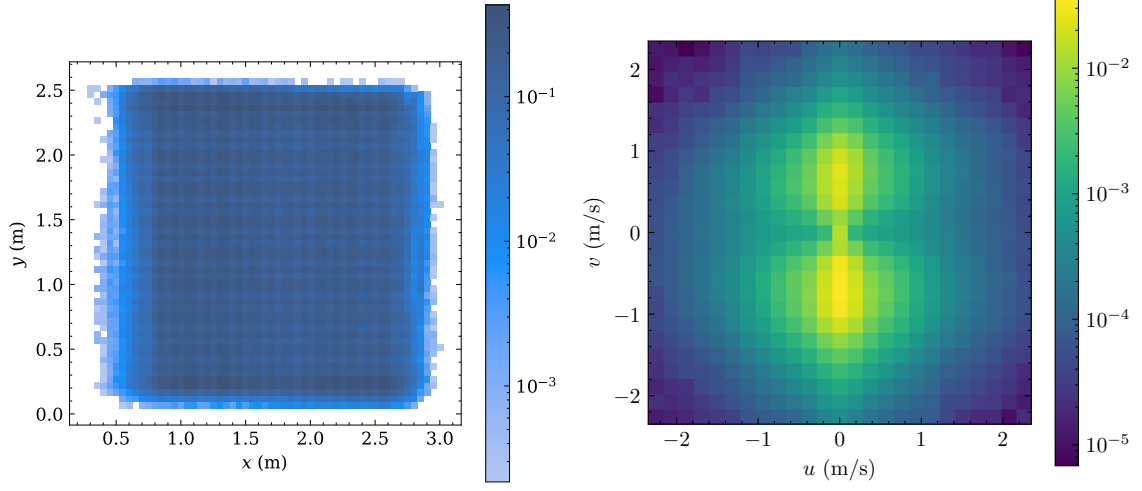
37

(a) Probability distribution function of measured pedestrian positions.

(b) Probability distribution function of measured pedestrian velocities.

Figure 7.1: Measured probability distributions for pedestrian positions and velocities.
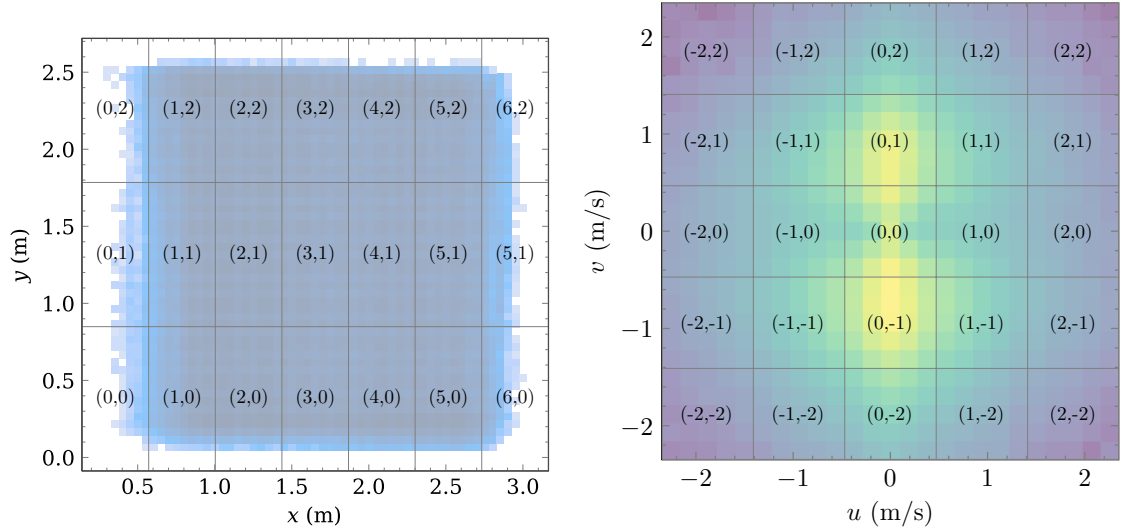
(a) Probability distribution function of measured pedestrian positions.

(b) Probability distribution function of measured pedestrian velocities.

Figure 7.2: Discretisation of phase space, illustrating subdomains and directions.

7.3 Conditional Estimation of Equilibrium Distribution Functions

The measured conditional equilibrium distribution functions are displayed in Figure 7.3 and Figure 7.4. In the former, the measured distributions are shown as a function of the pedestrian's previous coarse velocity $\vec{u}(t-1)$, while in the latter the measured velocity distribution function is given as a function of the coarse position $\vec{x}_c$.
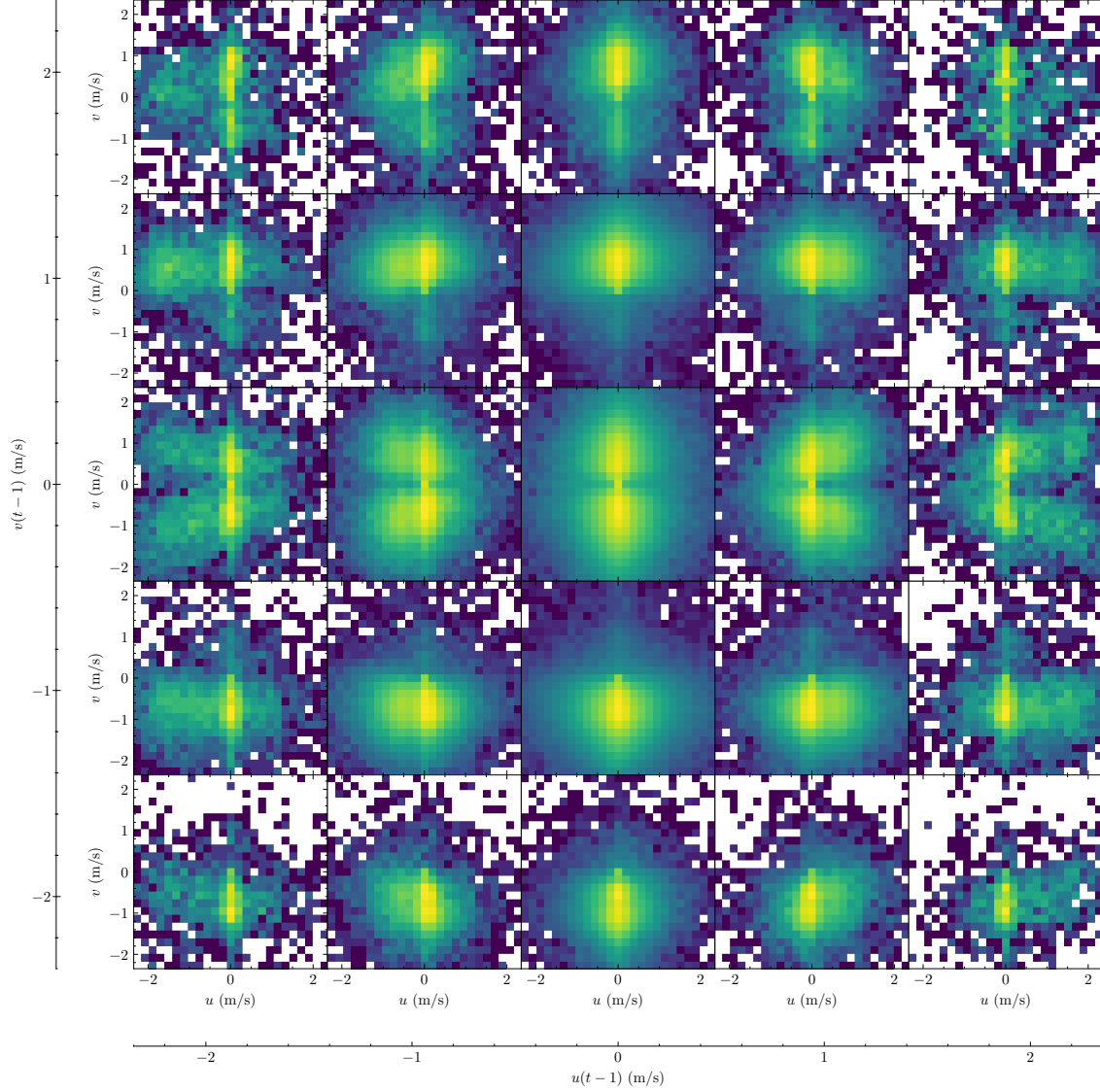
Figure 7.3: Velocity probability distributions (25x25) as a function of previous velocity (5x5). Every distribution is coloured independently on a logarithmic scale. This figure serves a qualitative description, for this reason the colourbar per distribution has been omitted.
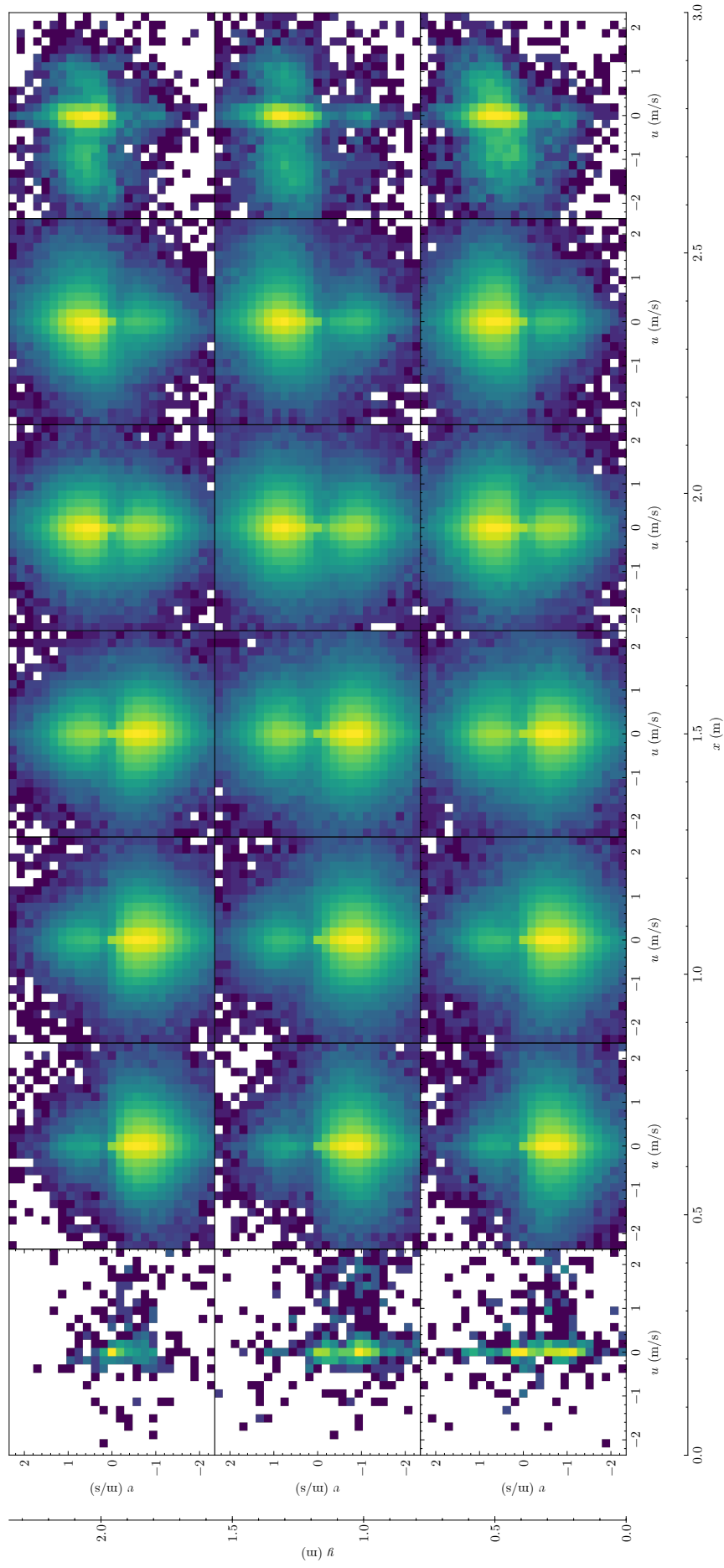
Figure 7.4: Velocity probability distribution functions for every subdomain part. Every distribution is coloured independently on a logarithmic scale. This figure serves a qualitative description, for this reason the colourbar per distribution has been omitted.

40

Chapter 8

Generation of Artificial Trajectories

In this chapter, artificial trajectories are generated using the Amsterdam trajectory data, following the microscopic simulation method proposed in Section 5.6. This chapter aims to illustrate that the proposed simulation method is able to generate realistic synthetic trajectory data. As the Amsterdam dataset features three distinct types of flows (as discussed in Section 8.1), the assessment is performed for each individual flow type. A simulation assessment of the complete flow is provided as well. In Section 8.2, a comparison metric is introduced which is used to quantify the performance of the simulation model. This metric consists of five elements, namely: (1) trajectory shape, (2) distribution of velocities, (3) distribution of positions, (4) horizontal difference distribution, and (5) the Lagrangian second-order structure function. The simulation performance assessments are given in Section 8.3. This chapter provides an in-depth look into the performance of the simulation model, while Chapter 9 is used to give a more overall look into the approach in terms of generalisability and applicability.

8.1 Assessment of Different Pedestrian Flows

As illustrated in Chapter 6, the Amsterdam measurement dataset features a clear bidirectional flow. From the overhead perspective, people walk downwards on the left side of the domain and people walk upwards on the right side of the domain. This corresponds with the local (Dutch) law of walking and driving on the right-hand side of the road. From the measurements and phase space discretisation, it can be seen that three different areas within the measurement domain can be distinguished:

1. On the left side of the domain ($0 \leq x \leq 1.55$ m), 81% of the measured trajectories, beginning in this part of the domain, consist of people walking from the top to bottom the domain. Hence, these people are following the local agreement to walk on the right side of the road. 19% perform the opposite behaviour, walking from the bottom to the top of the domain. Due to this strong difference, the pedestrian flow in this part of the domain is considered to be a unidirectionally downward flow.

2. The same argument holds for the right side of the domain ($1.80 \leq x \leq 3.00$ m), where 77% of the measured trajectories in this part of the domain walk upward, while only 23% of the measured persons walk downward. Hence, the flow is again considered to be unidirectional, now in the opposite direction.

3. A more complicated behaviour arises in the center part of the domain ($1.55 \leq x \leq 1.80$ m). In this area, an equal amount of upward and downward trajectories has been measured, hence forming a true bidirectional flow.

   In order to perform a complete validation of the simulation results, both the unidirectional flows (1&2) and the bidirectional flow (3) will be validated independently. By individually validating that every flow type can be simulated correctly, it is possible to perform a validation of a simulation of the entire domain as well.

   The validation of the unidirectional flows serves as a base scale for the simulation model. This allows for validation of simple pedestrian motion, in which a single pedestrian crossing the domain goes along with the flow. The validation of the bidirectional flow is more challenging, as the simulation model needs to account for two different pedestrian velocities in the same part of the domain. Especially for

the Amsterdam data, the simulation model needs to be able to distinguish between pedestrians moving upward and downward (cf. Figure 6.4b), ensuring the pedestrians are able to maintain momentum when traversing through the domain.

In order to make a thorough comparison between the measured and simulated trajectories, a data-based simulation initialisation is used. For every measured trajectory, the position x and velocity v at its first appearance are noted. In the simulation model, an equal amount of virtual pedestrians are initialised using these measured initial positions and velocities. The virtual pedestrians are then propagated using the simulation model, creating a dataset of artificial trajectories. Using the comparison metric detailed in the following section, the similarity of the measured and generated trajectory dataset is assessed.

Following these arguments, four different simulations are performed, which are all individually validated in Sections 8.3.1 until 8.3.4:

I. Unidirectional downward flow: pedestrians entering the domain in the top left area ($0.00 \le x \le 1.55\,\mathrm{m}$, $2.25\,\mathrm{m} \le y \le 2.55\,\mathrm{m}$) are selected for simulation.

II. Unidirectional upward flow: pedestrians entering the domain in the bottom left area ($1.80 \le x \le 3.00\,\mathrm{m}$, $0\,\mathrm{m} \le y \le 0.30\,\mathrm{m}$) are selected for simulation.

III. Bidirectional flow: pedestrians entering the domain in the middle of the domain ($1.55 \le x \le 1.80\,\mathrm{m}$), both at top ($2.25\,\mathrm{m} \le y \le 2.55\,\mathrm{m}$) and the bottom ($0.00\,\mathrm{m} \le y \le 0.30\,\mathrm{m}$) are selected for simulation.

IV. Complete flow: no restrictions are based on the pedestrian initialisation conditions.

In Figure 8.1, an overview of the four performed simulations is given.
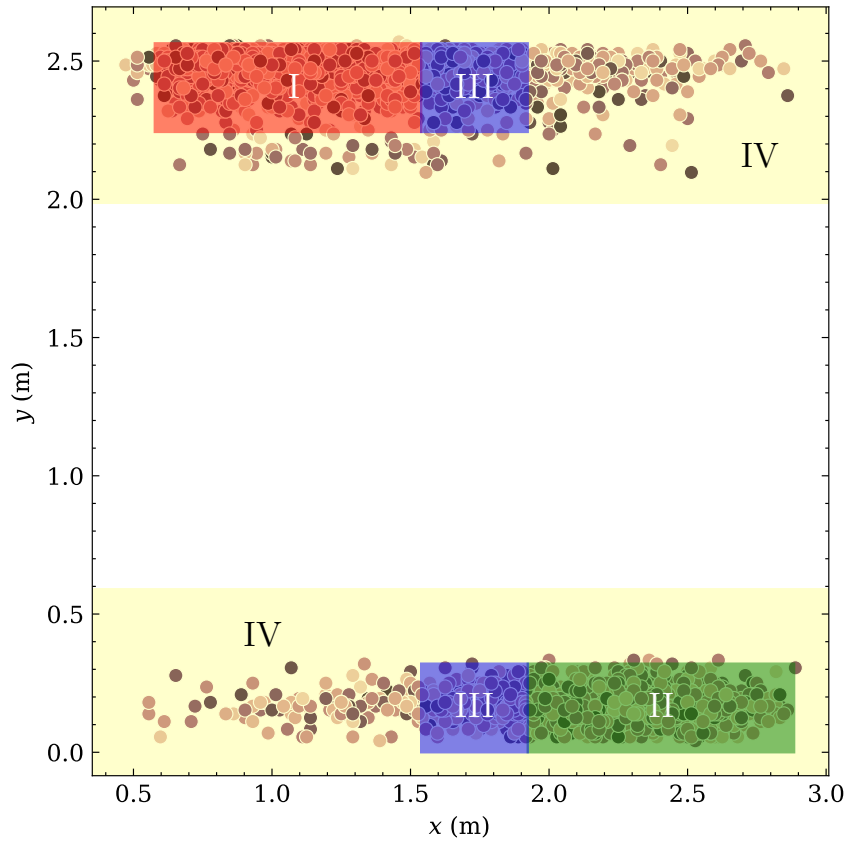
Figure 8.1: Overview of the measurement and simulation domain. The individual dots represent measured initial positions of pedestrians. The coloured areas: red (I), green (II), blue (III) and yellow (IV) correspond with the four performed simulations.

8.2 Comparison Metric

In order to quantify the comparison between the measured and generated pedestrian data, the similarity of trajectory data is analysed using a comparison metric. This metric consists of five quantities and is based on a comparison of distributions. In this way, random fluctuations are omitted from the comparison. The following five quantities are included in the comparison metric:

1. **Trajectory shapes**: a simple visual overhead comparison of trajectory shapes, providing a simple qualitative comparison method. This allows for verification if pedestrians show similar behaviour.

2. **Velocity distribution**: a two-dimensional histogram of all pedestrian velocities is generated. In addition to numerical comparison (mean, covariance), this allows for a complete review of similarity of pedestrian velocities.

3. **Position distribution**: a similar method is applied for the distribution of pedestrian positions: a two-dimensional histogram gives insight into the spacial distribution of positions.

4. **Horizontal difference distribution**: as the Amsterdam measurements feature a vertical bidirectional flow, it is of interest to consider the horizontal difference between a pedestrian's initial and final position. This is done by considering the difference d between the initial and final position: $d = \sqrt{(x_i - x_f)^2}$, where x_i and x_f respectively denote the initial and final position, which is visualised in Figure 8.2. Again, a comparison is made using histogram distributions.

5. **Lagrangian structure function**: lastly, for every trajectory its second-order Lagrangian structure function is calculated, a concept taken from turbulence physics. The equation is given by: $S_p(\tau) = \langle (\delta_\tau x)^p \rangle = \langle [x(t + \tau) - x(t)]^p \rangle$, in which x denotes a single position component, τ represents a variable time difference and p represents the order of the structure function. By analysing the structure functions, it allows for an analysis of the motion scales present in the domain. For two-dimensional diffusive motion ($x = \sqrt{2Dt}$), a linear relation for the structure function is expected: $S_x^2 \sim \tau$. Similarly, for ballistic motion ($x = vt$), a quadratic relation $S_x^2 \sim \tau^2$ is expected.

Figure 8.2: Illustration of the horizontal difference parameter, which is defined by: $d = \sqrt{(x_i - x_f)^2}$, where x_i and x_f respectively denote the trajectory's initial and final position. In this figure, two trajectories displaying upward and downward motion are included. The horizontal difference between the two positions are illustrated by the black arrow d.

8.3 Comparison of Flow Types

8.3.1 Unidirectional Downward Flow (I)
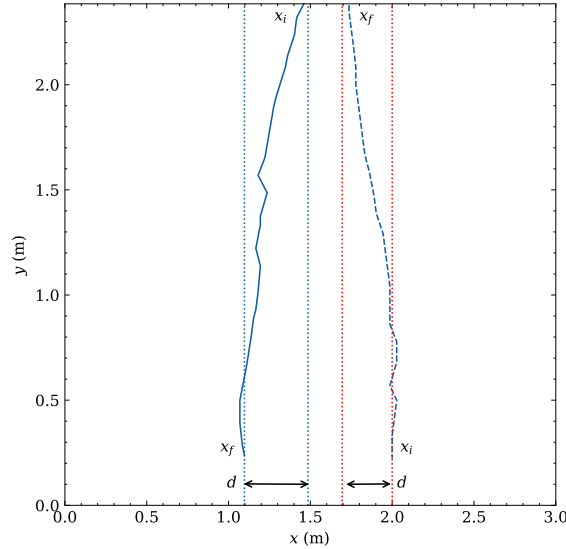
The first validation is performed on the left part of the domain, where the flow is considered to be approximately unidirectional. To initialise pedestrians, the following trajectory selection criteria are used:

$$t = 0 : \ 0 \le x \le 120 \text{ and } 160 \le y \le 190.$$

From the measured trajectory dataset, xxx trajectories are found to meet these conditions. An identical amount of trajectories has been generated using the simulation model. Every simulated trajectory has been initialised using a measured initial appearance. Using the aforementioned comparison metric, the following results are created.

In Figure 8.3, a visual comparison between measured and simulated trajectories is given. Both datasets show identical behaviour, as pedestrians enter the domain in the top left corner and navigate towards the bottom of the domain. Although the movements are mainly longitudinal, some transversal movement is visible as well.

(a) Measured trajectories. **(b)** Simulated trajectories.
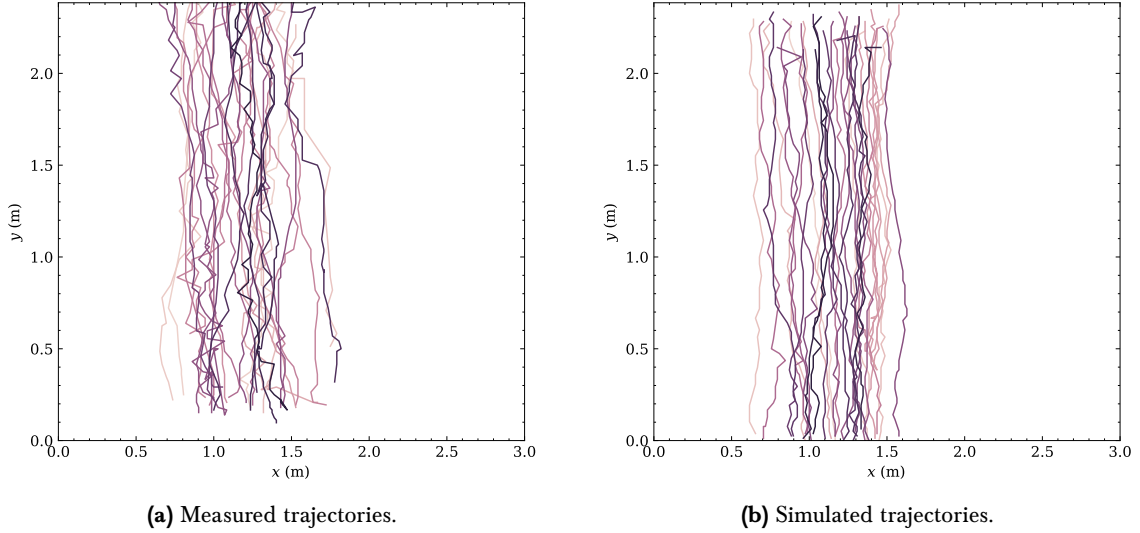
Figure 8.3: Illustrative comparison between measured and simulated trajectories. In both figures, 25 random trajectories have been selected from the respective datasets. It can be seen that the overal shapes of the trajectories is very similar, displaying clear unidirectional downward motion.

A two-dimensional of histogram of the measured pedestrian positions is given in Figure8.4, in which it can be seen that both datasets show a very similar spatial occupancy of the domain. One difference is visible in the right half the domain ($x > 1.5 \, \text{m}$). Some pedestrians are measured to have left the domain at the lower right edge of the domain, deviating from the bulk downward flow, and this behaviour is not reconstructed using the simulation model.

In Figure 8.5, the velocity distributions for both datasets is displayed. More differences arise here: the velocity distributions as seen in the simulation clearly follow normal distributions, but this does not hold for the distribution of u in the measured trajectories. Most probably, this occurs due to the fact that the domain is rather narrow: the top and bottom edges ($y = 0 \, \text{m}$) and ($y = 2.4 \, \text{m}$) are open, whereas the left and right edges are closed. Due to this, pedestrians are not able to walk freely, leading to a normal distribution that features a higher peak around $u = 0$.

In Figure 8.6, the distributions of the horizontal difference d are given, in which a clear difference between the measurement and the simulation is visible. Although the means of both distributions match, the variance of the measured distribution of d is higher than its simulated counterpart. Since the horizontal velocity distribution (Figure 8.5 is similar, it can be concluded that the simulation model is able to include horizontal variance within the trajectories itself, but does not create diagonally walking pedestrians very well. This can also be seen in Figure 8.4, where the position distribution is more narrow for the measured trajectories.

(a) Measured trajectories.

(b) Simulated trajectories.

Figure 8.4: Comparison between measured and simulated position distributions. A two-dimensional histogram binning is performed, visualising the area of the domain that is covered by the pedestrians. It can be seen that the beam is more narrow for the simulated pedestrians, and that the measurements include a few more outliers which cross the domain diagonally.

(a) Measured trajectories.

(b) Simulated trajectories.

Figure 8.5: Comparison between measured and simulated velocity distributions. Although both peaks are visible at the same location, a small difference visible in the width of the distributions. This is especially true for the distribution of the measured horizontal velocities, which does not follow a Gaussian distribution.

The final element of the comparison metric is the Lagrangian structure function S_x^2, which is given for the transversal ($\hat{x}$) and the longitudinal ($\hat{y}$) direction. As indicated in the comparison metric, the slope of the structure function on a log-log scale is an indication of the measured pedestrian motion. In the horizontal direction, both the measured and the simulated structure function show a linear slope ($S_x^2 \sim \tau$) on a log-log plot. The linear slope implies that the measured motion consists of diffusive motion. For the longitudinal motion, the slope is doubled, indicating a quadratic relation ($S_x^2 \sim \tau^2$) on the logarithmic axes. This slope is a clear indication of ballistic motion in the longitudinal direction, as is expected from the measured pedestrian behaviour. In short, the structure functions indicate that both the diffusive and ballistic motion are reconstructed correctly by the simulation model.

45

Figure 8.6: Comparison between the distributions of horizontal trajectory difference d in the measured and simulated trajectories. Again, the distributions clearly show an identical mean of $\approx 0.10\,\mathrm{m}$, but the measured trajectories show a much larger variance.
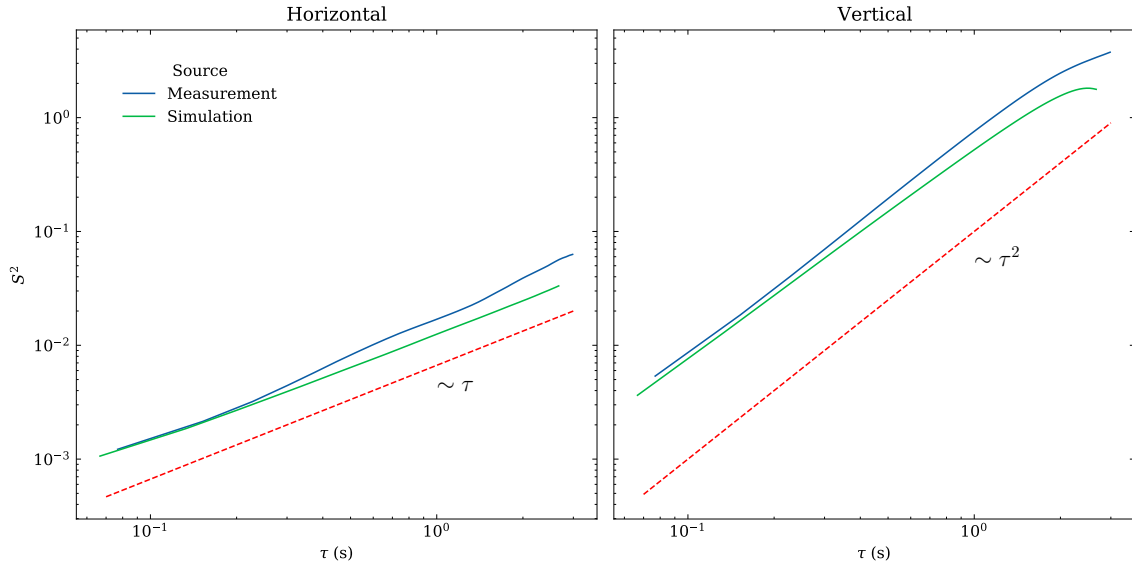
Figure 8.7: Comparison of the Lagrangian structure function S_x^2 for the measured and simulated trajectories. Both the transversal ($\hat{x}$) and longitudinal ($\hat{y}$) components are given for both datasets. A clear difference is visible in the slope of both plots: the transversal shows a slope of ≈ 1 indicating diffusive motion, while the longitudinal motion is ballistic, following the slope of ≈ 2.

8.3.2 Unidirectional Upward Flow (II)

The second validation is performed on the right side of the domain, where the flow is again considered to be approximately unidirectional. As the comparison is very similar to the unidirectional downward flow, a detailed discussion for every figure is omitted. However, an overall review of the similarity is given. The following trajectory initialisation criteria are used:

$$t = 0: \ 0 \le x \le 120 \text{ and } 160 \le y \le 190.$$

In the following figures, it can be seen that the simulation model is able to correctly generate trajectories that match the measured upward behaviour.

(a) Measured trajectories.　　　　　　　　**(b)** Simulated trajectories.
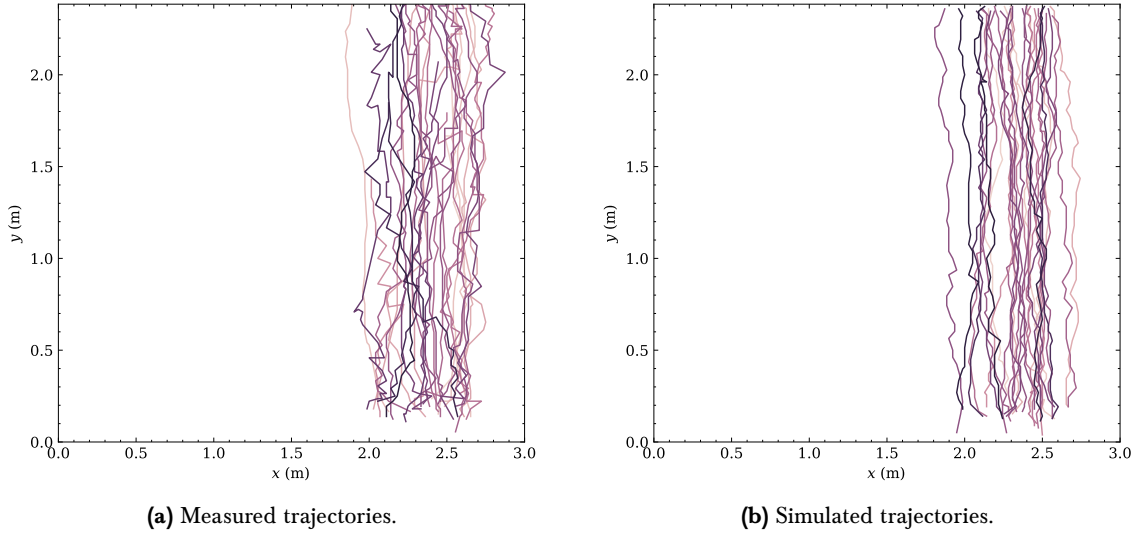
Figure 8.8: Illustrative comparison between measured and simulated trajectories. In both figures, 25 random trajectories have been selected from the respective datasets. It can be seen that the overal shapes of the trajectories is very similar, displaying clear unidirectional upward motion.

(a) Measured trajectories.　　　　　　　　**(b)** Simulated trajectories.
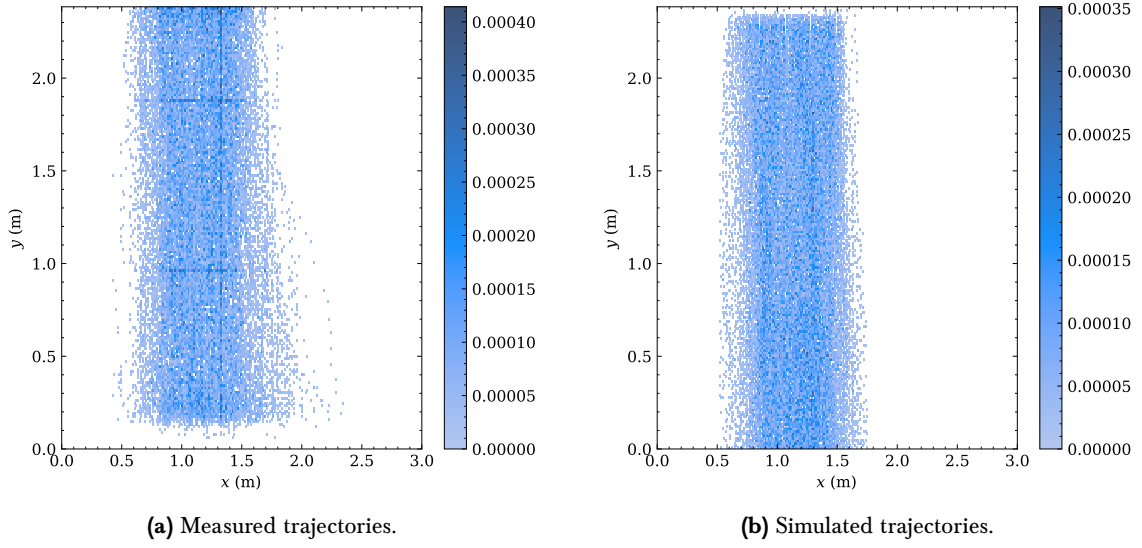
Figure 8.9: Comparison between measured and simulated position distributions. A two-dimensional histogram binning is performed, visualising the area of the domain that is covered by the pedestrians. It can be seen that the beam is more narrow for the simulated pedestrians, and that the measurements include a few more outliers which cross the domain diagonally.
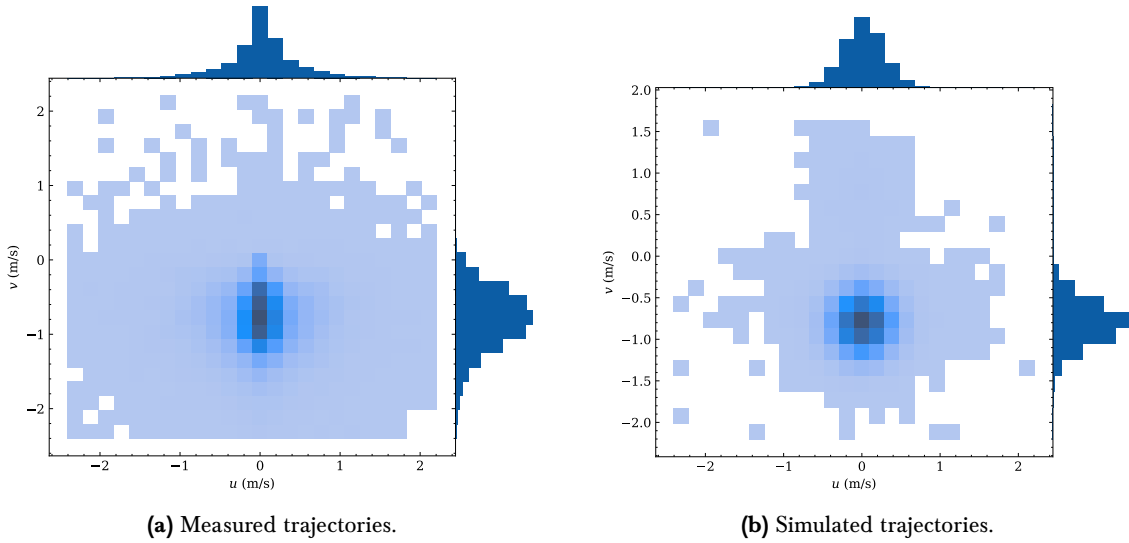
47

(a) Measured trajectories.

(b) Simulated trajectories.

Figure 8.10: Comparison between measured and simulated velocity distributions. Although both peaks are visible at the same location, a small difference visible in the width of the distributions. This is especially true for the distribution of the measured horizontal velocities, which does not follow a Gaussian distribution.
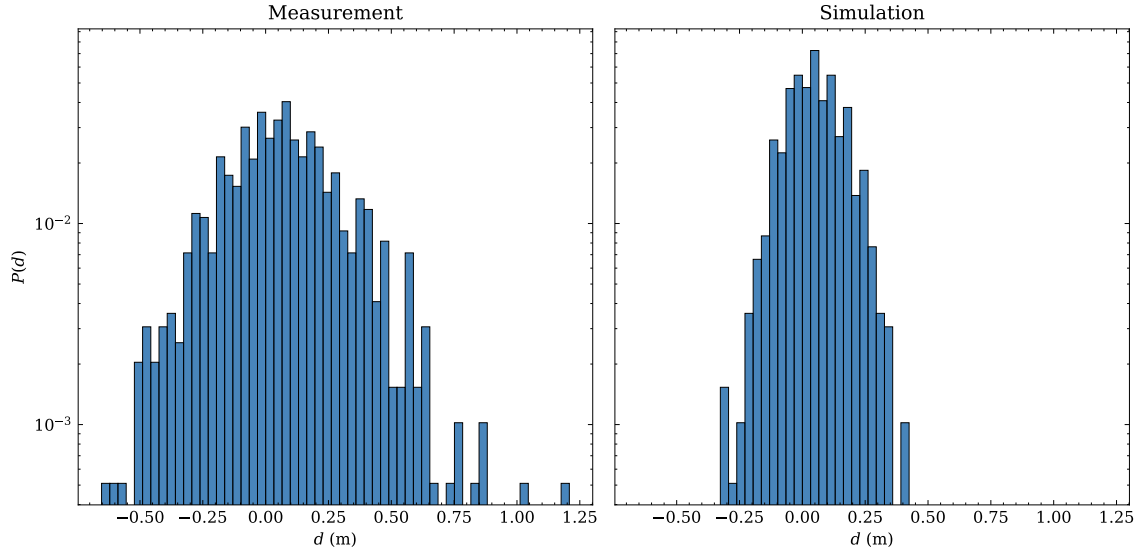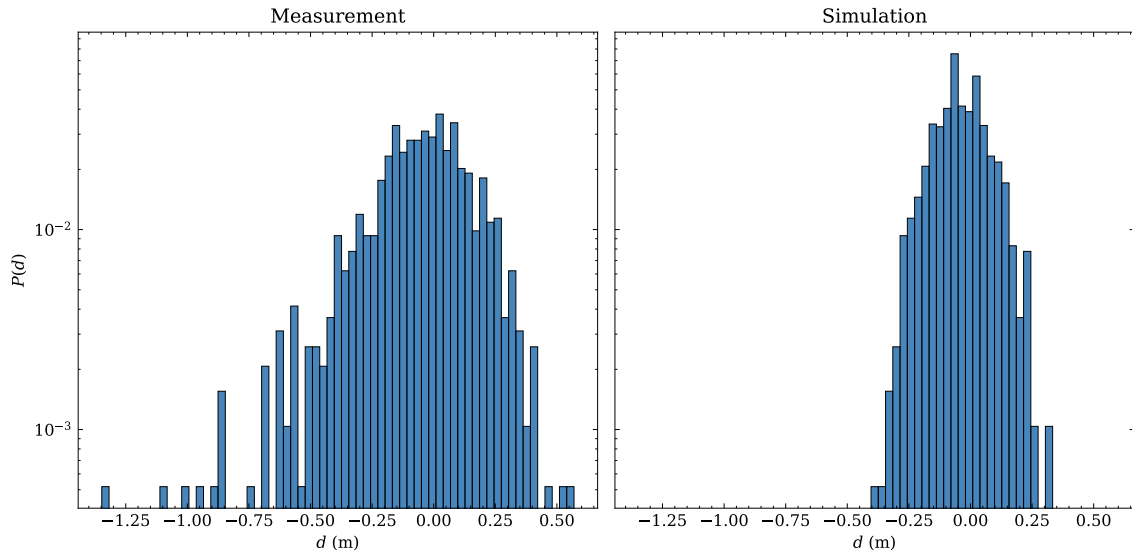

Figure 8.11: Comparison between the distributions of horizontal trajectory difference d in the measured and simulated trajectories. The distributions show an identical mean of $\approx -0.05\,\text{m}$, but the measured trajectories show a much larger variance.

The largest differences arise in the figures describing the comparison of the horizontal difference distribution and the Lagrangian structure function. For the structure functions, a decoupling between $x(t)$ and $x(t+\tau)$ arises for $\tau \gtrsim 3\,\text{s}$. As the typical domain crossing time is given by:

$$t = \frac{y_{domain}}{\bar{v}} = \frac{2.3\,\text{m}}{0.75\,\text{m}\,\text{s}^{-1}} = 3.07\,\text{s},$$

this decoupling is to be expected. It is unclear why the structure function shows a greater difference between measurement and simulation for the unidirectional upward flow, as opposed to the unidirectional downward flow.

Figure 8.12: Comparison of the Lagrangian structure function S_x^2 for the measured and simulated trajectories. Both the transversal ($\hat{x}$) and longitudinal ($\hat{y}$) components are given for both datasets. A clear difference is visible in the slope of both plots: the transversal shows a slope of ≈ 1 indicating diffusive motion, while the longitudinal motion is ballistic, following the slope of ≈ 2.

8.3.3 Bidirectional Flow (III)

A more complicated validation is performed in the center part of the domain. As shown in the previous chapter, the pedestrian flow is clearly bidirectional in this part of the domain: the measured velocity distribution shows two distinct peaks corresponding to pedestrians either travelling upward or downward through the domain. The following trajectory selection criteria are used:

$$t = 0 : 80 \leq x \leq 140,$$
$$t = 0 : 0 \leq y \leq 20 \text{ and } 160 \leq y \leq 180.$$

As the simulation model is able to take the previous velocity into account, the model is able to make a distinction between the two different behaviours displayed in the same domain. Again, an equal amount of measured trajectories is generated using the simulation model.

In Figure 8.13, 25 random trajectories have been plotted for both trajectory datasets. Again, it can be seen that the simulation model is able to correctly generate virtual pedestrians displaying the bidirectional motion. As mentioned in the introduction of this chapter, it is non-trivial that the simulation model is able to accurately reproduce this behaviour.

(a) Measured trajectories. **(b)** Simulated trajectories.
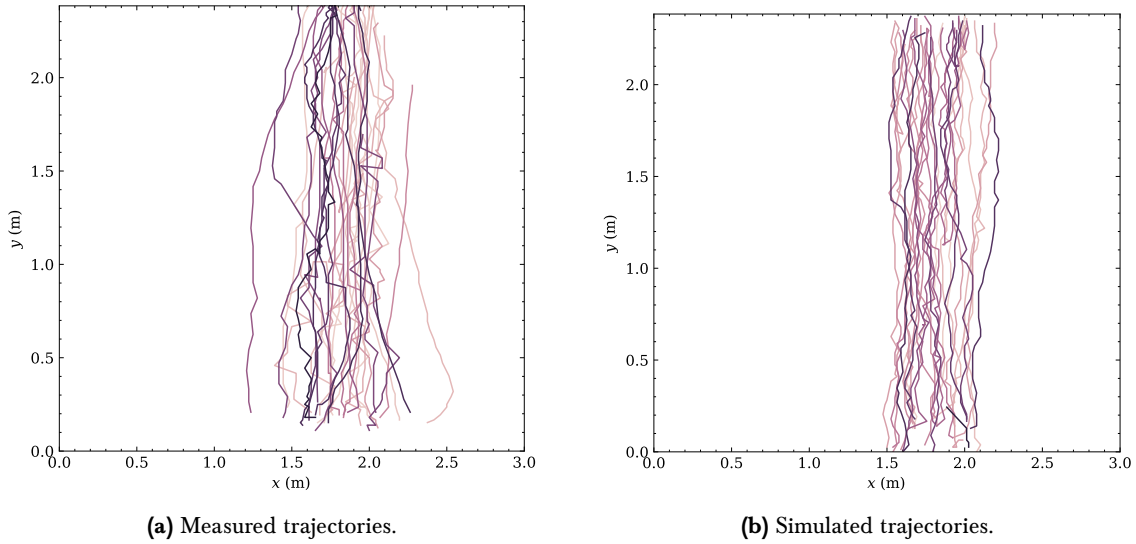
Figure 8.13: Illustrative comparison between measured and simulated trajectories. In both figures, 25 random trajectories have been selected from the respective datasets.

Similarly, in Figure 8.14 it is shown that the simulated position distribution is more narrow than its measured counterpart. Again, the measured trajectories are found to show more diagonal crossings than in the simulated trajectories.

As in the unidirectional flows, the measured trajectories show a clear horizontal peak around $u = 0 \, \text{m s}^{-1}$, where the simulated trajectories show a Gaussian distribution of horizontal velocities. This is shown in Figure 8.15.

In Figure 8.16, the horizontal difference distribution is shown. From this figure the differences visible in the previous figures can be explained. Where the simulated trajectories are found to have less horizontal difference between the initial and final positions, the opposite is true for the measured trajectories. In other words, the measurements show a much larger amount of diagonal domain crossings.

To conclude the comparison metric, in Figure 8.17, the Lagrangian structure function is given for both trajectory datasets. Contrary to the unidirectional flows, the bidirectional flow shows a clear knick in the measured structured functions, around $\tau = 0.4 \, \text{s}$. It is unclear what causes this knick. Another measurement error is visible towards the larger timescales of approximately $\tau \geq 2.0 \, \text{s}$.

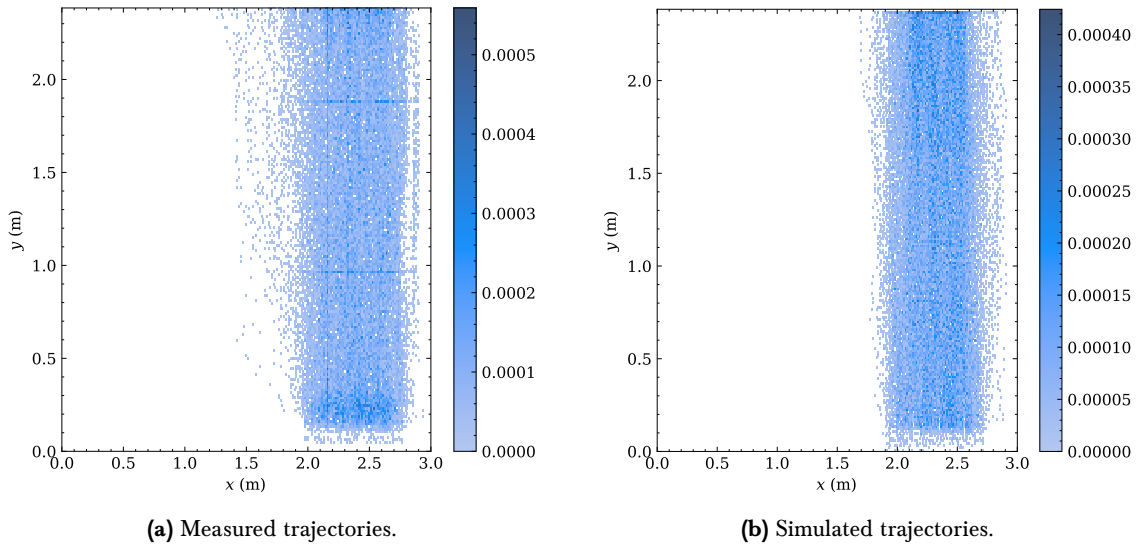**(a)** Measured trajectories.

(b) Simulated trajectories.

Figure 8.14: Comparison between measured and simulated position distributions. A two-dimensional histogram binning is performed, visualising the area of the domain that is covered by the pedestrians. It can be seen that the beam is more narrow for the simulated pedestrians, and that the measurements include more outliers which cross the domain diagonally.
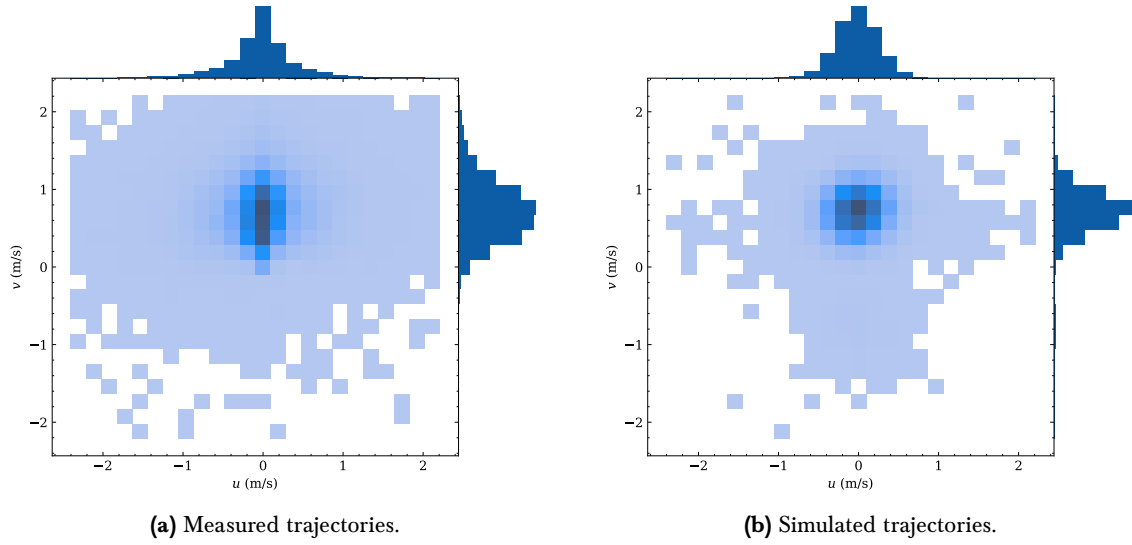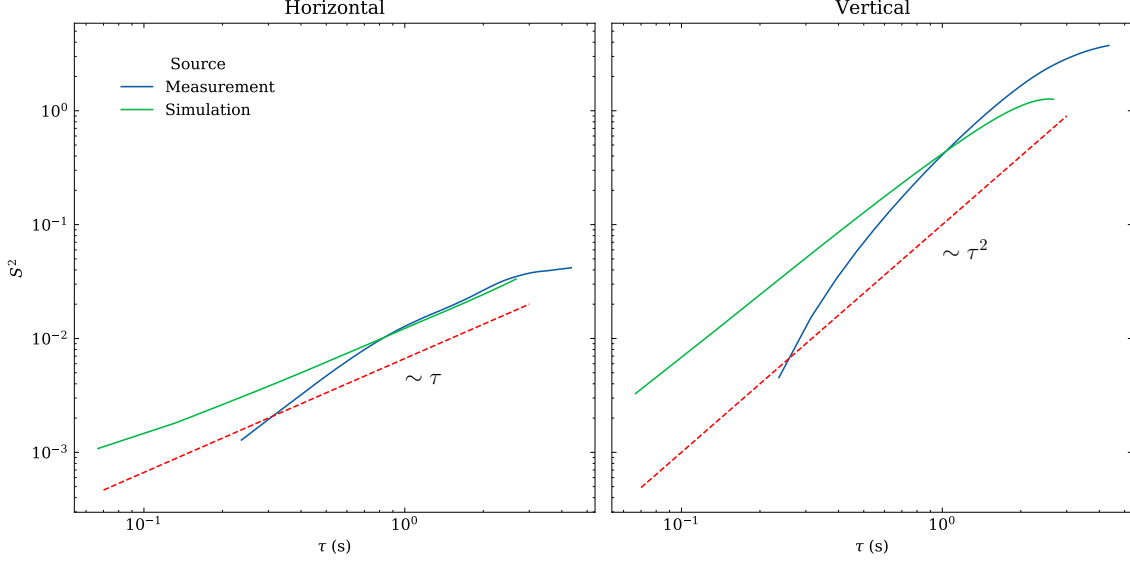

(a) Measured trajectories.

(b) Simulated trajectories.
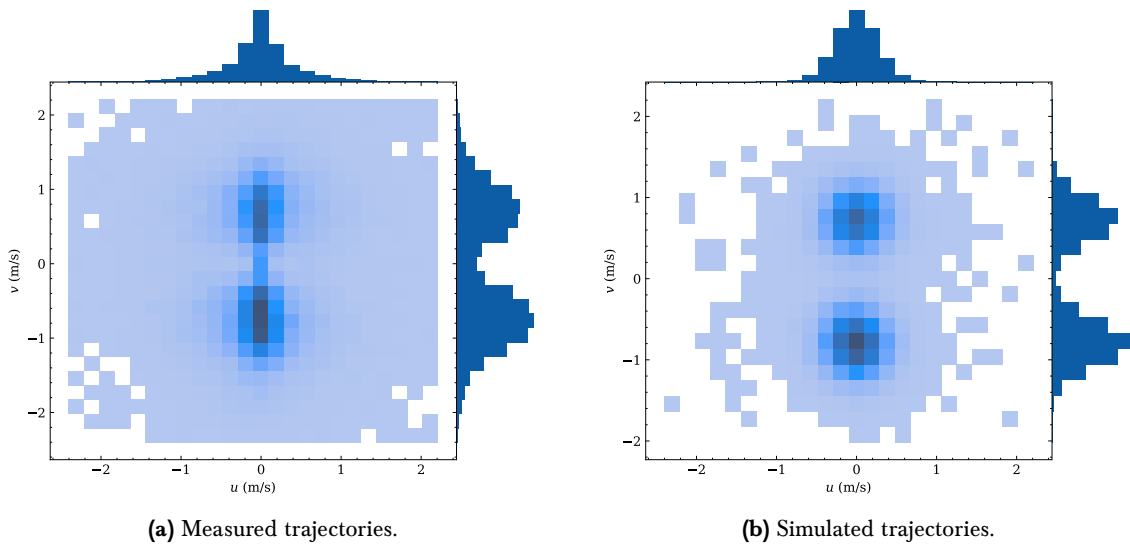
Figure 8.15: Comparison between measured and simulated velocity distributions. As the simulation model is able to distinguish upward and downward travelling pedestrians, two distinct velocity peaks are visible in both distributions. A more clear separation of peaks is visible in the simulation. Additionally, the measured horizontal distribution does not resemble a Gaussian distribution, while the simulated distribution does.
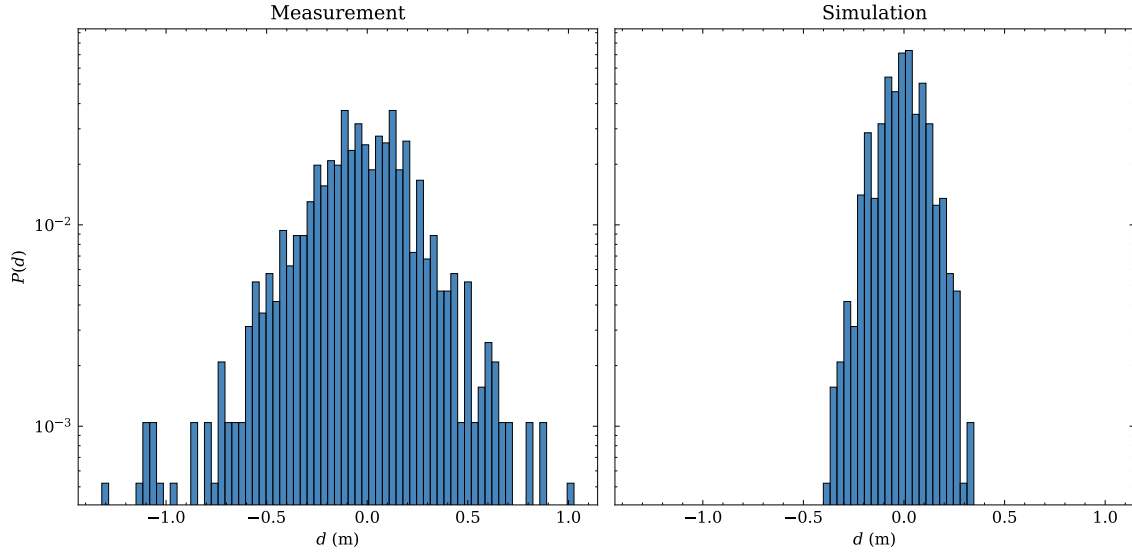
Figure 8.16: Comparison between the distributions of horizontal trajectory difference d in the measured and simulated trajectories. The distributions show an identical mean of $\approx -0.05\,\mathrm{m}$, but the measured trajectories show a much larger variance.

Figure 8.17: Comparison of the Lagrangian structure function S_x^2 for the measured and simulated trajectories. Both the transversal ($\hat{x}$) and longitudinal ($\hat{y}$) components are given for both datasets. Again, a clear difference is visible in the slope of both plots: the transversal shows a slope of ≈ 1 indicating diffusive motion, while the longitudinal motion is ballistic, following the slope of ≈ 2.

8.3.4 Comparison of Complete Flow (IV)

In order to make a complete review of the simulation performance, the last validation is performed by considering the entire measurement domain. Essentially, this simulation is a superposition of the two unidirectional flows and the bidirectional flow in the center of the domain. The simulation is performed by performing a simulation without taking any trajectory selection criteria into account: the entire measured dataset is taken as the initialisation for the simulation. Again, it can be seen that the model is able to generate the correct pedestrian behaviour as seen in the measurements, albeit the known flaws from the uni- and bidirectional flows.

(a) Measured trajectories.　　　　　　**(b)** Simulated trajectories.
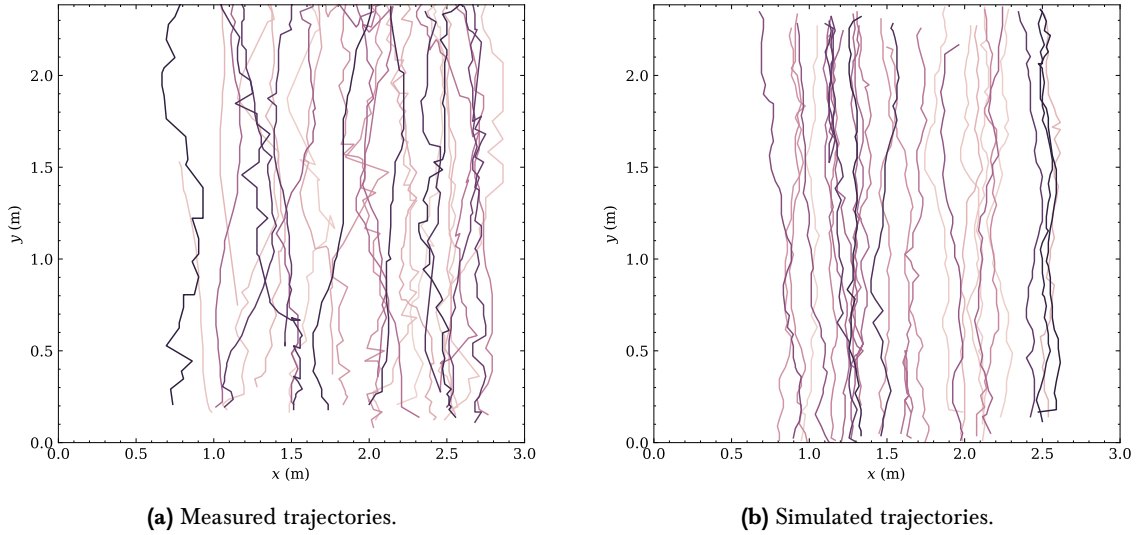
Figure 8.18: Illustrative comparison between measured and simulated trajectories. In both figures, 25 random trajectories have been selected from the respective datasets.

In Figure 8.19, the distributions of pedestrians positions are visualised. No clear differences between the measured and simulated trajectories are apparent in this comparison, apart from having slightly more defined walking lanes in the simulated trajectories. Most probably, this is due to the lack of diagonal trajectories which are less prominent in the simulated dataset.

(a) Measured trajectories.　　　　　　**(b)** Simulated trajectories.
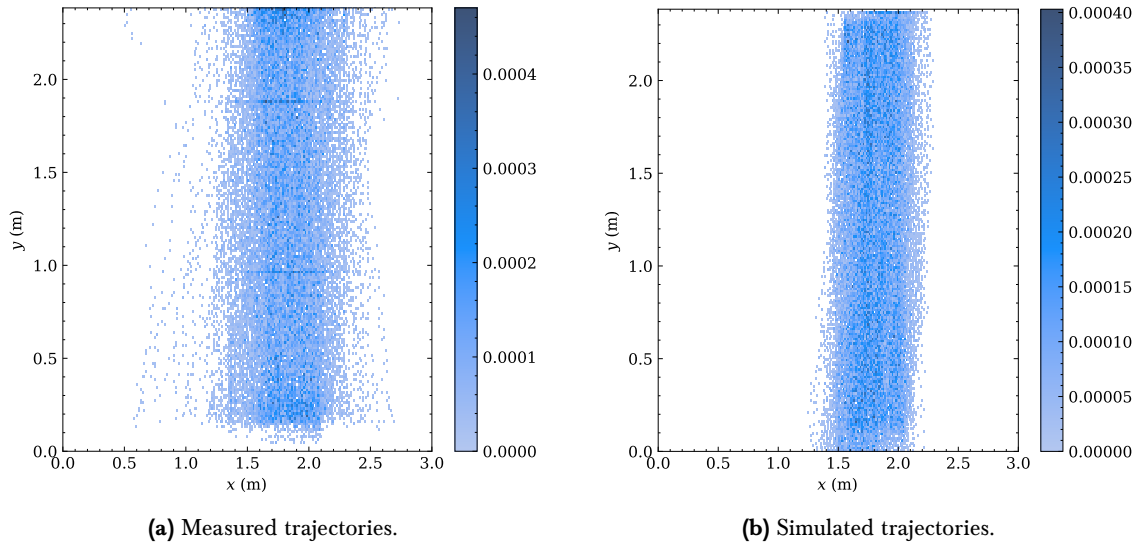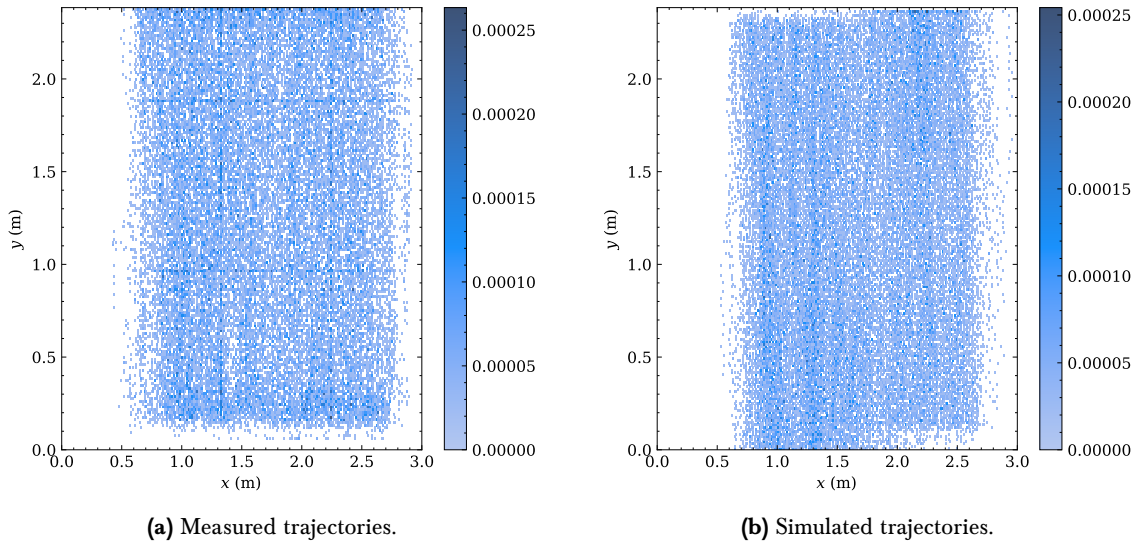
Figure 8.19: Comparison between measured and simulated position distributions. A two-dimensional histogram binning is performed, visualising the area of the domain that is covered by the pedestrians. As the measured pedestrian locations are used the initialisation for the simulation, two distinct lanes are visible in the simulation histogram.

Although the simulated velocity peaks are identical to the measured velocities, the same differences

between the measured and simulated trajectories are present in Figure 8.20. Primarily, the simulation model does not generate the same amount of standstill moments for trajectories as the peak at $\vec{v} = 0$ is clearly absent. Moreover, the simulated horizontal velocities follow a normal distribution, which is not the case for the measured trajectories.

(a) Measured trajectories. (b) Simulated trajectories.
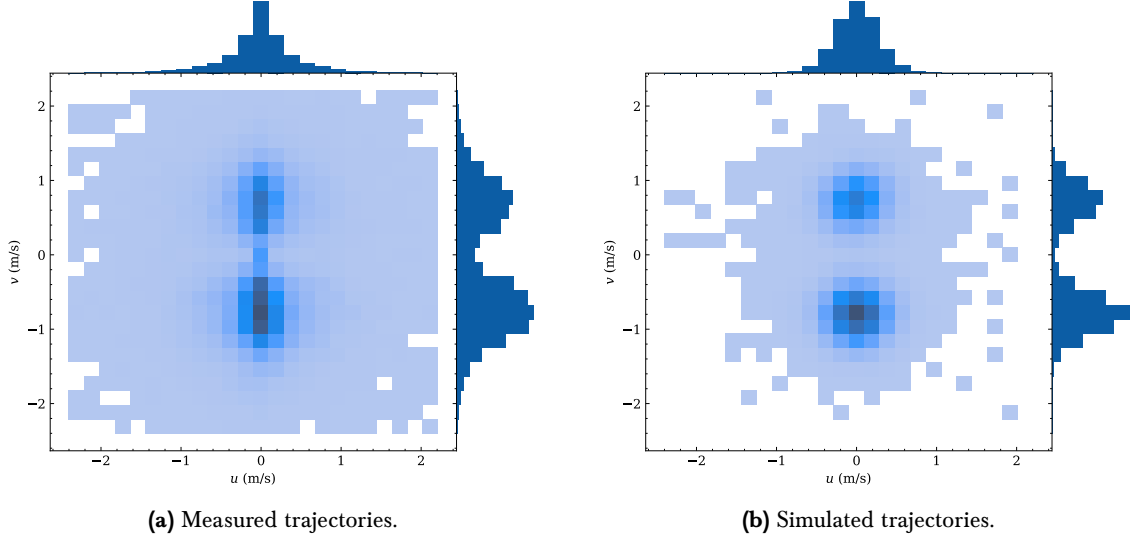
Figure 8.20: Comparison between measured and simulated velocity distributions. As the simulation model is able to distinguish upward and downward travelling pedestrians, two distinct velocity peaks are visible in both distributions. A more clear separation of peaks is visible in the simulation. Additionally, the measured horizontal distribution does not resemble a Gaussian distribution, while the simulated distribution does.

Similar to the unidirectional and bidirectional flows, the horizontal difference distribution is much more narrow for the simulated trajectories.
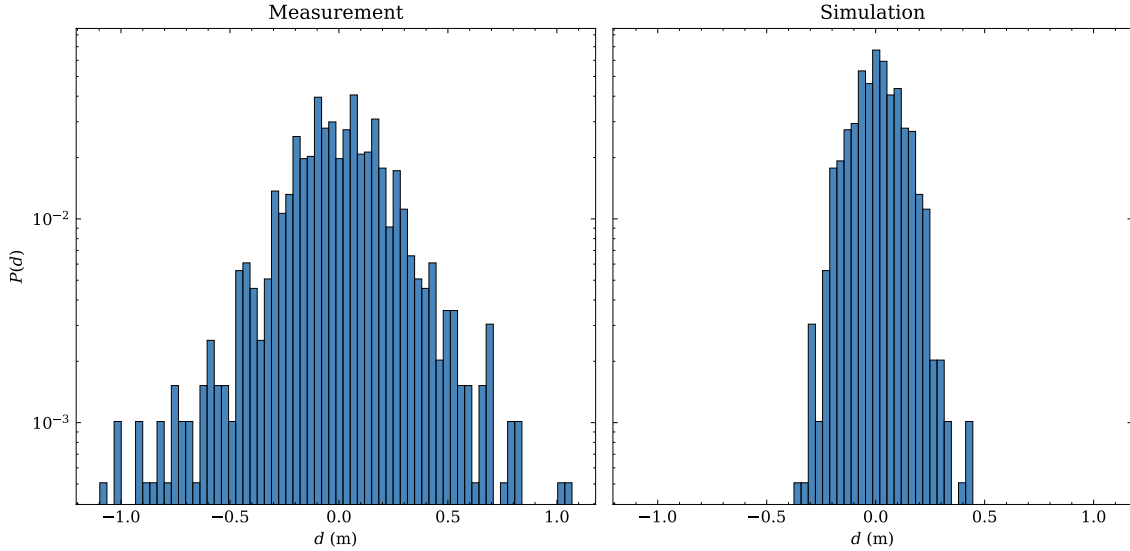
Figure 8.21: Comparison between the distributions of horizontal trajectory difference d in the measured and simulated trajectories. The distributions show an identical mean of $\approx 0.00\,\mathrm{m}$, but the measured trajectories show a much larger variance.

To conclude, the Lagrangian structure functions are calculated for the final simulation, which is visualised in Figure 8.22. Again, the clear difference between the horizontal diffusive motion and the vertical ballistic motion is strongly present. The erratic behaviour as seen in the measurement of the bidirectional center flow (Figure 8.17) is absent in this figure.
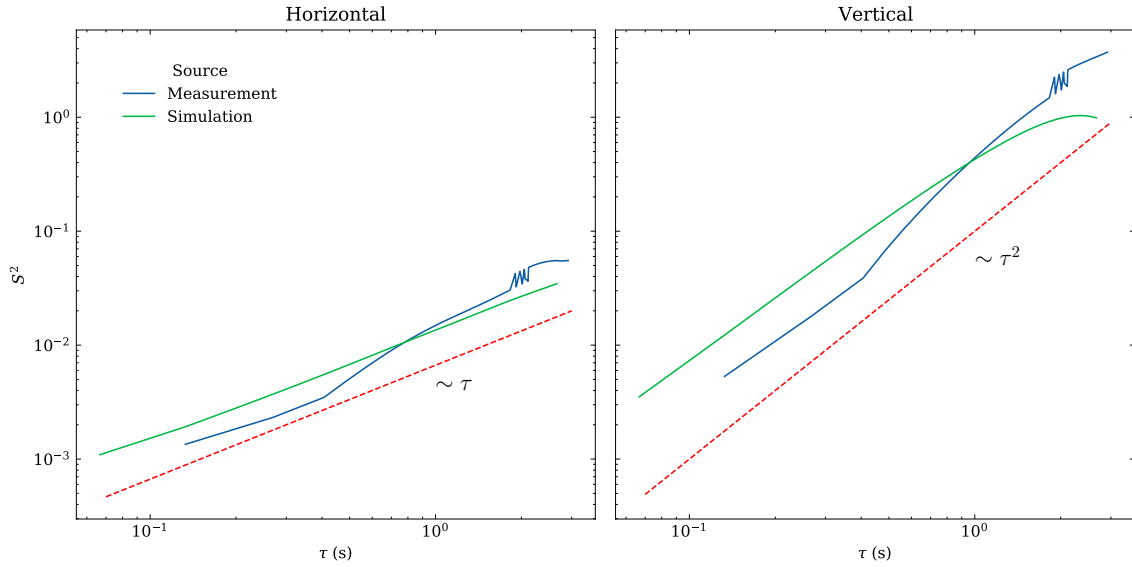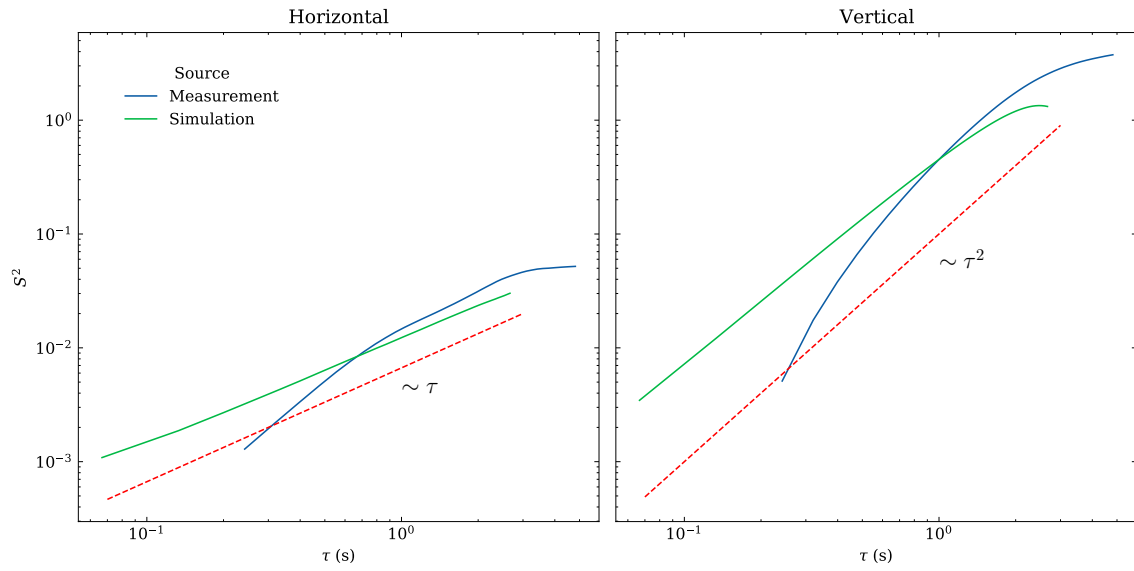
Figure 8.22: Comparison of the Lagrangian structure function S_x^2 for the measured and simulated trajectories. Both the transversal ($\hat{x}$) and longitudinal ($\hat{y}$) components are given for both datasets. Again, a clear difference is visible in the slope of both plots: the transversal shows a slope of ≈ 1 indicating diffusive motion, while the longitudinal motion is ballistic, following the slope of ≈ 2.

Chapter 9

Discussion

In this work, the issue of information assimilation from large-scale high-statistics data in the context of pedestrian dynamics has been considered. Closely following the Lattice Boltzmann Method (LBM) discretisation principles, a data-driven estimation of the equilibrium distribution functions in phase space has been created. As recent technological developments have enabled the anonymous en masse acquisition of pedestrian trajectories, it is the right time to investigate pedestrian dynamics from a mesoscopic perspective.

The considered analysis method is based on a lattice Boltzmann-inspired discretisation of the measured pedestrian behaviour in phase space. A double discretisation method has been applied, resulting in a high-resolution description of the phase space dynamics. The secondary discretisation enables the conditional calculation of equilibrium distribution functions. This approach is necessary to distinguish varying pedestrian behaviour in different parts of the domain and is also used to perform separate analyses of pedestrian walking in different directions. Following the similarity between kinetic theory and pedestrian dynamics, it is expected that, on average, the distribution of pedestrians in phase space is equal to the equilibrium distribution function. The discretised analysis approach naturally creates an environment suitable for the simulation of artificial trajectories. For this reason, a Monte Carlo-like simulation model has been developed, allowing for the generation of individual trajectories based on the measured equilibrium distribution functions.

The novel method of the data-driven analysis and simulation of pedestrian trajectories has been applied to a measurement site in Amsterdam, the Netherlands. Measurements took place during the first seven evenings of September 2019, before the outbreak of the COVID-19 pandemic. A large amount of trajectory data was acquired at this location, consisting of over $12 * 10^5$ measured trajectories. As the recorded trajectory data suffered from dataset quality issues, a random forest classification model was developed to discard invalid measurements. After this reduction, the dataset consisted of approximately $5 * 10^4$ trajectories, showing a clear bidirectional pedestrian flow. It is shown that the analysis model was able to memorise the statistics of the captured pedestrian motion efficiently. Furthermore, the double discretisation approach offers a good balance between an accurate description of measured behaviour and computational efficiency. The finest possible discretisation has been applied, as a 1:1 px/cell mapping has been used. If higher-resolution depth images become available in future work, a coarser discretisation can be applied. In the measured trajectory dataset, the measured velocity distributions are approximately Gaussian, showing more significant deviations at a greater distance from the velocity means. This imperfect agreement can be explained by the simplification of studying pedestrian dynamics using kinetic theory. Additionally, this might result from the measured pedestrian dynamics being of high complexity following the narrow, crowded flow in the alley.

In this work, a data-driven estimation of equilibrium distribution functions in the context of pedestrian dynamics has been provided. A microscopic Monte Carlo-based simulation method has been proven to show a fair agreement with the measured pedestrian dynamics. Following the similarity between the used discretisation method and LBM-based approaches, fully mesoscopic simulation possibilities can be considered for future research. As such approaches feature many benefits over microscopic simulation methods (e.g. highly efficient in parallel computing and robust handling of complex geometries), this should strongly be considered for further research. While these methods would come with difficulties – e.g. designing an appropriate collision term in the Boltzmann equation is highly non-trivial – this could lead to promising novel data-driven simulation approaches.

Bibliography

[1]   A. Corbetta, L. Bruno, A. Muntean and F. Toschi, "High statistics measurements of pedestrian dynamics", *Transportation Research Procedia*, vol. 2, pp. 96–104, 2014.

[2]   A. Corbetta, J. Meeusen, C. min Lee and F. Toschi, "Continuous measurements of real-life bidirectional pedestrian flows on a wide walkway", *Proceedings of Pedestrian and Evacuation Dynamics (Special Issue on Collective Dynamics)*, pp. 18–24, 2016.

[3]   A. Corbetta, W. Kroneman, M. Donners, A. Haans, P. Ross, M. Trouwborst, S. Van De Wijdeven, M. Hultermans, D. Sekulovski, F. Van Der Heijden *et al.*, "A large-scale real-life crowd steering experiment via arrow-like stimuli", *Collective Dynamics*, vol. 5, pp. 61–68, 2020.

[4]   A. Mayer, T. Skowroneck and R. Blanton, *The Science of Walking*. University of Chicago Press, 2020. doi: `10.7208/chicago/9780226352480.001.0001`. [Online]. Available: `https://doi.org/10.7208/chicago/9780226352480.001.0001`.

[5]   F. Martinez-Gil, M. Lozano, I. García-Fernández and F. Fernández, "Modeling, evaluation, and scale on artificial pedestrians", *ACM Computing Surveys*, vol. 50, no. 5, pp. 1–35, Nov. 2017. doi: `10.1145/3117808`. [Online]. Available: `https://doi.org/10.1145/3117808`.

[6]   D. C. Duives, W. Daamen and S. P. Hoogendoorn, "State-of-the-art crowd motion simulation models", *Transportation Research Part C: Emerging Technologies*, vol. 37, pp. 193–209, Dec. 2013. doi: `10.1016/j.trc.2013.02.005`. [Online]. Available: `https://doi.org/10.1016/j.trc.2013.02.005`.

[7]   G. Le Bon, *The Crowd: A Study of the Popular Mind*. Courier Corporation, 2002.

[8]   C. A. S. Pouw, F. Toschi, F. van Schadewijk and A. Corbetta, "Monitoring physical distancing for crowd management: Real-time trajectory and group analysis", *PLOS ONE*, vol. 15, no. 10, D. R. Chialvo, Ed., e0240963, Oct. 2020. doi: `10.1371/journal.pone.0240963`. [Online]. Available: `https://doi.org/10.1371/journal.pone.0240963`.

[9]   J. Adrian, N. Bode, M. Amos, M. Baratchi, M. Beermann, M. Boltes, A. Corbetta, G. Dezecache, J. Drury, Z. Fu, R. Geraerts, S. Gwynne, G. Hofinger, A. Hunt, T. Kanters, A. Kneidl, K. Konya, G. Köster, M. Küpper, G. Michalareas, F. Neville, E. Ntontis, S. Reicher, E. Ronchi, A. Schadschneider, A. Seyfried, A. Shipman, A. Sieben, M. Spearpoint, G. Sullivan, A. Templeton, F. Toschi, Z. Yücel, F. Zanlungo, I. Zuriguel, N. van der Wal, F. van Schadewijk, C. von Krüchten and N. Wijermans, "A glossary for research on human crowd dynamics", English, *Collective Dynamics*, vol. 4, pp. 1–13, 2019, issn: 2366-8539. doi: `10.17815/CD.2019.19`.

[10] R. S. Sobel and N. Lillith, "Determinants of nonstationary personal space invasion", *The Journal of Social Psychology*, vol. 97, no. 1, pp. 39–45, 1975.

[11] L. D. Vanumu, K. R. Rao and G. Tiwari, "Fundamental diagrams of pedestrian flow characteristics: A review", *European Transport Research Review*, vol. 9, no. 4, Sep. 2017. doi: `10.1007/s12544-017-0264-6`. [Online]. Available: `https://doi.org/10.1007/s12544-017-0264-6`.

[12] C. Feliciani, H. Murakami and K. Nishinari, "A universal function for capacity of bidirectional pedestrian streams: Filling the gaps in the literature", *PLOS ONE*, vol. 13, e0208496, Dec. 2018. doi: `10.1371/journal.pone.0208496`.

[13] S. Bandini, B. Chopard and M. Tomassini, Eds., *Cellular Automata*. Springer Berlin Heidelberg, 2002. doi: `10.1007/3-540-45830-1`. [Online]. Available: `https://doi.org/10.1007/3-540-45830-1`.

[14] V. J. Blue and J. L. Adler, "Emergent fundamental pedestrian flows from cellular automata microsimulation", *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1644, no. 1, pp. 29–36, Jan. 1998. doi: `10.3141/1644-04`. [Online]. Available: `https://doi.org/10.3141/1644-04`.

[15] V. J. Blue and J. L. Adler, "Cellular automata microsimulation of bidirectional pedestrian flows", *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1678, no. 1, pp. 135–141, Jan. 1999. doi: `10.3141/1678-17`. [Online]. Available: `https://doi.org/10.3141/1678-17`.

[16] L. Yang, D. Zhao, J. Li and T. Fang, "Simulation of the kin behavior in building occupant evacuation based on cellular automaton", *Building and Environment*, vol. 40, no. 3, pp. 411–415, Mar. 2005. doi: `10.1016/j.buildenv.2004.08.005`. [Online]. Available: `https://doi.org/10.1016/j.buildenv.2004.08.005`.

[17] A. Varas, M. Cornejo, D. Mainemer, B. Toledo, J. Rogan, V. Muñoz and J. Valdivia, "Cellular automaton model for evacuation process with obstacles", *Physica A: Statistical Mechanics and its Applications*, vol. 382, no. 2, pp. 631–642, Aug. 2007. doi: `10.1016/j.physa.2007.04.006`. [Online]. Available: `https://doi.org/10.1016/j.physa.2007.04.006`.

[18] Y. F. Yu and W. G. Song, "Cellular automaton simulation of pedestrian counter flow considering the surrounding environment", *Physical Review E*, vol. 75, no. 4, Apr. 2007. doi: `10.1103/physreve.75.046112`. [Online]. Available: `https://doi.org/10.1103/physreve.75.046112`.

[19] S. Wei-Guo, Y. Yan-Fei, W. Bing-Hong and F. Wei-Cheng, "Evacuation behaviors at exit in CA model with force essentials: A comparison with social force model", *Physica A: Statistical Mechanics and its Applications*, vol. 371, no. 2, pp. 658–666, Nov. 2006. doi: `10.1016/j.physa.2006.03.027`. [Online]. Available: `https://doi.org/10.1016/j.physa.2006.03.027`.

[20] J. Dijkstra, H. J. P. Timmermans and A. J. Jessurun, "A multi-agent cellular automata system for visualising simulated pedestrian activity", in *Theory and Practical Issues on Cellular Automata*, Springer London, 2001, pp. 29–36. doi: `10.1007/978-1-4471-0709-5_4`. [Online]. Available: `https://doi.org/10.1007/978-1-4471-0709-5_4`.

[21] J. Wąs, B. Gudowski and P. J. Matuszyk, "Social distances model of pedestrian dynamics", in *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 2006, pp. 492–501. doi: `10.1007/11861201_57`. [Online]. Available: `https://doi.org/10.1007/11861201_57`.

[22] M. Bierlaire, "Biogeme: A free package for the estimation of discrete choice models", in *Swiss transport research conference*, 2003.

[23] R. Lubaś, J. Wąs and J. Porzycki, "Cellular automata as the basis of effective and realistic agent-based models of crowd behavior", *The Journal of Supercomputing*, vol. 72, no. 6, pp. 2170–2196, Apr. 2016. doi: `10.1007/s11227-016-1718-7`. [Online]. Available: `https://doi.org/10.1007/s11227-016-1718-7`.

[24] C Burstedde, K Klauck, A Schadschneider and J Zittartz, "Simulation of pedestrian dynamics using a two-dimensional cellular automaton", *Physica A: Statistical Mechanics and its Applications*, vol. 295, no. 3-4, pp. 507–525, Jun. 2001. doi: `10.1016/s0378-4371(01)00141-8`. [Online]. Available: `https://doi.org/10.1016/s0378-4371(01)00141-8`.

[25] A. Schadschneider, A. Kirchner and K. Nishinari, "Ca approach to collective phenomena in pedestrian dynamics", in *International Conference on Cellular Automata*, Springer, 2002, pp. 239–248.

[26] D. Hartmann, "Adaptive pedestrian dynamics based on geodesics", *New Journal of Physics*, vol. 12, no. 4, p. 043032, Apr. 2010. doi: `10.1088/1367-2630/12/4/043032`. [Online]. Available: `https://doi.org/10.1088/1367-2630/12/4/043032`.

[27] A. Kirchner, H. Klüpfel, K. Nishinari, A. Schadschneider and M. Schreckenberg, "Simulation of competitive egress behavior: Comparison with aircraft evacuation data", *Physica A: Statistical Mechanics and its Applications*, vol. 324, no. 3-4, pp. 689–697, Jun. 2003. doi: `10.1016/s0378-4371(03)00076-1`. [Online]. Available: `https://doi.org/10.1016/s0378-4371(03)00076-1`.

[28] A. Kirchner, K. Nishinari and A. Schadschneider, "Friction effects and clogging in a cellular automaton model for pedestrian dynamics", *Physical Review E*, vol. 67, no. 5, May 2003. doi: `10.1103/physreve.67.056122`. [Online]. Available: `https://doi.org/10.1103/physreve.67.056122`.

[29] T. Kretz, "Pedestrian traffic: On the quickest path", *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2009, no. 03, P03012, Mar. 2009. doi: 10.1088/1742-5468/2009/03/p03012. [Online]. Available: https://doi.org/10.1088/1742-5468/2009/03/p03012.

[30] M. J. Seitz and G. Köster, "Natural discretization of pedestrian movement in continuous space", *Physical Review E*, vol. 86, no. 4, Oct. 2012. doi: 10.1103/physreve.86.046108. [Online]. Available: https://doi.org/10.1103/physreve.86.046108.

[31] A. Schadschneider and A. Seyfried, "Empirical results for pedestrian dynamics and their implications for cellular automata models", in *Pedestrian Behavior*, Emerald Group Publishing Limited, Nov. 2009, pp. 27–43. doi: 10.1108/9781848557512-002. [Online]. Available: https://doi.org/10.1108/9781848557512-002.

[32] B. Leng, J. Wang, W. Zhao and Z. Xiong, "An extended floor field model based on regular hexagonal cells for pedestrian simulation", *Physica A: Statistical Mechanics and its Applications*, vol. 402, pp. 119–133, May 2014. doi: 10.1016/j.physa.2014.01.039. [Online]. Available: https://doi.org/10.1016/j.physa.2014.01.039.

[33] M. Muramatsu, T. Irie and T. Nagatani, "Jamming transition in pedestrian counter flow", *Physica A: Statistical Mechanics and its Applications*, vol. 267, no. 3-4, pp. 487–498, May 1999. doi: 10.1016/s0378-4371(99)00018-7. [Online]. Available: https://doi.org/10.1016/s0378-4371(99)00018-7.

[34] M. Muramatsu and T. Nagatani, "Jamming transition in two-dimensional pedestrian traffic", *Physica A: Statistical Mechanics and its Applications*, vol. 275, no. 1-2, pp. 281–291, Jan. 2000. doi: 10.1016/s0378-4371(99)00447-1. [Online]. Available: https://doi.org/10.1016/s0378-4371(99)00447-1.

[35] A. Lerner, Y. Chrysanthou and D. Lischinski, "Crowds by example", *Computer Graphics Forum*, vol. 26, no. 3, pp. 655–664, Sep. 2007. doi: 10.1111/j.1467-8659.2007.01089.x. [Online]. Available: https://doi.org/10.1111/j.1467-8659.2007.01089.x.

[36] J. Porzycki, R. Lubaś, M. Mycek and J. Wąs, "Dynamic data–driven simulation of pedestrian movement with automatic validation", in *Traffic and Granular Flow 13*, Springer International Publishing, Nov. 2014, pp. 129–136. doi: 10.1007/978-3-319-10629-8_15. [Online]. Available: https://doi.org/10.1007/978-3-319-10629-8_15.

[37] E. Ju, M. G. Choi, M. Park, J. Lee, K. H. Lee and S. Takahashi, "Morphable crowds", *ACM Transactions on Graphics*, vol. 29, no. 6, pp. 1–10, Dec. 2010. doi: 10.1145/1882261.1866162. [Online]. Available: https://doi.org/10.1145/1882261.1866162.

[38] S. Lemercier, A. Jelic, R. Kulpa, J. Hua, J. Fehrenbach, P. Degond, C. Appert-Rolland, S. Donikian and J. Pettré, "Realistic following behaviors for crowd simulation", *Computer Graphics Forum*, vol. 31, no. 2pt2, pp. 489–498, May 2012. doi: 10.1111/j.1467-8659.2012.03028.x. [Online]. Available: https://doi.org/10.1111/j.1467-8659.2012.03028.x.

[39] S. Kim, A. Bera, A. Best, R. Chabra and D. Manocha, "Interactive and adaptive data-driven crowd simulation", in *2016 IEEE Virtual Reality (VR)*, IEEE, Mar. 2016. doi: 10.1109/vr.2016.7504685. [Online]. Available: https://doi.org/10.1109/vr.2016.7504685.

[40] K. H. Lee, M. G. Choi, Q. Hong and J. Lee, "Group behavior from video: A data-driven approach to crowd simulation", in *Proceedings of the 2007 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, ser. SCA '07, San Diego, California: Eurographics Association, 2007, 109118, isbn: 9781595936240.

[41] J. Willems, A. Corbetta, V. Menkovski and F. Toschi, "Pedestrian orientation dynamics from high-fidelity measurements", *Scientific Reports*, vol. 10, no. 1, Jul. 2020. doi: 10.1038/s41598-020-68287-6. [Online]. Available: https://doi.org/10.1038/s41598-020-68287-6.

[42] T. Fernando, S. Denman, S. Sridharan and C. Fookes, "Tracking by prediction: A deep generative model for mutli-person localisation and tracking", in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, IEEE, 2018, pp. 1122–1132.

[43] S. P. Hoogendoorn, W. Daamen and P. H. Bovy, "Extracting microscopic pedestrian characteristics from video data", in *Transportation Research Board Annual Meeting*, Citeseer, 2003, pp. 1–15.

[44] A. Johansson and D. Helbing, "Analysis of empirical trajectory data of pedestrians", in *Pedestrian and Evacuation Dynamics 2008*, Springer Berlin Heidelberg, Dec. 2009, pp. 203–214. doi: `10.1007/978-3-642-04504-2_15`. [Online]. Available: `https://doi.org/10.1007/978-3-642-04504-2_15`.

[45] J. Kerridge, R. Kukla, A. Willis, A. Armitage, D. Binnie and L. Lei, "A comparison of video and infrared based tracking of pedestrian movements", in *Traffic and Granular Flow '03*, Springer Berlin Heidelberg, pp. 383–391. doi: `10.1007/3-540-28091-x_37`. [Online]. Available: `https://doi.org/10.1007/3-540-28091-x_37`.

[46] J. Kerridge, S. Keller, T. Chamberlain and N. Sumpter, "Collecting pedestrian trajectory data in real-time", in *Pedestrian and Evacuation Dynamics 2005*, Springer Berlin Heidelberg, pp. 27–39. doi: `10.1007/978-3-540-47064-9_3`. [Online]. Available: `https://doi.org/10.1007/978-3-540-47064-9_3`.

[47] S. Seer, N. Brändle and C. Ratti, "Kinects and human kinetics: A new approach for studying pedestrian behavior", *Transportation Research Part C: Emerging Technologies*, vol. 48, pp. 212–228, Nov. 2014. doi: `10.1016/j.trc.2014.08.012`. [Online]. Available: `https://doi.org/10.1016/j.trc.2014.08.012`.

[48] D. Brščić, T. Kanda, T. Ikeda and T. Miyashita, "Person tracking in large public spaces using 3-d range sensors", *IEEE Transactions on Human-Machine Systems*, vol. 43, no. 6, pp. 522–534, 2013.

[49] A. Corbetta, V. Menkovski and F. Toschi, "Weakly supervised training of deep convolutional neural networks for overhead pedestrian localization in depth fields", in *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, IEEE, 2017, pp. 1–6.

[50] Orbbec. (2016). Orbbec persee product specification, [Online]. Available: `https://orbbec3d.com/product-persee/` (visited on 04/05/2020).

[51] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified, real-time object detection", in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[52] W. Kroneman, A. Corbetta and F. Toschi, "Accurate pedestrian localization in overhead depth images via height-augmented HOG", *Collective Dynamics*, vol. 5, Mar. 2020. doi: `10.17815/cd.2020.30`. [Online]. Available: `https://doi.org/10.17815/cd.2020.30`.

[53] O. software consortium. (2012). OpenPTV: Open source particle tracking velocimetry, [Online]. Available: `http://www.openptv.net/` (visited on 04/05/2020).

[54] J. Willneff, "A spatio-temporal matching algorithm for 3d particle tracking velocimetry", PhD thesis, ETH Zurich, 2003.

[55] D. B. Allan, T. Caswell, N. C. Keim, C. M. van der Wel and R. W. Verweij, *Soft-matter/trackpy: Trackpy v0.5.0*, 2021. doi: `10.5281/ZENODO.4682814`. [Online]. Available: `https://zenodo.org/record/4682814`.

[56] J. C. Crocker and D. G. Grier, "Methods of digital video microscopy for colloidal studies", *Journal of Colloid and Interface Science*, vol. 179, no. 1, pp. 298–310, Apr. 1996. doi: `10.1006/jcis.1996.0217`. [Online]. Available: `https://doi.org/10.1006/jcis.1996.0217`.

[57] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: A system for large-scale machine learning", in *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, 2016, pp. 265–283.

[58] F. Chollet *et al.* (2015). Keras, [Online]. Available: `https://github.com/fchollet/keras`.

[59] T. Krüger, H. Kusumaatmaja, A. Kuzmin, O. Shardt, G. Silva and E. M. Viggen, *The Lattice Boltzmann Method*. Springer International Publishing, 2017. doi: `10.1007/978-3-319-44649-3`. [Online]. Available: `https://doi.org/10.1007/978-3-319-44649-3`.

[60] AT5. (Jul. 2020). Weer grote drukte op de wallen: 'alsof corona niet bestaat' (dutch), [Online]. Available: `https://www.at5.nl/artikelen/203216/weer-grote-drukte-op-de-wallen-alsof-corona-niet-bestaat`.

[61] T. Danka and P. Horvath, "ModAL: A modular active learning framework for Python", available on arXiv at `https://arxiv.org/abs/1805.00979`. [Online]. Available: `https://github.com/cosmic-cortex/modAL`.

[62] K. Fitzpatrick, M. A. Brewer and S. Turner, "Another look at pedestrian walking speed", *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1982, no. 1, pp. 21–29, Jan. 2006. doi: `10.1177/0361198106198200104`. [Online]. Available: `https://doi.org/10.1177/0361198106198200104`.