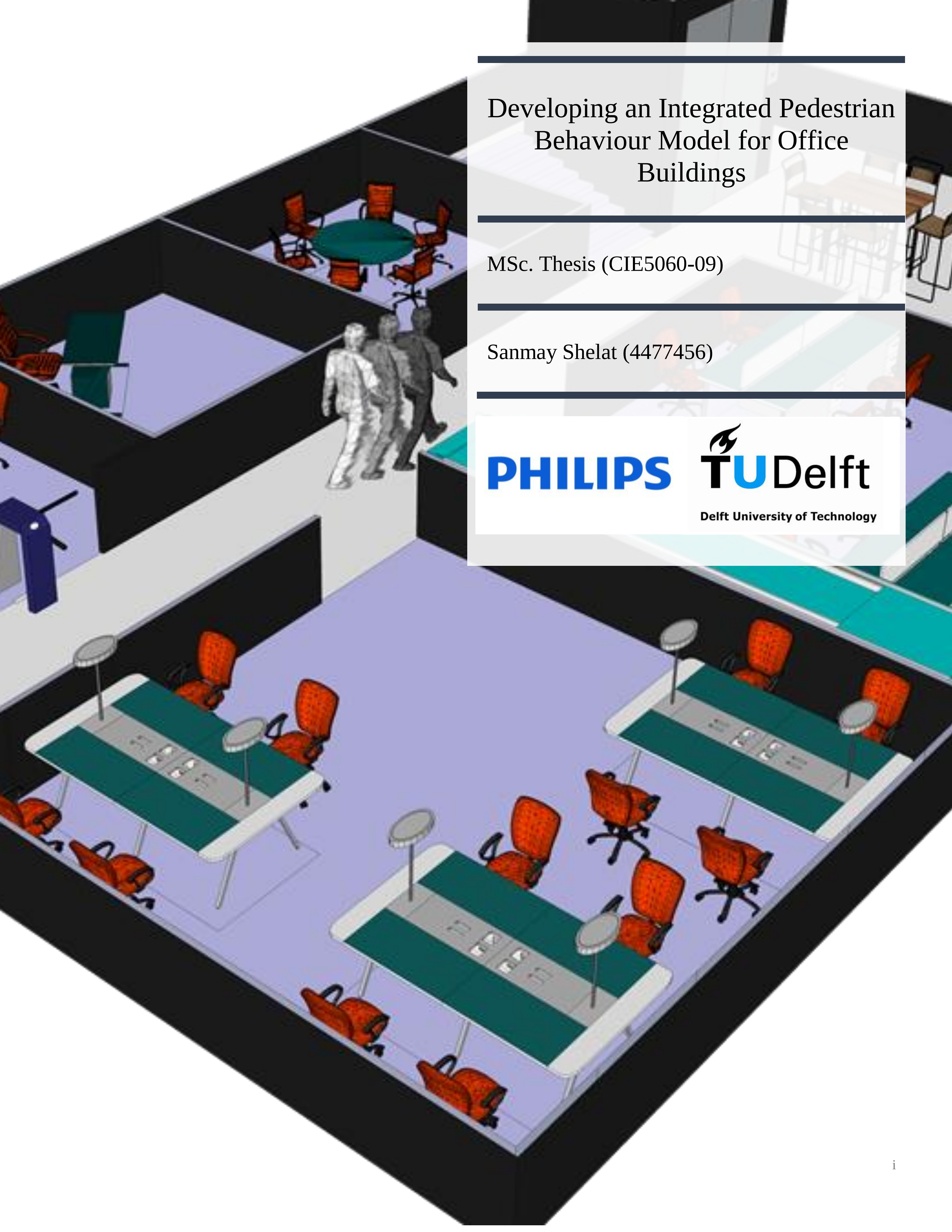

Developing an Integrated Pedestrian Behaviour Model for Office Buildings

MSc. Thesis (CIE5060-09)

Sanmay Shelat (4477456)

Cover image: A pedestrian trail in the case study office used in this thesis (Author).

Developing an Integrated Pedestrian Behaviour Model for Office Buildings

by

Sanmay Shelat

Defended on

29 November 2017

Thesis submitted in partial fulfilment of the regulations for the degree of Master of Science

At the Delft University of Technology,

Faculty of Civil Engineering and Geosciences,

Department Transport and Planning.

Student Number

4477456

Graduation Committee

Prof. Dr. ir. Serge Hoogendoorn	Chairperson, TU Delft (CiTG)
Dr. ir. Winnie Daamen	Daily Supervisor, TU Delft (CiTG)
Dr. ir. Stefan van der Spek	External Supervisor, TU Delft (BK)
Dr. ir. Dorine Duives	External Supervisor, TU Delft (CiTG)
Mr. Bjorn Kaag	Company Supervisor, Philips Lighting

Delft University of Technology

Acknowledgements

As I finish this master's degree and move on to starting as a doctoral candidate, I would like to show my deepest gratitude to the following:

Dr Anurag Kandya	*for fostering a research environment under an unlikely situation and always encouraging the researcher in me*
The Transport Institute, TU Delft	*for their generous grant that allowed me to have one less thing to worry about*
Dr. Oded Cats	*for supervising my Honours programme at TU Delft and being a mentor to me for the major part of my time here*
Dr. Winnie Daamen	*for her unwavering support and encouragement throughout this master thesis*
Dr. Niels van Oort	*for offering the doctoral position which helped me keep confidence in myself and gave me something to look forward to*
Philips Lighting	*for providing an interesting topic for this master thesis, all the support I needed during my time there, and demonstrating their belief in my work*
Friends here at TU Delft and at Philips Lighting	*for helping me break away from the solitude of work*
Kavita Rathore	*for always being there to listen to what is sometimes an unending stream of complaints*
My Parents	*for always being supportive of my decisions*

Executive Summary

Background and Research Objective

As more and more people work in offices, humans are spending a significant proportion of time in office buildings. Therefore, there is a need to optimize building services to achieve energy efficiency whilst improving the well-being of occupants. For the research and evaluation these data is required. Specifically, since building services are dependent on the behaviour of occupants in the buildings, data on the locations and movements of occupants in office buildings is required. Given that recent technologies allow tracking of individual occupants, the data for testing control systems making use of these advanced detection techniques should also be of such high resolution. However, obtaining this data is not straightforward for three reasons: (i) the need to test control system in buildings or situations that are not yet commissioned; (ii) the lack of appropriate detection system on-site require expensive sensor installations; and (iii) privacy concerns often prevent data sharing even when appropriate data is available. Hence, data scarcity is a core issue in the development of building service controls.

Simulation offers a solution to the above problems as data can be generated for different situations without actual implementation in the real-world and minimal privacy concerns. Therefore, a pedestrian behaviour model that simulates the movements and locations of occupants in office buildings over a complete day is sought. The pedestrian behaviour model will be a research platform that produces the data required for external applications such as the testing of a newly designed control system.

While many studies focus on simulating pedestrian movements, (operational level of pedestrian behaviour (Hoogendoorn, 2001)), only a few deal with decisions driving those movements such as activity choice (strategic level), activity scheduling, and location and route choice (tactical level). Furthermore, almost no studies integrate the three levels, as is required here, especially in the context of office buildings. There is, thus, a potential to fulfil, both, a practical need as well as a scientific gap in pedestrian modelling studies. Hence, the objective of this study is to develop a pedestrian behaviour model that integrates the strategic, tactical, and operational behaviour levels in order to generate data on occupant movements in office buildings, thereby acting as a research platform to other, external applications. To achieve the above objective the following research question is proposed:

What is an appropriate pedestrian behaviour model of occupants in office buildings that integrates the three pedestrian behaviour levels so that it can describe movement patterns of different organizations and generate high spatial and temporal resolution data on individual occupant positions?

In order to fulfil the research objective, the model development process is broken down into 4 research steps: (i) model requirements and scoping, (ii) model conceptual design, (iii) model implementation and verification, and (iv) model assessment.

Model requirements are elicited, both, directly from the stakeholders at and indirectly from the expected applications. Pedestrian movement data in offices has to be generated for applications ranging from testing the layouts of occupancy sensors in a building to developing building automation strategies using individual tracking. Thus, the output level-of-detail requires locations of individual occupants. In addition to these requirements, the model is designed based on three important guidelines: (i) low and simple model data requirements, (ii) flexibility to work with a wide range of building plans and organizations, and (iii) extensibility to connect this research platform to other workplace related studies.

Conceptual Model: Operational Level

The extensive literature available for the operational level is reviewed to select approaches suitable to model pedestrian movement behaviour for non-emergency, low density situations that are likely in office buildings. Based on this, a social forces model in continuous space that is contextualised by a navigation graph is selected to model occupant movements that include horizontal movement with collision avoidance, formation of vertical queues for waiting, and vertical movements and passages through turnstiles. The social forces model uses attractive and repulsive forces to simulate pedestrian movements. Although this approach is comparatively slow, it is chosen because it has been found suitable to model a wide range of actions even beyond those required here which can be useful if the model is to be further extended. The navigation graph guides occupants between different locations in the building. The method used generates a sparse graph using a set of rules where the nodes in the graph are derived from obstacles corners and edges are based on clear line of sight and minimum angle of deviation. For vertical movements and passages through turnstiles, pedestrians are modelled in a mesoscopic manner, such that they are represented individually but their exact location during the movement is unknown, to avoid the difficulties faced when modelling these movements using the social forces model.

Conceptual Model: Strategic Level

The strategic level decides which activities occupants would like to carry out during their time in the office building. In order to decide which activities they perform, each individual occupant is associated with a profile of attributes related to the different activities they perform. For this, first a classification of five activity types is presented so that each category can be assigned a methodology for its generation and scheduling. The classification is done so that it can accommodate various activities occurring in office buildings in most situations. Next, these activity categories are associated with the individual occupants such that the attributes of the activities performed by an occupant make up its profile. These profiles are used to prepare activity schedules for the occupants in the tactical level. Lastly, since many activities in offices are related to the organization, the representation of the organization in the model is described to generate such activities. Representing organizations in a structured manner within the model allows users to pre-define templates for various organization types from empirical observations which can be used when data is not available for the organization to be modelled. Also, completely modelling the organization structure in parallel to individual spatial movements enables connecting the model to other organization related studies such as interaction dynamics which may be driven by team membership.

Conceptual Model: Tactical Level

The tactical level model focuses on scheduling activities to be performed throughout the day. Based on a review of various approaches used to schedule movements it is found that no single approach is directly suitable for the given model requirements. Therefore, the activity scheduling approach proposed here combines the advantages of two approaches by using an activity-based framework with the less complex location-based, Markov chain model proposed by Wang et al. (2011). While the simple input needs of the Wang model fulfil the data parsimony model requirement, the activity-based framework makes the model more flexible and intuitive. In this model, time-independent activities are considered to be the states of an ergodic Markov chain and the activities are generated and scheduled by simulating the Markov chain. The associated transition matrices for each individual are derived from two inputs for each activity: (i) the average duration and (ii) either one of the following: the percentage of time spent doing it or the average duration between two instances of performing it in a single day. Using these inputs, a linear least-squares problem is setup, the solution of which gives the transition matrix associated with the attributes input. One of the advantages of using this model is that as and when more information is available it can be taken into consideration by formulating additional constraints in the linear least-squares problem.

Unlike other models using the Wang model, here it is noted that a solution is not available for all input combinations and when a solution is available there are infinitely many other solutions around it as well. Therefore, formulae are derived through empirical observations for the range of valid input combinations and solution ranges. Apart from giving an intuition of the working of the model, the formulae can be used to reduce input errors by prompting model users regarding valid input and additional constraint possibilities. In the model implementation, the Markov chains are simulated repeatedly for a maximum time period of a day whereas the stationary properties, on which the approach is based, hold true for time limiting to infinite. Thus, the limited time simulations add stochasticity to the results. It is found that the presence of this stochasticity does not eliminate the effects of using different transition matrices obtained from the same input. However, it is observed that for states with a high sojourn time, in absolute terms, the transitions from that state will produce similar results, in terms of simulated transition probabilities, for different transition matrices while those with a low sojourn time will be impacted more strongly by the transition matrix choice.

The Markov chain scheduler is supplemented by three other schedulers: (i) an event scheduler for time-dependent activities, such as meetings or lunches, which uses information from the strategic level for the duration and (preferred) starting time of activities; (ii) a movement scheduler that uses activity location and routing decisions to schedule movement times between activities; and (iii) a re-scheduler that modifies the planned schedule obtained using the above three schedulers during the execution of the operational level to accommodate, in the schedule, en-route activity and location decision changes as well as delays in movement time. Finally, a resource handler is used that interacts with the above schedulers. The event scheduler uses it to book rooms and find the availability of team activity members for meetings and the re-scheduler uses it to check the occupancy or queue lengths at various activity locations. For activity location choices and routing decisions assumptions of nearest activity location choice and shortest route selection are based on the hypotheses that different locations provide similar opportunities, occupants have full spatial knowledge of the building, and are purpose-driven as opposed to leisurely loitering.

Model Integration and Case Study Results

A conscious effort is made to keep the integrated framework as modular as possible so that, if required by a particular application, different model components can either be replaced or extended. Six modules are identified in the framework: (i) organization, (ii) individual, (iii) activities, (iv) schedulers, (v) functional spaces, and (vi) physical spaces. The model takes as input a building plan, and occupant and organization characteristics. Physical spaces in the building plan are used by the operational level to model movements from origin to destination locations through a sequence of navigation points obtained from the routing decisions at the tactical level. To create functional spaces, the physical spaces are associated with activities and capacity values for the tactical level location choice decisions. Different activity categories are scheduled by different schedulers leading to a planned schedule that is passed from the tactical level to the operational level. Through a feedback loop, this planned schedule is changed as the operational level is executed. Occupant profiles and organization characteristics input are used by the strategic level to assign activities to individuals. While the strategic level also generates episodes of time-dependent activities, the other activities are both generated and scheduled by the tactical level. This framework is implemented in MATLAB and the implementation is verified using a case study.

The case study consists of an imaginary office with a simple layout and 18 occupants. The case study is conducted to verify the implementation of the pedestrian behaviour model and to assess the model performance. In the absence of appropriate data for validation, model assessment can be carried out in two ways. First, by comparing simulated distributions against the distributions that are mathematically expected from the inputs; to this end, the first two measures for model assessment are presented: (i) duration distribution of breaks and (ii) away time distribution of the activity getting coffee. Second, assessments can also be carried out in a qualitative manner by comparing aggregated measures against expected patterns of behaviour; thus, the next two measures used are: (iii) average desk occupancy over a day and (iv) activations of a sensor placed at the entrance of the main part of the building. The first three measures are aggregated over 10 simulations whereas sensor activations are chosen for a particular day. The first two, time-related measures return expected exponential distributions but underestimate the mean values which may be due to the way they are scheduled or because a limited number of simulations were used. Both average desk occupancy and sensor activations show expected macroscopic patterns of arrivals, lunch breaks, and departures, thus, qualitatively validating the model output.

Main Contributions

This thesis makes two scientific contributions: the first is the integration of all the pedestrian behaviour levels into a single framework which can generate occupant locations throughout a day in office buildings using as input a building plan, and occupant and organization characteristics; secondly, a Markov chain based model is designed to make the generation and scheduling of activities much more straightforward whilst reducing the data requirements. This scheduler uses a previously proposed methodology and modifies it and supplements it with other schedulers to overcome its limitations. Moreover, the methodology is thoroughly analysed for valid input range, solution range, and simulation results, something that is missing in previous studies using this approach.

Even though an increasing number of people are working in offices and detection techniques in buildings have improved greatly the unavailability of data to researchers and designers impedes the development and testing of building service controls. The practical contribution of this thesis is towards the solution of this problem. Rather than address a specific issue the pedestrian behaviour model developed becomes a research platform which can generate data for various external applications. Furthermore, its data requirements are few and simple, it is flexible in terms of the building layouts and organizations it can be applied to, and lastly, it can be easily extended due to its modularity and the structured representation of the modelled organization in it.

Limitations and Future Research

The main limitations in the current research are due to the limited verification of assumptions by the Markov chain scheduler and the lack of validation of model results against empirical data. Future works can validate the model by collecting inputs, which for occupant profiles are in the form of stated preferences, running the model, and comparing results against revealed preferences obtained from sensors in the same office building. Apart from adding minor features, another important possibility for future research is to connect this model with studies on workplace interactions. For this, the pedestrian behaviour model should be able to simulate unplanned interactions which it, currently, is not able to do due to the fact that the Markov chain methodology used here assumes independence of individual schedules.

Contents

Acknowledgements v

Executive Summary vii

1 Introduction 1
1.1 Background and Motivation 1
1.2 Pedestrian Behaviour Levels 2
1.3 Research Objective and Question 2
1.4 Model Development Approach 3
1.5 Thesis Structure 4

2 Model Requirements and Scoping 6
2.1 Model Requirements 6
2.2 Scoping 7

3 Operational Level: Modelling Pedestrian Movements 9
3.1 Introduction 9
3.2 Literature Review 9
3.2.1 Model Features 10
3.2.2 Methodological Approaches 11
3.2.3 Actions Modelled 11
3.2.4 Non-functional Features 14
3.2.5 Model Selection 15
3.3 Methodology 17
3.3.1 Navigation Graph Generation 17
3.3.2 Social Forces Model 20
3.3.3 Wormholes 23
3.4 Building Representation for the Operational Level 23

4 Strategic + Tactical Level: Modelling Pedestrian Activities 24
4.1 Introduction 24
4.2 Strategic Level 24
4.2.1 Activity Classification 25
4.2.2 Activity-Occupant Association 26
4.2.3 Organization Structure 28
4.3 Activity Scheduler 29
4.3.1 Literature Review 29
4.3.2 Scheduler Framework 32
4.3.3 Markov Chain Theory 35
4.3.4 Markov Chain Scheduler 36
4.3.5 Applicability 38
4.3.6 Analysis 40
4.4 Building Representation for the Tactical Level 45
4.5 Location and Route Choice 45

5 Model Implementation and Case Study 47
5.1 Introduction 47
5.2 Framework 47
5.3 Model Input 49
5.3.1 Building Plan 49
5.3.2 Strategic Level Input 50
5.4 Operational Level 51
5.5 Strategic Level 52
5.6 Tactical Level 53
5.6.1 Event Scheduler 53
5.6.2 Movement Scheduler 55
5.6.3 Markov Chain Scheduler 55

5.6.4 Re-scheduler 59
5.7 Model Output 61
5.8 Case Study: Applications 61
5.9 Model Limitations 63

6 Conclusion 66
6.1 Summary 66
6.2 Answers to Research Questions 66
6.3 Main Contributions 69
6.4 Next Steps 70

7 References 71

A Appendix: Deriving System of Linear Equations for Transition Matrix Elements 77

B Appendix: Office Data Analysis 80

C Appendix: Preparing Building Plans 82

Figures

Figure 1: Thesis structure. Black and blue boxes show chapters and model development steps respectively. Numbers on arrows represent research questions answered in the connected chapter. 5
Figure 2: System architecture and project scope 8
Figure 3: Pedestrian movement base cases (from (Duives et al., 2013)) 13
Figure 4: Model design options 16
Figure 5: Class diagram of the operational level framework. Colours signify cohesive classes as described in section 5.2. 17
Figure 6: Navigation points (red circles) placement from corner points according to Gloor et al. (2004) (left) and Kneidl et al. (2012) (right) with convex, concave and unaligned corners. Dashed lines indicate convex angle bisectors. 18
Figure 7: Example of eliminating infeasible navigation points. Black circle and triangle indicate mid-point between obstacle and corner point, and minimum clearance distance from corner point respectively 18
Figure 8: Example of merging navigation points. Translucent red circle indicates merging position. Blue lines and arcs used to assist distance comparison. 19
Figure 9: Example of cone-based search to connect navigation points. Striped circle is the origin point, blue and black solid lines indicate prospective and final connections respectively, and black dashed lines represent the search cone. Based on (Kneidl et al., 2012). 19
Figure 10: Example describing the need to remove connections close to obstacle corners. Solid green circle is a pedestrian, translucent green circles indicate trajectory, translucent red circle indicates navigation point proximity area, and arrows represent approximate forces on the pedestrian. 20
Figure 11: Isofields around a pedestrian, in white, walking left to right, indicating the influence of other pedestrians at different distances and directions. Darker colours indicate larger influence. Two figures used to demonstrate effect of parameter. From (Campanella, 2016). 22
Figure 12: Example showing navigation points generated by a series of turnstiles. Black lines indicate obstacles, red lines turnstiles, red plus marks navigation points, and purple, smaller crosses wormhole nodes. 23
Figure 13: Class diagram of activities and association with individuals. Colours signify cohesive classes as described in section 5.2. 27
Figure 14: Class diagram of organization representation in the model. Colours signify cohesive classes as described in section 5.2.... 28
Figure 15: Class diagram of schedulers used in the model. Colours signify cohesive classes as described in section 5.2. 34
Figure 16: An example Markov chain 35
Figure 17: Workflow of Markov chain generation and simulation 38
Figure 18: Occupancy intervals fitted to an exponential distribution in single-occupant offices (top, from (Luo et al., 2017) (duration in hours); bottom-left, from (Wang et al., 2005); bottom-right desks in an open plan office) 40
Figure 19: Logistically distributed utility growth against time interval for recurrent activities performed with different daily frequencies and occurrences (from (Tabak, 2008)) 40
Figure 20: Objective function values for transition matrix generation with different combinations of 3-state steady state probabilities and sojourn times 41
Figure 21: Transition matrix elements range for $\pi = 0.25\ 0.35\ 0.40; \tau = [1\ 1\ 2]$. Horizontal and vertical red lines describe the threshold of acceptable objective function value (10^{-10}) and the maximum value the transition matrix element can take respectively. Red dot represents the solution found by the MATLAB solver. 42
Figure 22: Variation of transition element (red) and its effect on other elements for $\pi = 0.25\ 0.35\ 0.40; \tau = [1\ 1\ 2]$ 43
Figure 23: Simulated transition probabilities for four different Markov chains generated from $\pi = 0.25\ 0.35\ 0.40; \tau = [1\ 1\ 2]$ 44
Figure 24: Simulated transition probabilities for four different Markov chains generated from $\pi = 0.25\ 0.35\ 0.40; \tau = [10\ 10\ 20]$ .. 44
Figure 25: Class diagram of building representation for the tactical level. Colours signify cohesive classes as described in section 5.2. 45
Figure 26: Class diagram of the complete pedestrian model. Similar colours indicate cohesive classes. 48
Figure 27: Pedestrian behaviour levels and their interaction based on (Hoogendoorn, 2001). Arrow 1: Time-dependent activity episodes and occupant profiles for time-independent activities; arrow 2: feedback from tactical level (does not occur in this model); arrow 3: schedule of movements; arrow 4: feedback from operational level for re-scheduling 49
Figure 28: Case study: Building plan – physical spaces as obstacles and positions, and text annotations (left; distance marks in meters) and as functional spaces with functions embedded as layers in the AutoCAD plan (right) 50
Figure 29: Case study: Navigation graph of the office; blue lines are connections between locations (orange dots) to navigation points (red crosses); green lines are connections between navigation points 51
Figure 30: Flow diagram of the event scheduler for planned activities in a day 54
Figure 31: Flow diagram of the event scheduler for time-window activities in a day 54
Figure 32: Case study: Event schedule of agent 12. Dashed lines represent gaps of available time. 55
Figure 33: Flow diagram of the Markov chain scheduler 57
Figure 34: Case study: Markov chain agenda for agent 12 from simulation of transition matrix 58

Figure 35: Case study: Full planned schedule for agent 12. Dashed lines show the parts of the final schedule that come from the Markov chain agenda. ... 58
Figure 36: Flow diagram for on-the-fly location and activity choice by the re-scheduler ... 60
Figure 37: Case study: Rescheduling due to operational delay (agent 12; left) and activity rejection (agent 17; right). Activity rejection continues as after returning to the desk there is no specific movement to the next activity which is also at the desk. ... 61
Figure 38: Case study: Distribution of break durations (left) and coffee away times (right) ... 62
Figure 39: Case study: Desk occupancy over 10 model runs ... 62
Figure 40: Case study: Sensor field (dashed lines) over turnstiles ... 63
Figure 41: Case Study: Sensor activations and possible reasons for turnstile sensors ... 63
Figure 42: Example of defining areas (cyan) around desks to check occupancy status ... 80
Figure 43: Meeting room data analysis results (row-wise): (a) meeting start time minute, (b) meeting start time hour, (c) meeting duration in minutes, (d) meeting room capacity ... 81

Tables

Table 1: Model requirements and priorities 7
Table 2: Summary of activity types in the conceptual model 26
Table 3: Pros and cons of different movement scheduling approaches 33
Table 4: Case study: Team definitions 50
Table 5: Case study: Team activity definitions 50
Table 6: Case study: Activities modelled 51
Table 7: Case study Attributes for Markov chain activities 51
Table 8: Case study: Attributes for time-window activities 51
Table 9: Case study: Member agents of team activities 52
Table 10: Case study: Transition matrix for agent 12 58
Table 11: Building plan AutoCAD layers' properties 82

1 Introduction

As an increasing number of people work in offices, there is a growing need to provide better control strategies for more energy efficient buildings and higher user well-being. However, research and development efforts for this are often stymied by the unavailability of data for testing and evaluation. To this end, it is proposed that a pedestrian behaviour model be developed to simulate movements in offices and thus generate the data that is otherwise scarce. Such a model would act as a research platform that produces data for external applications. This chapter introduces the research objective, questions, and approach to developing this pedestrian model and sets up the remainder of this thesis by discussing its structure and the model requirements and project scope.

1.1 Background and Motivation

Around the world, especially in developed countries, an increasing proportion of the workforce is shifting to tertiary economies that often operate in offices buildings (Al Horr et al., 2016). For example, in the Netherlands, more than a third of the workforce works in offices (Bergs, 2002). In a national time-activity pattern survey of the U.S. it was found that those of working age spent 8.42% of their time in offices/factories with a mean duration of 414 minutes (Klepeis et al., 1996). Moreover, in the U.S., amongst commercial buildings, offices make up the highest number of buildings, have the largest floorspace, and are occupied by the highest number of workers (U.S. E.I.A., 2012). It is clear from these statistics that humans spend a large amount of time in office buildings.

Consequently, these buildings consume a large proportion of resources for building services such as lighting, heating, ventilation, and air-conditioning (HVAC), and vertical transportation. For example, in the U.S. commercial buildings make up about 20% of the overall energy consumption of the country (U.S. E.I.A., 2017a) of which offices are the second highest users at 14% closely following the first, mercantile and other services at 15% (U.S. E.I.A., 2017b). Thus, building services in offices need to be optimized to achieve energy efficiency and reduce pressure on, what are often, limited resources. Conversely, since people spend so much time inside buildings, these services, in turn, impact their comfort and productivity and therefore, better control of indoor environmental conditions is necessary (Al Horr et al., 2016). Such building services are not only dependent on physical factors such as available daylight, orientation and weather but also on the behaviour of occupants in the building. For example, the demand for HVAC and lighting depends on the spatial and temporal distribution of occupants within the building, and various factors, such as type of activity being carried out, determine the ideal indoor environmental conditions.

Philips Lighting seeks to use lighting control systems to increase energy savings as well as improve the well-being of occupants in office buildings to which end they would like to research these systems and evaluate their performance before implementation. As communication and detection capabilities grow, the potential brought by newer technologies, such as connected lighting and indoor tracking using visible light communication, for more complex controls to provide smarter building services need to be researched and developed. Furthermore, lighting controls involve several designs decisions such as the spatial level of occupancy detection (building, zone, or desk level), the detection technique or sensor type, and the intelligence level (reactive, anticipatory, or proactive) (de Bakker et al., 2017) and therefore various combinations need to be evaluated before implementation in order to be able to take optimal decisions.

Since building services are dependent on the behaviour of occupants in the buildings, it is an input for the research and evaluation of these services. According to the control system under question, occupant behaviour required as input may range (from simple to complex) from total building occupancy variation in a day, to the sequence of locations visited by occupants at different times, to the likelihoods of moving towards a given direction at different points in the building. Given that recent developments in detection technologies allow the tracking of individual occupants up to a resolution as high as 10 cm (from discussions with Philips Lighting employees), the research of control systems intending to fully exploit these capabilities also require data describing individual occupant movements in the building at a similar high level-of-detail. When such input is available, it can also, naturally, be aggregated as required for systems using more aggregated measures such as occupancy of different zones. For the research of control strategies in general, occupant behaviour inputs would be required from different offices to test the controls under different movement patterns whereas evaluating the performance on a specific building would require input for occupants of that particular building. It should be noted that during development and evaluation, data over large periods of time are used for testing whereas during the operational phase, control systems would make use, at least in part, of information collected in real-time from sensors installed in the respective buildings.

Inputs for the development phase of control systems may be collected through recording locations and movements of occupants in office buildings over large periods of time but often obtaining such data is not straightforward. The first issue arises when there is a need to evaluate performance in buildings that are not yet commissioned since there is no possibility of data collection. This is also the case when the impacts of rearrangements or other situational variations that do not exist need to be measured. When the buildings are already inhabited, the detection and data collection systems available on-site may not be appropriate and could require expensive installation of temporary sensors for data collection. Especially, when data from traditional passive infrared (PIR) sensors is not sufficient to test or evaluate the impact of newer technologies. Finally, even when suitable data is available privacy concerns may prevent sharing sensor data generated in offices with external companies that develop and implement the control systems. Hence, data scarcity is a core issue in the development of building service controls.

Simulation offers a method to solve the above problems as data can be generated for different situations without installation of sensors and with minimal privacy concerns. While data is still required to verify assumptions used in the design of a simulation model, as well as for its calibration and validation, once developed fewer inputs can be used to generate required data for a variety of situations and for long periods of time. However, conventional occupancy models (e.g. (Wang et al., 2005)) are likely to fall short of the data requirements of smarter controls that are made possible by the availability of newer technologies. Therefore, a pedestrian behaviour model that simulates the movements and locations of occupants in office buildings over a complete day is sought by Philips Lighting. By simulating a number of working days, such a model would produce the data required for the development and evaluation of different control strategies. In this manner the pedestrian behaviour model would be a research platform that produces the data required for external applications such as the testing of a newly designed control system.

The pedestrian behaviour model would not only simulate movements but also individual decisions driving these movements such as those regarding where to be at different times of a day. While many studies focus on simulating pedestrian movements, only a few deal with decisions driving those movements (Tabak, 2008). Furthermore, Schadschneider et al. (2009) suggests that the next step in research would be to develop an integrated model that can simulate complex situations for longer periods of time. There is, thus, a potential to fulfil, both, a practical need as well as a scientific gap in pedestrian modelling studies.

Next, first a brief background on pedestrian behaviour levels is discussed (section 1.2) to be able to place the subsequent sections on research objective and question (section 1.3), and the model development approach (section 1.4) in context. Finally the thesis structure is presented in section 1.5.

1.2 Pedestrian Behaviour Levels

Hoogendoorn (2001) proposes a classification of pedestrian behaviour levels that is widely used in pedestrian modelling: strategic, tactical, and operational. These levels represent decisions made by pedestrians and range from long-term at the strategic level to instantaneous at the operational level. The strategic level models which activities pedestrians want to carry out, thus giving them a reason to move. The tactical level schedules these activities, chooses locations where they are performed, and decides the route to be taken to reach these locations. Finally, the operational level uses this schedule of movements and routing decisions to simulate walking behaviour which involves instantaneous decisions by the pedestrians to, for example, avoid each other as well as other static obstacles. However, the three levels do not only interact in this hierarchical manner but there is also feedback interaction wherein operational level decisions affect tactical level behaviour and tactical level decisions affect strategic behaviour.

1.3 Research Objective and Question

Based on the above background, the objective of this study is to develop a pedestrian behaviour model that integrates the strategic, tactical, and operational behaviour levels in order to generate data on occupant movements in office buildings, thereby acting as a research platform to other, external application layers. To achieve the above objective the following research question is proposed:

What is an appropriate pedestrian behaviour model of occupants in office buildings that integrates the three pedestrian behaviour levels (Hoogendoorn, 2001) so that it can describe movement patterns of different organizations and generate high spatial and temporal resolution data on individual occupant positions?

Hence, the model development process involves selecting, developing, and integrating appropriate conceptual models for the three pedestrian behaviour levels. The pedestrian behaviour model should be able to simulate movements in different offices which may have different movement patterns and return individual occupant positions at fine spatial and temporal resolution. This data will be used by external applications that are not part of this research. In order to assess the appropriateness of the model, it should fulfil the model requirements that are to be collected from the stakeholders. Moreover, after verifying its functionality, its performance should be judged through some form of validation.

The next section presents the model development approach used to fulfil the research objective. Moreover, to help answer the research question it is broken down into sub-questions representing research steps which in turn are split into supporting sub-questions.

1.4 Model Development Approach

The model development approach is broken down into 4 research steps: (i) model requirements and scoping, (ii) model conceptual design, (iii) model implementation and verification, and (iv) model assessment. Next, these steps are explained alongside their respective sub-question and their supporting sub-questions.

1. *Model Requirements and Scoping*: In this step, the objective is to derive model requirements and project scope from the stakeholders at Philips Lighting. The model requirements will create the basis for and guide the design of the pedestrian behaviour model. The pedestrian behaviour model is to act as a research platform producing data for external applications. Therefore, first possible applications need to be known which will lead to possible model inputs and the requirements about the model outputs. Specifically, the output level-of-detail is an important factor in deciding which approaches can be used to model pedestrian behaviour. Since this is a research platform rather than a model for a very specific problem, all possible applications that can be connected to it are not decided. Therefore, for possible applications, pedestrian behaviours that stakeholders specifically wish to be represented in the model should also be taken into account.

   Use cases and applications are obtained through interviews with various interested persons at Philips Lighting. Since the topic is relatively novel for the company, the approach selected for the elicitation of requirements and potential applications involves explaining the basics of the topic and possible solution directions for present applications. Further, several requirements are developed introspectively (by the author), based on the stated use cases, and then put forward for approval by the stakeholders. Finally, the verification of requirements is also carried out throughout the development phase of the project making the requirement elicitation procedure similar to the linear and iterative combination described in Martin et al. (2002).

   Sub-question *1(1)*:

   What are the stakeholder requirements for the pedestrian behaviour model so that it can generate data for present and potential external applications?

   Supporting sub-questions:
   a. What are the expected applications of the model?
   b. What are the potential model inputs and outputs?
   c. What should be the level-of-detail of the model output?
   d. Which pedestrian behaviours should be specifically included in the model?

   Furthermore in this step, the scope of the project is also decided. This defines which peripheral features need to be included in the scope of the project.

   Sub-question *1(2)*:

   Which features, peripheral to the pedestrian behaviour model are within the scope of this project?

2. *Model Conceptual Design*: The conceptual design of the model is at the core of the model development. A conceptual model is developed for each of the three pedestrian behaviour levels keeping in mind the model requirements derived in the previous step. The first step is to review the state-of-the-art for all three pedestrian behaviour levels in order to evaluate various approaches and select those that are most appropriate given the model requirements and their compatibility with others. For each level, several approaches may have to be selected to bring about a complete conceptual model of that level. Since it is unlikely that these approaches can be directly adopted for the integrated model and given requirements, their limitations or drawbacks have to be identified. Having identified these limitations, the approaches are modified and additional components as appropriate. Finally, by bringing together the approaches of a level the conceptual model for that level is complete.

   Sub-question *2*:

   Based on the model requirements, which approaches can be brought together to produce a conceptual model for each of the three pedestrian behaviour levels?

   Supporting sub-questions:
   a. From the state-of-the-art, which approaches and combinations thereof are suitable to model different parts of each of the pedestrian behaviour levels?
   b. What are the limitations and drawbacks of these approaches given the model requirements and the need to integrate all behaviour levels?
   c. Which modifications or additions can be made to the approaches from literature in order to suit the model requirements?

3. *Model Implementation and Verification*: In this step, the conceptual models of the different pedestrian behaviour levels are brought together under a common framework. For this the interfaces between the different conceptual models have to be defined. Next, each module in the framework is converted to an executable computer model and the implementation is verified to check its functioning.

   Sub-question *3*:

   What should be the framework to integrate the conceptual models of all three pedestrian behaviour levels and implement them in a computer language?

   Supporting sub-questions:
   a. What are the various interfaces between the conceptual models of the three pedestrian behaviour levels?

4. *Model Assessment*: Finally, in this last step the performance of the model has to be assessed. Since this model is being developed as a research platform than to model a particular situation some generic assessment criteria have to be proposed. Due to unavailability of time and appropriate data, calibration and validation of the model on the basis of real-world data is not done. In its place, a case study of an imaginary office is used for qualitative validation against expected patterns of behaviour. Lastly, model limitations and recommendations for future implementations are made.

   Sub-question *4(1)*:

   Is the pedestrian behaviour model able to produce sufficiently realistic outputs as per the model requirements?

   Supporting sub-questions:
   a. Which metrics can be used to assess the performance of the pedestrian behaviour model?

   Sub-question *4(2)*:

   What are the limitations of the pedestrian behaviour model?

1.5 Thesis Structure

Figure 1 shows the thesis structure: the model development steps dealt with and the research questions answered in each chapter. After this introduction, the next chapter discusses the first research step: model requirements and scoping (chapter 2) which guide the design of the conceptual model and its implementation. The next two chapters are dedicated to the design of the conceptual model and answer the same research questions but for different behaviour levels. For this first the operational level that deals with modelling pedestrian movement behaviour is discussed in a single chapter (chapter 3). This chapter thoroughly reviews the extensive state-of-the-art available, selects appropriate models, applies suitable modifications and describes the model design. Since many decisions at the strategic and tactical levels are intertwined these two levels are discussed together in chapter 4. However, while several design decisions for the levels are made together they are still described separately. For these levels, too, appropriate literature is reviewed to place the current application in context with past approaches and synthesize approaches from different perspectives to understand their pros and cons. While the strategic level is concerned with the representation of organizations, individuals, and the activities performed by them the tactical level mainly deals with scheduling activities for individuals with lesser focus on location and route choice. The approaches for both levels are described separately and the methodology used for scheduling activities at the tactical level is analysed in depth. Throughout the discussion of the conceptual models in this and the preceding chapter, class diagrams of different parts of the model are presented for better understanding of the modules. The conceptual models for the different pedestrian behaviour levels are put together in chapter 5 under a single integrated and modular framework by connecting the class diagrams of different modules through appropriate interfaces. The implementation of the model is described alongside a case study for easy understanding as well as verification of the implementation. For model assessment, results from the case study are presented to qualitatively validate the model and lastly, limitations of the pedestrian behaviour model are discussed. The thesis is concluded (chapter 6) by answering the research questions, followed by a discussion on the main contributions of this study, and a note on possible future directions for this study.

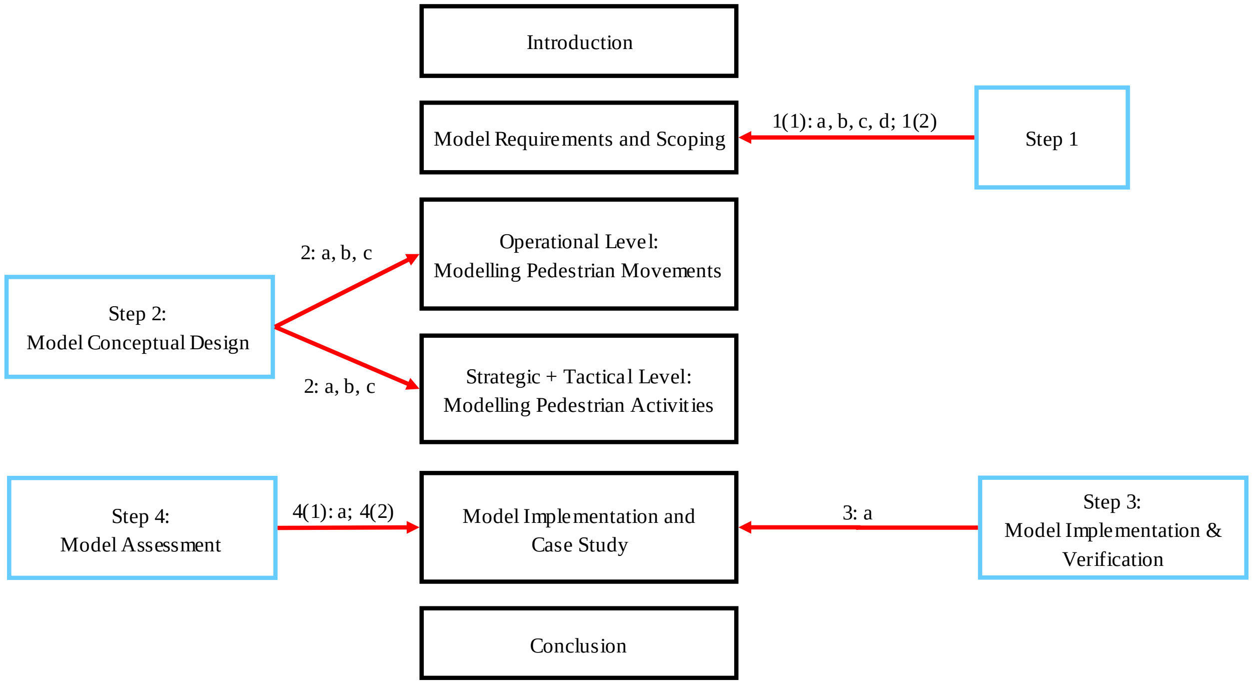

Figure 1: Thesis structure. Black and blue boxes show chapters and model development steps respectively. Numbers on arrows represent research questions answered in the connected chapter.

2 Model Requirements and Scoping

This chapter discusses the model requirements for the pedestrian behaviour model. Collecting model requirements is an essential part of the model development process as these requirements form a basis for design decisions. Here, the requirements are elicited through discussions with stakeholders regarding current and potential applications for which the pedestrian behaviour model will produce data. From these applications as well as direct inputs from the stakeholders the possible inputs, the required outputs, and pedestrian behaviours to be represented in the model are obtained. Based on this a list of functional and non-functional requirements is prepared. At the end of the chapter, the research platform with the pedestrian behaviour model and peripheral components is discussed along with scoping of the work to be included in this project.

2.1 Model Requirements

Model requirements or specifications are defined as statements which translate or express a need and its associated constraints and conditions (IEEE, 2011), guide the design and implementation and are, therefore, integral to the development of the model (Aurum and Wohlin, 2005). Zowghi and Coulin (2005) list a series of activities for the elicitation of requirements: understanding the application domain; identifying the sources of requirements; analysing the stakeholders; selecting the techniques, approaches and tools to use for elicitation; and eliciting the requirements from stakeholders and other sources. Here, the stakeholder is Philips Lighting and the sources of requirements are the employees with applications that can make use of the data that can be generated by the pedestrian behaviour model. Requirements are elicited mainly through interviews and discussions but are also developed introspectively based on the stated use cases.

Next, the range of current and future applications that will make use of the pedestrian behaviour model as a research platform is discussed. Based on these, next, possible inputs are discussed. From this the types of inputs the model users are likely to have access to and which types might be difficult to obtain can be understood. This is useful to guide the model design decisions accordingly. The output requirements consist of the type of outputs, the level-of-detail, and data storage. From these, the level-of-detail of the output is an important property in selecting modelling approaches. Specific behaviour to be modelled, obtained as direct input from the stakeholders, is then discussed followed by some specifications for the executable model to be implemented. From these discussions a list of requirements and their priorities is presented to aid the model development process and clearly check whether all requirements are being met.

Applications: The pedestrian behaviour model developed here will be used to generate data for various applications ranging from selecting optimal spatial configuration of sensors for switching lights on and off automatically to developing predictive building automation controls based on individual tracking. Moreover, the sensors deployed in reality can be simple presence detectors such as PIR sensors, indoor positioning using the lighting grid based on visible light communication with handheld devices, or Doppler- and image-based sensors that can detect individual positions. The model should be such that data for all these different detection techniques can be generated. Furthermore, the model should be flexible with respect to handling different types of offices (closed, open plan, or flexible), building plans and organizations and since the model acts as a research platform it should be easily extensible such that it can connect with other workplace related studies.

Inputs: Building plans are necessary to define the boundaries of walking and defining the functions of different spaces. It is expected that computer-aided design (CAD) files of building plans are easily available and hence the model should be able to make use of them directly. However, as discussed in the introduction revealed preference data on movement patterns from installed sensors are unlikely to be available and therefore their use as input is discouraged. Since revealed preference data is unavailable, it is expected that occupant and organization characteristics will be inputs to the model. Even these inputs should be as few as possible as extensive data collection is not expected to be viable. Moreover, the inputs should be intuitive and easily estimated or filled in by office building occupants.

Output: The level-of-detail of the output is an important aspect of the model that determines the modelling approaches that can be adopted. From the applications and the real-world detection techniques for which data needs to be generated from the model, it can be seen that the spatial level-of-detail ranges from individual location to zonal occupancy. Therefore, the model should output individual locations with a high spatial resolution of about 10 cm which is found to be, from interviews with Philips Lighting employees, the highest resolution of sensors used. Given the individual locations, it is straightforward to calculate and generate lower resolution data such as zonal occupancy. Furthermore, a methodology to store the data generated by the model should be present but it does not need to be in a specific format as long as there is appropriate documentation explaining the same.

Model: The pedestrian behaviour model should be able to simulate movement patterns in office building for a complete day. Moreover, it should be done at sufficient detail so that the position of the person can be known. For example, it is not enough to know that a person is working in a particular room, rather the assigned or chosen desk of the person must be known. In order to develop smart controls, applications may also be interested in knowing the different individual and interpersonal activities carried out by occupants throughout the day at different times. Further, based on direct stakeholder inputs, the following movement actions are required to be represented in the model:

- Horizontal movement with collision avoidance: Pedestrians should be able to walk through a facility without colliding into obstacles or other pedestrians.
- Trajectories should be smooth and 'natural looking'.
- Waiting: Pedestrians should be able to form queues in front of services such as coffee machines.
- Walking in groups: Pedestrians should be able to form groups, that is, two or more people walking together.
- Vertical movement across floors: Pedestrians should be able to go from one floor to another using vertical movement facilities such as stairs, elevators, and escalators.

Regarding the model implementation, MATLAB is the preferred tool so that existing sensor models and other simulation framework in Philips Lighting can be easily connected to the model. The stakeholder requirement is that model running time be faster than real-time. However, since in fact the applications would require running the model for several (simulation) days in order to generate sufficient data, the model running time should actually be a small factor of the simulated time.

(Aurum and Wohlin, 2005) define functional requirements as those that describe "what the system will do" while non-functional requirements are those that place "constraints on the types of solutions that will meet the functional requirements" such as accuracy, running time, extensibility, etc. Based on the above discussions, Table 1 lists the functional and non-functional requirements alongside their importance which is helpful in deciding priorities when designing the model. These priorities are based on the author's understanding from various discussions with the stakeholders. The priorities which are categorised as high, medium, and low indicate which requirements are strict, which have some leeway, and which are more a wish than a requirement, respectively. The three, high priority, non-functional requirements make up important guidelines that are to be considered during the design of the conceptual models.

Table 1: Model requirements and priorities

Model Requirements	**Priority**
Functional Requirements	
Layered approach with interchangeable inputs and applications	High
High output level-of-detail with individual positions and smooth trajectories	High
Model functionality: Horizontal movement with collision avoidance	High
Model functionality: Waiting behaviour	Medium
Model functionality: Walking in groups	Medium
Model functionality: Vertical movement across floors	Low
Model functionality: Represent individual and interpersonal activities	High
Non-functional requirements	
Flexibility to work with different office types, building plans, and organizations	High
Extensible model to connect with other workplace related studies	High
Low and simple data requirements for the model	High
Preferred tool of implementation: MATLAB	Medium
Fast execution time	Medium

2.2 Scoping

While the pedestrian behaviour model is the core of the research platform, several peripheral components also exist. This section discusses which additional components are within the purview of this project and which are not. The ecosystem of the research platform is shown in Figure 2 with the model as a black box along with the required input, output and peripheral tasks related to the core assignment. To describe the project scope, different components are outlined so that those with solid, dashed and no outlines are fully, partially and not within the purview of this project respectively.

Inputs: For CAD building plans, ideally, geometric information and attributes should be extracted automatically by the model from these files but often drawings from one drafting team are quite different from those of another making this difficult. Therefore, drawing rules that enable the required input to be easily extracted by the model should be defined. Further, either a programme or methodology must be developed to read in commonly used CAD files. Apart from CAD files, users may also wish to use a graphical user interface (GUI) to enter or modify the required inputs. However, it is out of scope of the project but a basic input interface or methodology should be created.

Outputs: Regarding the output interface, a simple graphical output of the pedestrian behaviour model is required to visualize the movements in the office for demonstration purposes. The visualization of the output is also important for the verification and face validation of the model. An output GUI to allow tasks such as zooming in to a particular area, following a particular person and the possibility to interactively hide or show particular details may be considered later but is currently considered to be out of scope.

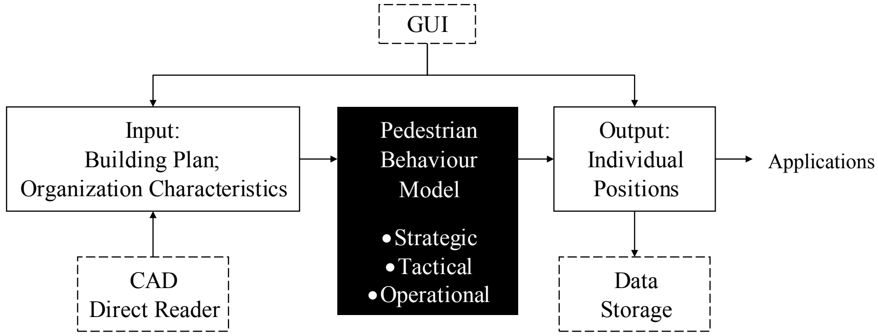

Figure 2: System architecture and project scope

3 Operational Level: Modelling Pedestrian Movements

The operational level of the pedestrian behaviour hierarchy is concerned with short-term movement related decisions. This chapter develops an operational level model which accepts as input a building plan and a schedule of activities, and returns realistic individual pedestrian trajectories. Given the large number of studies at this behavioural level, a detailed literature review is carried out to identify suitable models based on the model requirements resulting in the selection of a social force model in continuous space contextualised by a network of navigation points.

3.1 Introduction

The operational level lies at the bottom of the pedestrian behaviour hierarchy accepting activity schedule, location choices, and routing decisions from the tactical level and making related short-term choices involving locomotion of pedestrians (Hoogendoorn, 2001). From a pedestrian's perspective some wayfinding tasks, especially under exploratory conditions, and choices between activity location in close proximity based on availability may also occur at this level of short-term decisions. However, while modelling such behaviour, these decisions may be modelled either deterministically at the tactical level or dynamically by allowing information from the operational level to affect changes in tactical level decisions on-the-fly. Thus, from the viewpoint of modelling, the operational level, only executes the schedule decided at the tactical level. The exact output type and level-of-detail depend on the choice of the pedestrian model.

In the past four decades, the operational level has been studied extensively and a large number of models has been developed either for general use or to model specific situations or behaviours. Models have focussed on large public or semi-public spaces such as transportation hubs (Campanella, 2016; Daamen, 2004; Hoogendoorn and Bovy, 2004b) as well as indoor spaces such as museums (Turner and Penn, 2002) although the bulk of the models are applied to the former. Similarly, most operational level models have been developed to analyse evacuations (Gwynne et al., 1999) or crowd disasters (Helbing et al., 2007) since longer-term decisions play a reduced role in emergency situations than when simulating longer times and more complex scenarios (Schadschneider et al., 2009). To take advantage of the vast literature available at this behaviour level, an extensive review of past developments is undertaken to identify model components suitable for application in offices under non-emergency, low-density conditions. This is done by deriving a structure for the classification of models such that the model requirements can be directly used to make design decision for the operational level model (section 3.2).

Having identified promising methodologies from the literature review, the selected model components and suitable modifications are described in section 3.3. Finally, in section 3.4, the representation of the building for the operational level is discussed. Given a building plan, a schedule of origins and destinations, and routing decisions, the model components together are able to simulate pedestrian movements.

3.2 Literature Review

Several existing literature reviews on operational level models structure models based on certain criteria. (Gwynne et al., 1999) classify early evacuation models based on the representation of space and population, and underlying behaviour paradigms assumed. Kuligowski (2004) adds more attributes to include routing behaviour, movement methodology, validation of the model, and more practical indicators such as the ability to import CAD files and visualize results. Further some specific features are included in the review such as the possibility to define groups, model elevator use, or change detailed pedestrian behaviour such as impatience. Both Gwynne et al. (1999) and Kuligowski (2004) focus on models for fire evacuations. Schadschneider et al. (2009) and Zheng et al. (2009) classify models on the basis of commonly used methodological approaches. Zheng et al. (2009) further cross-classify these models on features such as population homogeneity and scale, continuity of space and time, modelling situations, and typical phenomena that the model describes. Zhou et al. (2010) focus on model application and categorize models according to the number of pedestrians and timescale that the model is suitable for. Further classification is based on the representation of pedestrians and factors, such as physical, social, and psychological, assumed to affect movements. Campanella (2016) uses the PAGE – Perception, Action, Goals, and Environment – concept by Russell and Norvig (1995) to classify these methodological approaches and their combinations into behavioural paradigms and describe features of the agents therein. Concentrating on movement behaviour of large crowds, Duives et al. (2013) evaluate whether models are able to represent specific movement cases and self-organizing phenomena.

Based on these reviews, here, operational level models are described using the following properties: (i) model features, (ii) methodological approaches, (iii) actions modelled, and (iv) non-functional features. In the following sub-sections, each property is discussed by expanding on the various design decisions within it and reviewing the choices made in previous studies. Once this general overview of all the options available in the different properties is completed, the requirements for the current model are used to make appropriate choices for the model design which, in turn, can be used for the selection of model components.

3.2.1 Model Features

Model features describe how the real world is represented in the model. The main model features at this level include: (i) individual representation of pedestrians or scale of the model, (ii) representation of space, and (iii) homogeneity of pedestrians. Depending on the model application some representations may be more suitable while some may be unusable.

Scale: Scale represents whether pedestrians are modelled individually or collectively and consists of three categories: (i) macroscopic, (ii) microscopic, and (iii) mesoscopic. Collective representation of pedestrians, or macroscopic models, use aggregated parameters such as speed, flow and density. These models are useful to analyse the spatial distribution of these parameters in pedestrian facilities under different conditions but by nature are unable to output individual pedestrian behaviour. Moreover, due to this limitation they cannot represent or take into account a number of observed phenomena although recent developments are being made towards remedying this (Hoogendoorn et al., 2014). Since they are unable to output individual positions, these models are not of interest and are, therefore, not discussed further in this review. Microscopic models, on the other hand, describe individual pedestrian behaviour, that is, each pedestrian has a specific position. Since individual pedestrians are modelled, microscopic models can be more precise (Campanella, 2016; Hoogendoorn et al., 2014) and represent a wider range of individual pedestrian behaviour as well as emerging collective phenomena. Yet, microscopic models are generally slower and since they represent individual movements, rather than aggregate parameters, they are more difficult to calibrate. The last category within this classification, mesoscopic models, covers a broader range of definitions than the above two as definitions in literature differ based on the modelling context (see (Ijaz et al., 2015) versus (Biedermann and Borrmann, 2016)). However, what these models primarily aim for is to balance the trade-off between computational cost and the ability to simulate required details (Abdelghany et al., 2012). It is also possible to include, at this level, hybrid models that make such a trade-off by spatially dividing the pedestrian facility and applying microscopic models where accuracy is required and using a macroscopic model otherwise (Biedermann and Borrmann, 2016; Xiong et al., 2010). In this review, the focus will remain on microscopic models as they describe the individual behaviour described in the model requirements.

Space Representation: Depending on the methodological approach, previous studies have used 3 types of space representations to describe walkable areas: (i) networks, (ii) discrete cells, and (iii) continuous space. Networks represent the walking facility as nodes and edges where edges represent the space available for walking and nodes represent activity locations, edge intersections at corners, and entry/exit points (Borgers and Timmermans, 1986; Daamen, 2004; Løvås, 1994). Walking behaviour in networks is simulated in one-dimension along the edges in these models (Bierlaire et al., 2003). These models are able to take into account the effects of congestion on such edges, thus describing aggregated parameters, but are unable to explicitly model pedestrian interactions which is why some reviews classify these models as mesoscopic (Abdelghany et al., 2012; Campanella, 2016) or even macroscopic (Hoogendoorn et al., 2014). Cellular discretization relies on division of the walking facility into varying levels of resolution has also been used. At the lowest resolution, several pedestrians are accommodated in a cell (Asano et al., 2007) while at the highest resolution a single pedestrian covers more than one cell in order to represent 'compressibility' of pedestrians in high density crowds (Song et al., 2006). Most cellular representations, however, discretize the space such that each cell can contain only one pedestrian. Such cells are usually squares of side 40-45cm (Bierlaire et al., 2003; Blue and Adler, 1999) but they may also be of other shapes such as hexagons (Davidich et al., 2013). In such models, pedestrian can only occupy certain locations and furthermore, the shape of the cell and the farthest cell that can be reached in a time step restrict the angular movements of pedestrians to certain values; for example, a model using a rectangular lattice and allowing movements to the Moore neighbourhood of range 1, the available angles are only multiples of 45 degrees. Larger neighbourhoods would, naturally, decrease this value but movements would still be constrained to a set of angular values (Gloor et al., 2004). Continuous spaces, however, allow pedestrians to occupy any location in a walkable area and turn at any angle, and do not have any constraints on the movement resolution in itself. However, it should be noted that while the other two representations provide some context for routing namely, node-to-node or cell-to-cell, continuous space does not. In literature, such context has been provided in the form of overlaying networks (Gloor et al., 2004; Karamouzas et al., 2009; Kneidl et al., 2012; Sud et al., 2008), Bezier splines (Nasir et al., 2014; Zhou et al., 2010), discrete visual fields (Moussaïd et al., 2011; Turner and Penn, 2002), and potential fields (Gloor et al., 2004; Hoogendoorn, 2001). From the above space representations, only cells of high resolution and continuous spaces are able to represent unique trajectories of individual pedestrians produced through their interaction with obstacles and other pedestrians.

Population Homogeneity: Pedestrians may be considered as homogeneous or heterogeneous individuals (microscopic) or groups (macroscopic) depending on the complexity of individual behaviour to be modelled. Individual characteristics may control walking behaviour either (i) directly or (ii) indirectly. Direct control characteristics are attributes such as free speed or acceleration but indirect control characteristics may be divided into (i) personal, (ii) psychological, or (iii) cultural behavioural characteristics (Zhou et al., 2010). Personal characteristics include gender, age, body size and ability. For example, Versluis (2010) reports effects of various personal characteristics on walking behaviour from literature and also experimentally determines significant effects of gender on passing and interaction behaviour. Psychological characteristics represent effects such as travel purpose and urgency. For instance, Versluis (2010) finds that pedestrians in a hurry tend to take larger lateral but smaller longitudinal evasive maneuvers when interacting with other pedestrians than pedestrians not in a hurry. Finally, cultural characteristics dictate what is 'normal' for a pedestrian in a particular context. Cultural effects are usually described by personal and psychological characteristics such as the tendency to give more space to the elderly. However, some effects such as personal space or the tendency to pass on a particular side (Versluis, 2010) are strictly cultural. While most models vary free-speed and many methodological approaches in microscopic models allow varying other parameters in order to reflect reality as closely as possible (Campanella, 2016), it should be noted that introducing pedestrian heterogeneity on several fronts makes it difficult to calibrate the model while not having a large impact on the overall behaviour of pedestrians, especially, in low density, non-emergency situations.

3.2.2 Methodological Approaches

In order to review various methodological approaches, they are grouped into four categories: (i) cellular automata, (ii) force-based, (iii) velocity-based, and (iv) others.

Cellular Automata: In the basic implementation of this approach, space is discretized into cells such that each pedestrian occupies a particular cell and in fixed time steps their positions are updated synchronously. Pedestrian movements are based on rules defining transition probabilities to neighbouring cells. The probabilities depend on desired direction and the status of neighbouring cells which may be either empty or occupied by other pedestrians or obstacles (Blue and Adler, 1999; Schadschneider et al., 2009). In order to include longer range interactions, Burstedde et al. (2001) introduce floor fields which modify local transition probability rules. Static floor fields guide pedestrians towards their goal and are equivalent to assigning different transition probabilities to each cell while dynamic floor fields are used to model attraction to pedestrians of the same group or herding behaviour using virtual trails left behind whenever a pedestrian walks off a cell.

Force-based Models: These models make use of physical forces to induce pedestrian movements. Okazaki and Matsushita (1993) use magnetic forces while Helbing and Molnár (1995) consider Newtonian forces. In these models, attractive and repulsive forces guide movements. Attractive forces act towards destinations and pedestrians in the same social group whereas repulsive forces act between pedestrians and obstacles and other pedestrians preventing collisions. However, unlike the rule-based cellular automata models, collisions are not a theoretical impossibility and unlike velocity-based models collision avoidance is a result of the model forces rather than an active search for collision-free paths. The social forces models also include normal and frictional contact forces.

Velocity-based Models: These models are goal-based in that they specifically choose paths to avoid collisions. These models are often sight-based, such as (Moussaïd et al., 2011), where information about the environment is passed on to agents as the perception of sight. The general methodology used in these models involves pedestrians, at every time step, seeking out a collision-free path that would lead them, in the most direct manner, to their respective destinations. Other models, such as those reviewed here (Geraerts and Overmars, 2004), first determine collision-free paths from origin to destination and then make the pedestrian follow it through a series of closely placed nodes.

Others: Other models include those that do not fit in the above categories but are not common enough to have their own classification. For example, the visual-field based pedestrian model built on space syntax theory (Penn and Turner, 2001), or behavioural models using discrete choices such as (Robin et al., 2009).

3.2.3 Actions Modelled

The different actions that are modelled at the operational level can be broadly classified into (i) walking and (ii) waiting behaviour. These are distinct actions that need to be modelled explicitly (Campanella, 2016; Daamen, 2004) and, depending on the application or need to model a specific situation, can be modelled in less or greater detail. Both walking and waiting can be further broken down into individual actions and collective phenomena. Moreover, another level between these two is the behaviour of social groups where small groups of people formed by social ties, such as kinship, intentionally make collective decisions.

Walking

Walking involves moving from an origin to a destination whilst interacting with the environment. The origin and destination are known from the tactical level but may not be modelled as deterministic; meaning that some models, such as (Hoogendoorn and Bovy, 2004b), allow the operational level to communicate to the tactical level on-the-fly route and destination choices. Routing decisions and the related feedback mechanisms from the operational level are discussed in chapter 4 while this section concentrates on physical actions carried out by pedestrians. Based on the nomenclature used by Daamen (2004), the environment consists of other pedestrians, obstacles, joints, and walkways. Obstacles include walls and furniture; joints are obstacles with which pedestrians interact to move from one space to another, such as turnstiles or doors; and walkways are where the walking takes place, for example, a flat surface, stairs, or escalators. While the method of moving individuals through space depends on the methodological approach the following describe different actions within walking that may need to be modelled. Individual actions include (i) collision avoidance with other pedestrians and obstacles, (ii) physical contact in crowds, and (iii) movement on walkways other than stationary flat surfaces; collective phenomena found in literature are (i) motion cases and (ii) self-organization phenomena; and finally walking behaviour in social groups which has been observed to have different walking behaviour than individuals. Below, each of these actions and the model features and methodological approaches suitable to model them are discussed. It should be noted that the performance of microscopic models in simulating motion base cases and the self-organization phenomena depends on how individual behaviour such as acceleration and collision avoidance is modelled.

Collision Avoidance: Collision avoidance between pedestrians and other pedestrians, and pedestrians and static obstacles is a fundamental behaviour modelled in some form by nearly all models at the microscopic level (Campanella, 2016). Duives et al. (2013) classify the performance of models on collision avoidance as being either intrinsic to the model such that avoidance measures are taken before reaching a conflict or only being present at the local operative level. Similarly, Campanella (2016) classifies such interactions into reactive and anticipation based models. Reactive models depend on local and contemporary interactions to avoid collisions whereas anticipatory or predictive models pre-compute possible collisions on the current path in order to avoid them. Campanella (2016) reports, based on studies by Sparnaaij (2015), that anticipatory models perform better in low to medium density flows whereas reactive models are more useful in high density flows. One reason for this may be that in low densities pedestrians are able to observe other pedestrians and obstacles from afar and perform collision avoiding manoeuvres early on whereas in high density crowds, pedestrians do not anticipate but react to their immediate surroundings. According to Duives et al. (2013), velocity and force-based models are able to model collision avoidance more realistically because of their ability to factor in objects at larger distances enabling pedestrians to adjust their paths several meters upstream of the collision. Yet, pure force-based models are reactive in nature and need to be combined with some anticipation strategy for a more realistic simulation over a range of pedestrian densities (Campanella, 2016). With regards to interaction with obstacles specifically, it is suggested that pedestrians maintain a certain separation (Bosina et al., 2016) or 'shy-away' (Campanella, 2016; Daamen, 2004) distance to walls and other obstacles. Moreover, this distance decreases with increase in pedestrian density (Bosina et al., 2016) and varies with material types and function (Daamen, 2004). As opposed to discrete-space models such as cellular automata, spatially continuous models are better able to incorporate this information (Campanella, 2016).

Physical Contact: In overcrowded situations, physical contact with other pedestrians and obstacles are often inevitable. The forces on pedestrians due to such contact can be decomposed into normal forces that cause body compression and frictional forces that slow down relative tangential motion (Helbing et al., 2000). Such forces are responsible for the formation of arches in front of bottlenecks ('faster-is-slower' effect (Helbing and Johansson, 2011)), and large crowd pressure on individuals in turbulent pedestrian flows (Helbing and Johansson, 2011). Thus, these forces are essential in the assessment of safety of large crowds. Cellular automata models are generally unable to represent these forces (Campanella, 2016) and overcome the inability to reduce velocity arbitrarily by adding probability parameters that decide whether pedestrians in contact move in a given time step (Schadschneider et al., 2009; Wei-Guo et al., 2006). but force-based models and models that specifically add a physical contact component (such as (Moussaïd et al., 2011)) are able to simulate these effects.

Non-stationary & flat Walkways: The use of walkways such as stairs or escalators for level changing is not modelled in most pedestrian simulation models and has been identified as a gap in literature by Campanella (2016). SimPed (Daamen, 2004) and NOMAD (Campanella, 2016) simulate pedestrians on stairs and escalators as being on flat surfaces but with a lower speed as indicated from previous empirical studies. Additionally, NOMAD applies certain other behavioural hypotheses such as less overtaking and no back-stepping for more realistic individual behaviour.

Motion Cases: In order to assess the performance of various models in their representation of crowd motion, Duives et al. (2013) present eight distinct motion base cases, the combination of which can represent the whole range of pedestrian movements as shown in Figure 3. The cases are derived such that unidirectional flows are classified based on the change in space available for walking and direction of walking while multi-directional flows are divided based on the angle of interaction and number of flows interacting. Amongst the models reviewed, they find that only force-based models (Helbing and Molnár, 1995), utility-based NOMAD (Hoogendoorn and Bovy, 2004b), and some cellular automata based models are able to realistically model all eight motion base cases. Even methodological approaches that model individuals but are at a more macroscopic level, such as network-based models, are unable to realistically describe the motion cases as the description of the environment and individual movements is limited in such models (Duives et al., 2013).

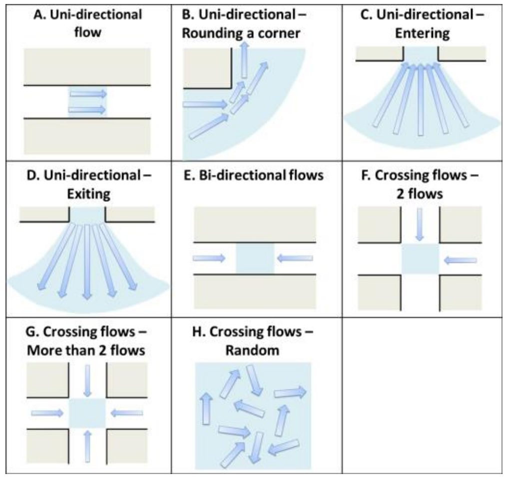

Figure 3: Pedestrian movement base cases (from (Duives et al., 2013))

Self-organization Phenomena: Self-organization phenomenon in pedestrian behaviour is defined by Helbing and Johansson (2011) as the formation of various spatio-temporal patterns of motion resulting from the interactions between many objects. Stop-and-go pedestrian movement waves (Helbing et al., 2007), slower movement of uncoordinated crowds ('faster-is-slower' effect), and zipper-like arrangement of pedestrians in narrow passages (Daamen, 2004) are all self-organized patterns observed in unidirectional flows. In bidirectional flows, these patterns include formation of stripes of pedestrian density waves in angular crossing, lane formation in parallel crossing, and oscillation of flows from either side through a bottleneck (Helbing and Johansson, 2011). Here too, Duives et al. (2013) find that force-based models and NOMAD outperform most other models in their ability to simulate different types of self-organization phenomena. While the ability to simulate self-organization phenomena is often used as a face validation test for microscopic models, the phenomena are more relevant for medium to high density crowds (Helbing and Molnar, 1998) that are normally found in public or semi-public spaces than for the low density interactions normally observed in office buildings.

Social Groups Walking: While most models simulate pedestrians as individuals, according to empirical analyses conducted by Moussaïd et al. (2010) for two public areas, the majority of pedestrians were part of a social group, that is, they walk together intentionally because of certain social ties. Their findings also indicate that the speed and arrangement of such social groups depend on their size, the density of pedestrians in the facility, and the trade-off made between speed and facilitation of communication amongst the group. Furthermore, empirical analyses by Versluis (2010) found that individual pedestrians reacted differently to those walking in groups in terms of yielding and extent of evasion. In order to model such social groups (Helbing and Molnár, 1995) propose using attractive force between pedestrians in their social force model. Moussaïd et al. (2010) make use of this model for their mathematical model of social groups, Seitz et al. (2014) use potential fields in a cellular automata setting, and Duives et al. (2013) make the assumption that force-based models are capable of simulating the group behaviour. Although the above empirical findings indicate the need to model social groups, both Moussaïd et al. (2010) and Campanella (2016) remark that social groups are more likely to be found in leisure areas than in areas visited for other purposes.

Waiting

Waiting for services or entering restricted walkways is a common phenomenon in pedestrian facilities. Most models simulate pedestrians as either inactive or continuously walking towards a destination and do not explicitly model waiting behaviour (Campanella, 2016; Davidich et al., 2013). While unorganized queues emerge in front of bottlenecks as a form of self-organization with the simulation of microscopic walking behaviour, organized queues or queues with peculiar characteristics cannot be produced unless explicitly modelled. Kneidl (2016) differentiates the arch form emerging from self-organization of walking pedestrians in front of bottlenecks from waiting pedestrians in a similar situation by defining that pedestrians in a queue have limited physical contact, which, as described above, is a reason for the arch formation. The various behaviours of waiting pedestrians generally depend on the reason for waiting, cultural aspects, and the environment (Kneidl, 2016; Köster and Zönnchen, 2016). Here individual actions include the movements related to queuing while the collective phenomenon is the overall organization of the queue. Notably, unlike the phenomena emerging in walking, the overall organization of queues is intentional and explicit depending upon the situation. Finally, how social groups wait also has to be modelled over and above individual waiting behaviour. Next, each action is explained along with the model properties suitable for its simulation.

Queue Movements: Waiting pedestrians perform various kinds of movements such as approaching, joining, moving within, and leaving a queue (Okazaki and Matsushita, 1993). Furthermore, waiting pedestrians affect those walking and their interactions have also been modelled (Campanella, 2016; Davidich et al., 2013). To model such movements involving waiting pedestrians the concept of a waiting area has been used in both continuous (Campanella, 2016) and discrete (Davidich et al., 2013) space models. For winding queues, Kneidl (2016) and Köster and Zönnchen (2016) propose using the last waiting pedestrian in the queue as guidance for approaching pedestrians. To use this model, one would need to use a model with a fine spatial resolution such continuous or local discretization (Seitz and Köster, 2012) (as is used in the said studies) models so that there are no artificial restrictions on the angle at which pedestrians join queues. Competition within queues is modelled by Köster and Zönnchen (2016) using a switching strategy that changes potential field distribution to simulate behaviour change from cooperation to aggression or vice versa. While the above models use magnetic and social forces or potential fields, Daamen (2004) uses a less complex, network-based approach by assigning some links for queuing in front of services.

Queue Organization: Different queuing patterns are observed under different conditions in real-life which have to be specified as such when simulating their formation. Okazaki and Matsushita (1993) define three different types of queues based on the reason for waiting: (i) service queues where a person reaches and gets a service and then gets out of the queue, such as a coffee machine; (ii) queues for passing through an entry point, such as a turnstile, and entering another walkway; and (iii) queues where pedestrians want to enter another walkway but first need to wait for those on the other side to leave from the same entry point, as happens in front of elevators and train doors. On the other hand, Kneidl (2016) concentrates on the organization of queues and identifies four patterns: (i) organized queues demarcated by an authority, for example, those found in front of security checkpoints at airports; (ii) organized queuing without demarcations resulting in naturally winding queues; (iii) unordered queues waiting in an area, such as those in front of elevators; and (iv) loosely ordered queues in front of bottlenecks with limited competition. Okazaki and Matsushita (1993) use their magnetic forces model to simulate all the queue types but do not expand on their methodology while Kneidl (2016) describes a different model for each queue pattern she identifies.

Social Groups Waiting: Finally, social group behaviour is also observed for waiting pedestrians although models replicating such behaviour are rare. For service queues, groups may wait in queue together or separately where a part of the group waits in line while the others wait separately. Kneidl (2016) models the first case while models allowing the definition of waiting zones may allow defining the separate areas used by the latter case. On the other hand, pedestrians waiting to pass through a restricted entrance are likely to wait together.

3.2.4 Non-functional Features

In section 2.1, model requirements were divided into functional – what needs to be modelled – and non-functional – how should the model be (Aurum and Wohlin, 2005). This section discusses features related to the latter, that is, those describing the performance of operational level models. Zhou et al. (2010) divide these criteria into those that are important from the design and development perspective and those that are concerned with the execution of the model. The former includes (i) flexibility and (ii) extensibility of the model while the latter consists of (iii) execution efficiency and (iv) scalability (Zhou et al., 2010). Next, these features and the performance of different methodological approaches are discussed.

Flexibility: Flexibility refers to how well models are able to represent pedestrian behaviour under a range of different conditions. Different conditions may refer to changes such as crowd densities or pedestrian facility layout leading to the emergence of varying pedestrian behaviour. In most microscopic models, the definition of pedestrian behaviour is independent of the pedestrian facility and other pedestrians which instead act as boundary conditions. Therefore, these models are able to handle a variety of situations. Some models, however, are designed for a specific situation making them unusable for others (Zhou et al., 2010) or represent the environment in a manner not making them amenable to a variety of situations (Duives et al., 2013).

Extensibility: Extensibility is the ability to add extra features or behavioural insights to the model (Zhou et al., 2010). The extensibility of models depends to a large extent on how it represents pedestrians. Models representing pedestrians as simple physical entities (Helbing and Molnár, 1995) are somewhat limited in extensibility as opposed to agent-based approaches (Bandini et al., 2005) which are able to represent complex pedestrian behaviour as well as population heterogeneity.

Execution Efficiency: Model runtime efficiency is an important consideration when selecting a model. Model complexity and detail usually increase the execution time making the trade-off between model accuracy and running time an important design decision (Zhou et al., 2010). As stated previously, macroscopic models run the fastest but have the least level-of-detail; followed by mesoscopic and microscopic models that are increasingly slower. Within microscopic models, rule-based cellular automata models are significantly faster than force-based models (Duives et al., 2013; Gloor et al., 2004). Furthermore, anticipatory velocity-based models relying on vision fields and behavioural models using discrete choices perform even worse (Campanella, 2016; Duives et al., 2013). Some models such as cellular automata define discrete positions that pedestrians can accommodate and use logical rules to make it theoretically impossible for pedestrians to overlap other pedestrians or obstacles. However, models using force-based approaches to prevent such overlap (Helbing and Molnár, 1995; Moussaïd et al., 2011) in theory still allow overlaps to occur (Lakoba et al., 2005) leading to unexpected pedestrian behaviour. Because of this, solving such models using explicit numerical integration methods such as Euler’s, is very inefficient as it forces modellers to use very small time-steps and thus repeat calculations a larger number of times. In order to solve this problem, implicit routines have been suggested in literature that provide reasonable approximations and are reportedly much faster than explicit methods (Koster et al., 2013; Lakoba et al., 2005). Campanella (2016) employs variable time-steps in order to use the simpler Euler's method but overcome its time inefficiency. Depending on the 'isolation' of pedestrians, a global time-step is decided and further improvements are made by allowing isolated pedestrians to ignore other pedestrians and obstacles for a given 'time-in-isolation'. Duives et al. (2013) evaluate the model initialization time separately and find that models that need to define a spatial context, such as a grid, for the pedestrians perform poorly. For models using potential fields, a different potential field has to be generated for every destination possible and the initialization time increases with an increase in spatial resolution and number of obstacles (Gloor et al., 2004). On the other hand, the same study reports that network-based contextualisation depends only on the number of obstacles and although different origin-destination pairs have different routes, they use the same underlying network.

Scalability: Scalability is important when the model is expected to handle large crowds, such as during large-scale evacuations. It refers to the ability to distribute the computational burden of the model such that a large number of pedestrians can be modelled. Similar to the findings of Zhou et al. (2010), it was found that amongst the studies reviewed none reported on the ability to distribute computational burden although many did report on the efficiency of execution and thereby the ability to simulate large crowds.

3.2.5 Model Selection

Figure 4 describes the various design options available based on the structure developed in the previous sections. These options and the model requirements outlined in section 2.1 are used to arrive to the model components to be used to model the operational level behaviour.

Model Features: Since the position of each pedestrian is required as an output of the model, the model has to represent individual pedestrians. Therefore, as stated previously, the model scale should be microscopic or at least hybrid of micro- and mesoscopic such that pedestrians are modelled as individuals for the major part. With regards to space representation, only discrete cells hold one or less pedestrian per cell and continuous representation are able to represent individual trajectories. Between these two options, continuous space is preferred as it does not have arbitrary restrictions on the angle of movement and can produce a smooth, 'natural looking', trajectories which also useful for more believable demonstrations. Network-based approach is selected to contextualize the space for wayfinding because of its execution efficiency during initialization as opposed to potential field methods. The visual field method used in (Turner and Penn, 2002) has problems with long corridors whereas the network-based approach is very robust in adapting to a range of building plans. Finally, this approach, being commonly used, has the advantage of being tried and tested unlike the use of splines and bezier curves. Finally, since there is no specific requirement for the population to be heterogenous, desired speed is the only heterogeneous factor among individuals' walking properties as other factors are considered to be too detailed for the purpose of this model.

Actions Modelled: From the model requirements it is known that horizontal movement with collision avoidance which is modelled by nearly all microscopic models is a strict requirement. Movement in social groups and vertical movements by stairs or escalators are also required but with a lower priority. Since physical contact is not expected to occur in non-emergency situations in offices, and because the pedestrian density is likely to remain low, the operational level model does not need to specifically reproduce collective phenomena such as the motion base cases or the various self-organization phenomena. Waiting behaviour, including various forms of queue organizations, are also required to be modelled.

Non-functional Features: While the operational level model should be flexible so that it can be used for different building layouts, pedestrian densities are not expected to vary by much in non-emergency situations within office buildings. Here, the execution efficiency is also important as the model is expected to simulate long periods of time in order to generate sufficient data for research applications.

Methodological Approach: Considering the above design decisions, it is possible to eliminate cellular automata approaches due to the decision to use continuous space representation. Further, considering that model execution time is an important factor, the worst performing methodological approaches, anticipatory velocity-based models relying on vision fields and behavioural models using discrete choices, can be eliminated. On the other hand, the social forces model can function in continuous spaces, can simulate all the actions required to be modelled and beyond, and has the advantage of having been widely used and calibrated in a variety of situations.

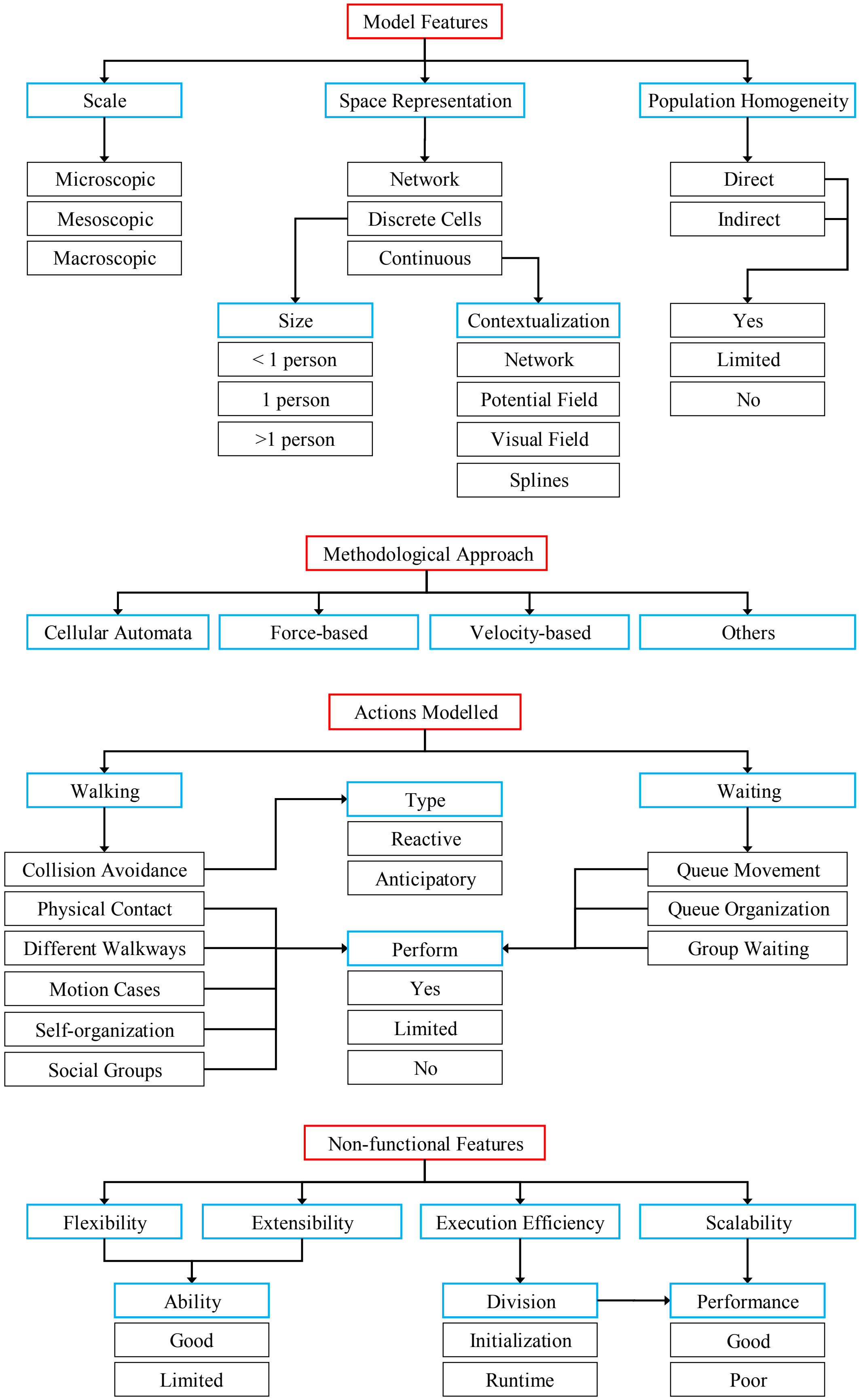

Figure 4: Model design options

3.3 Methodology

This section expands on the model components selected in the literature review. The operational model designed here represents space as a continuum and therefore a contextualisation layer is required for pedestrian agents in the model to navigate around obstacles. Having already decided to use a network based context, here, the navigation graph generation methodology by Kneidl et al. (2012) alongside suitable modifications employed is discussed in section 3.3.1. The social forces model (Helbing and Molnár, 1995) that is used to execute the movements of the pedestrians is described next in section 3.3.2. As noted in the literature review, the social forces model can be used to model different types of queues as well as walking in social groups. However, in the current design waiting behaviour is modelled only as vertical queues and walking in social groups is not implemented. To model vertical movements across floors, a mesoscopic methodology that uses the concept of wormholes is developed (section 3.3.3).

The framework of the operational level model is shown in Figure 5 through a class diagram. Each occupant is associated with a persona that assigns it walking characteristics, such as desired speed and preferred distance to obstacles, and controls its walking behaviour using the social forces model based on its current position and velocity. The building is represented by obstacle edges and nodes as is described in section 3.4. The navigation graph generated from these edges and nodes is used to guide pedestrians from origin to destination, around obstacles, based on routing decisions from the tactical level.

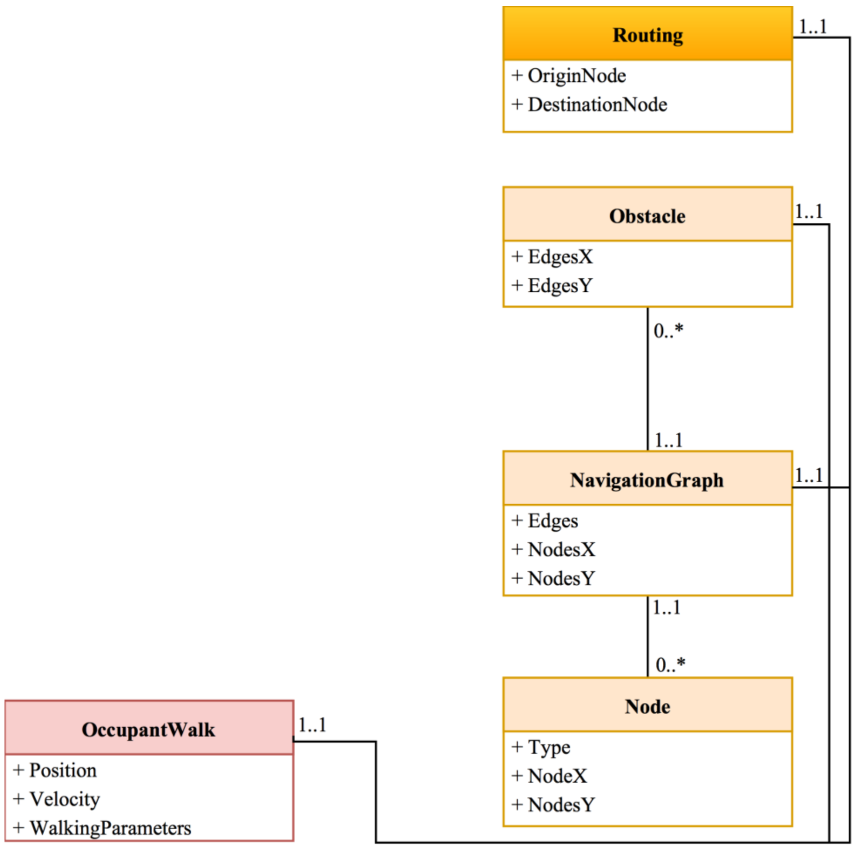

Figure 5: Class diagram of the operational level framework. Colours signify cohesive classes as described in section 5.2.

3.3.1 Navigation Graph Generation

Given a building plan consisting of activity locations and obstacles, the aim of this module is to automatically generate a navigation graph connecting all the locations through a series of mutually visible points called navigation points. Such a graph should be as sparse as possible while still being detailed enough to connect all points and represent distinct routes (Kneidl et al., 2012). Sparseness of the graph is important for the computational efficiency of route choice, especially when such choices are made dynamically, with complex choice models, and for a large number of people. Detailed spatial representation is necessary to allow pedestrians to use all parts of the buildings by connecting activity locations with a number of alternative routes. This module produces a navigation graph in which the nodes are composed of activity locations that are part of the building plan and navigation points defined by this module itself. Two types of edges exist; first, those that connect activity locations and navigation points, and second, those that connect navigation points. This definition of two edge types is necessary to ensure that pedestrians are guided only using navigation points and not locations. Below, the algorithm used to generate the navigation graph is described as a sequence of four steps.

i. Derive Navigation Points from Corner Points

The navigation graph is used to guide pedestrians around obstacles and therefore, the convex corners in the building plan can be used as a reference point. Gloor et al. (2004) propose placing 4-points around each corner and then eliminating those that cannot be reached as these points are inside the obstacle. However, this method would also connect, unnecessarily, concave corners, behave unexpectedly unaligned corners, and create at least three points for each corner less than 90° resulting in a graph with more edges. Instead, Kneidl et al. (2012) proposes placing a single navigation point on the angle bisector of the convex angle created by lines forming a corner. This way the navigation points created by concave corner points lies either inside the obstacle where it is not visible by any other navigation point or such that the navigation point and its corresponding corner are not mutually visible (Figure 6). Such points are removed in later steps. The distance at which the navigation point is placed is based on the body radius of pedestrians.

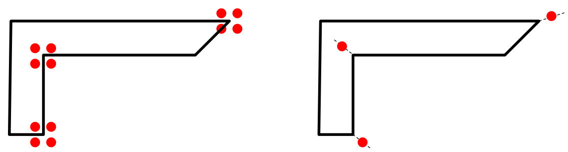

Figure 6: Navigation points (red circles) placement from corner points according to Gloor et al. (2004) (left) and Kneidl et al. (2012) (right) with convex, concave and unaligned corners. Dashed lines indicate convex angle bisectors.

ii. Eliminate Infeasible or Unnecessary Navigation Points

In order to eliminate infeasible points, two rules are considered for: (i) clear line of sight to corresponding corner, (ii) clearance to obstacles (Kneidl et al., 2012). These rules ensure, respectively, that the navigation points are accessible and that there is enough space for pedestrians to pass through them. While Kneidl et al. (2012) uses the distance between two obstacles to implement the second rule, the method is unclear as it does not describe the exact point of measurement. The method used here considers the distance between the navigation points and obstacles along the lines of sight between the former and their corresponding corners. For all navigation points with a clear line of sight to their corresponding corners, the clearance distance to other obstacles are calculated. If a minimum clearance is not available, the navigation point is moved to the minimum clearance distance from its corresponding corner. The minimum clearance distance is based on pedestrian body radius. Next, the first rule is checked by connecting a line between the navigation point and the corresponding corner and seeing if this line intersects any obstacle edges. If it does, the navigation point is placed halfway between the intersection point nearest to the corresponding corner and the corresponding corner. Finally, the minimum clearance rule is used to eliminate infeasible nodes. This is described through an example in Figure 7. The left figure indicates the navigation points derived from the corners. The algorithm sets the bottom navigation point to a point mid-way (black circle) the corresponding corner point and the obstacle along the convex angle bisector to achieve a clear line of sight, while the upper navigation point is moved to the point of minimum clearance from the corresponding corner (black triangle) because it is too near an obstacle. When the final check for minimum clearance is done, it is clear that the bottom navigation point is infeasible and is therefore eliminated, as shown in the right figure.

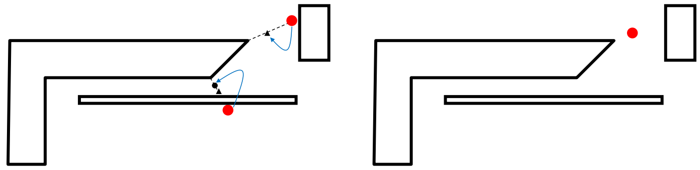

Figure 7: Example of eliminating infeasible navigation points. Black circle and triangle indicate mid-point between obstacle and corner point, and minimum clearance distance from corner point respectively

Even though only one navigation point is derived per convex corner, there may be still be several redundant points which may be eliminated. This is achieved by merging navigation points that are in close proximity to one another into one point lying between them. As noted, correctly, by Kneidl et al. (2012), using a fixed threshold to decide which navigation points are close enough to be merged could result in some points becoming inaccessible from others. Here two rules are used to define which navigation points should merge. The first differs only slightly from the single rule used by Kneidl et al. (2012) as it was observed to produce results consistent with expectations: if a corner point is closer to a navigation point than the navigation point's own corner point, the former corner point is added to the navigation point's list of corner points; if the respective corresponding corners of two navigation points can be found in each other's list then the two points must be merged. The additional second rule used here consists of merging two navigation points if they are closer to one another than their respective corner points. In theory, if very large clearance distances are used this may lead to similar problems of inaccessibility, as stated previously, but since such large values are not expected and would not be realistic, this problem is ignored in favour of having a more intuitive result. Figure 8 explains the working of the two rules. The first rule is applicable to the bottom set of navigation points where each navigation point lies closer to the other's corresponding corner point than their own and the second to the upper set where the points are closer to one another than their corresponding corner points. It should also be noted that the rules would not have been applicable to each other's case thus, indicating the need for two distinct rules.

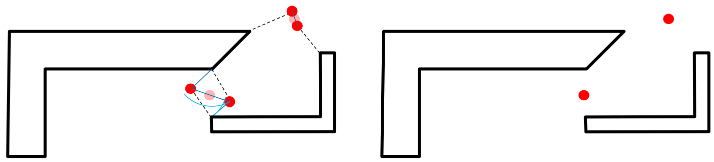

Figure 8: Example of merging navigation points. Translucent red circle indicates merging position. Blue lines and arcs used to assist distance comparison.

iii. Connect Navigation Points

One of the main contribution of Kneidl et al. (2012) on which this model component is based, is the sparse connection of navigation points unlike previous studies, such as (Gloor et al., 2004), which have connected all mutually visible points. This algorithm uses a cone-based search to find which points should be connected from a given navigation point (Kneidl et al., 2012). The method assumes that the angular resolution used during wayfinding is limited and therefore two navigation points lying within such resolution from a given point are equivalent; thus, only the nearer navigation point is connected and a new wayfinding decision is made from there. To create such a network, the algorithm starts with any navigation point as origin and orders the other points by their distance from the origin point. The origin point is connected with the nearest point and all points lying within the chosen half-cone-angle on either side of this connection are removed from consideration. The next-nearest point is connected, the angular elimination is carried out, and so on until no more connections are possible. This is repeated for all navigation points as the origin point to obtain a unidirectional graph of navigation points. An example of the cone-based search is shown in Figure 9.

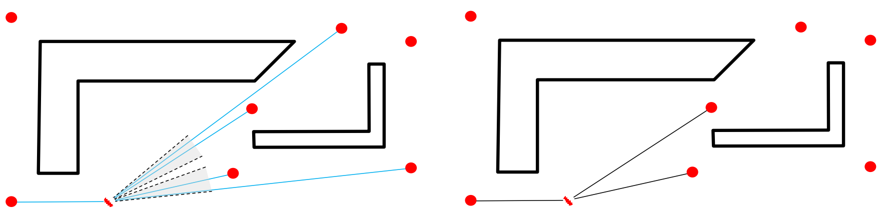

Figure 9: Example of cone-based search to connect navigation points. Striped circle is the origin point, blue and black solid lines indicate prospective and final connections respectively, and black dashed lines represent the search cone. Based on (Kneidl et al., 2012).

The half-cone-angle is responsible for determining the sparsity of the connects; larger angles result in greater sparsity (Kneidl et al., 2012). At the risk of stating the obvious, it should be noted that for unreasonably large angles, the graph's connectivity becomes increasingly dependent on the distance order which could lead to unexpected connections as well as fragmentation of the graph. Kneidl et al. (2012) arbitrarily use 9° as the half-cone-angle value. The same method is also used to connect activity locations to other activity locations and navigation points.

iv. Remove Infeasible Connections

This step is not used in (Kneidl et al., 2012) but is included in the algorithm here because the social forces model (section 3.3.2) is used to determine local pedestrian movements. Connections between navigation points that are just visible, lie very close to obstacle corners. When a pedestrian following such a connection comes close to the obstacle corner, the resulting forces on the pedestrian may make it impossible to reach the next navigation point resulting in unexpected behaviour. Such behaviour is especially observed when the obstacle involved is of relatively low thickness and the pedestrian approaches the navigation point proximity area (see section 3.3.2) on the side from which the next navigation point is not visible. Such an example is shown in Figure 10. Therefore, all connections which do not have a minimum clearance distance from all corners are removed. The minimum clearance distance is based on pedestrian body radius. Although unlikely, since such just visible connections are rare on their own, it should be noted that removing a connection can lead to fragmentation of the navigation graph. This is not remedied in the current implementation although a possible solution is to re-run the cone-based search while disallowing the previously removed connections from becoming prospective connections.

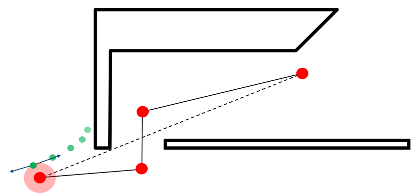

Figure 10: Example describing the need to remove connections close to obstacle corners. Solid green circle is a pedestrian, translucent green circles indicate trajectory, translucent red circle indicates navigation point proximity area, and arrows represent approximate forces on the pedestrian.

3.3.2 Social Forces Model

This module determines the local walking behaviour of pedestrians, that is, their movements towards their respective destinations and various interactions with the environment. The social forces model used here is based on (Helbing and Molnár, 1995) with modifications from (Campanella, 2016) and face validation of test runs. Considering the model is to be used in non-emergency situations some components of the model have been simplified. Since, pedestrians are highly unlikely to come into contact in low density office environment, a major simplification is the absence of normal and frictional contact forces. Consequently, the pedestrians are also simply represented by points rather than using a shape that considers pedestrian body dimensions. Because of this simplification, it is very likely that the model is unable to represent many self-organization phenomena but should, nevertheless, be able to represent all behaviour observed in non-emergency, low density situations that are common in offices. However, if required, the model can be modified to represent body radius and contact forces, the decision to not do so was only to simplify the current implementation. Further, as discussed in section 3.2.3 the use of anticipation strategies is suggested by Campanella (2016). Using no-acceleration assumptions to predict positions of pedestrians in the future, and using these positions to calculate social forces produced smoother trajectories not only in pedestrian interactions but also when manoeuvring through quick turns into bottlenecks. Therefore, anticipated positions are used to calculate all the forces acting on pedestrians. It is important to emphasis here that these choices are made based on prima facie observations and not through scientific calibration. These decisions can be easily changed through minor modifications. Finally, to overcome issues, characteristic of force-based models, related to pedestrians getting stuck when faced with head-on collisions with other pedestrians, an extra lateral force (Campanella, 2016) is added under certain circumstances.

To describe the social forces model used here the various forces acting on pedestrians in order to induce movements are described. The forces are (i) attraction towards the destination, (ii) repulsion from obstacles, (iii) repulsion from other pedestrians, (iv) extra lateral forces, and (v) fluctuation forces which are summed to calculate the total acceleration of each pedestrian:

$$a_i^{total}(t) = a_i^{dest}(t) + a_i^{obstacle}(t) + a_i^{pedestrian}(t) + a_i^{lateral}(t) + a_i^{fluctuation}(t) \qquad [1]$$

Where,

$\boldsymbol{a_i^{total}(t)}$	[m/s^2]	Total force on pedestrian *i* at time *t*
$\boldsymbol{a_i^{dest}(t)}$	[m/s^2]	Attractive force towards next destination
$\boldsymbol{a_i^{obstacle}(t)}$	[m/s^2]	Repulsion force from all obstacle edges
$\boldsymbol{a_i^{pedestrian}(t)}$	[m/s^2]	Repulsion force from all other pedestrians
$\boldsymbol{a_i^{lateral}(t)}$	[m/s^2]	Extra lateral forces due to other pedestrians approaching head-on
$\boldsymbol{a_i^{fluctuation}(t)}$	[m/s^2]	Random disturbance force

As stated above, the model uses anticipated positions of pedestrians in order to calculate various forces on pedestrians. A constant anticipation time is used throughout but the model may be modified so that anticipation time changes with the level of congestion in the walking facility. The anticipated position is given by:

$$\vec{\tilde{p}}_i(t) = \vec{p}_i(t) + \vec{v}_i(t) \cdot \tau^{anticipation} \qquad [2]$$

Where,

$\boldsymbol{\vec{\tilde{p}}_i(t)}$	[-]	Anticipated position (for time $t + \tau^{anticipation}$)
$\boldsymbol{\vec{p}_i(t)}$	[-]	Current position
$\boldsymbol{\vec{v}_i(t)}$	[m/s]	Velocity
$\boldsymbol{\tau^{anticipation}}$	[s]	Anticipation time

i. Attraction Towards Next Destination

Pedestrians seek to maintain a desired velocity that is in the direction of the shortest path to their next destination with a characteristic desired speed (Helbing and Molnár, 1995). This is modelled by an attractive force from the anticipated position of each pedestrian to their destination point. Each movement from an origin to a destination is associated with a route on the navigation graph consisting of a sequence of nodes, starting and ending with the origin and destination activity locations respectively. The destination of a pedestrian at a given time is the next point in the sequence of nodes the pedestrian is following. While the pedestrian is attracted to the destination node, the model assumes that the pedestrian has reached the point as soon as a certain proximity area around the point is entered. This is necessary to prevent artificial merging and diverging behaviour at navigation points (Gloor et al., 2004) and oscillations at destination activity locations as pedestrians try to approach a point under the influence of several forces. The proximity area is modelled as a circle of a given radius around nodes in the navigation graph. Once a pedestrian reaches the proximity area of a destination node, if the destination activity location has not been reached the destination node is updated by selecting the next node in the sequence, otherwise the movement is stopped and the pedestrian is moved directly to the location.

$$a_i^{dest}(t) = \left(\frac{1}{\tau^{relax}}\right)\left(v_i^{desired}\vec{e}_i^{dest}(t) - \vec{v}_i(t-1)\right) \quad [3]$$

$$\vec{e}_i^{dest}(t) = \frac{\vec{p}_i^{n+1}(t-1) - \vec{p}_i(t-1)}{\left\|\vec{p}_i^{n+1}(t-1) - \vec{p}_i(t-1)\right\|} \quad [4]$$

Where,

$\boldsymbol{\tau^{relax}}$	[s]	Relaxation time
$\boldsymbol{v_i^{desired}}$	[m/s]	Desired speed
$\boldsymbol{\vec{e}_i^{dest}(t)}$	[-]	Direction of shortest path to next destination point
$\boldsymbol{\vec{p}_i^{n+1}}$	[-]	Position of node *n+1*, indicating the destination, in the route associated with pedestrian *i* at time *t*

Relaxation time indicates the time in which pedestrians seek to get to their desired velocity such that lower relaxation times result in more aggressive and stubborn pedestrians. Here, the same relaxation time is assigned to all the pedestrians but desired speeds are chosen from a uniform distribution.

ii. Repulsion from Obstacles

Collisions with obstacles are avoided using a repulsive force directed in the direction from the nearest point of the obstacle edge towards the pedestrian. The magnitude of the force is calculated by a monotonically decreasing function of the distance between the pedestrian and the obstacle; here an exponential function is used (Helbing and Molnár, 1995). The total repulsion force from obstacles on a pedestrian is the sum of the forces from each obstacle.

$$a_i^{obstacle}(t) = \sum_{k\in K} -\rho^{obstacle} \cdot e^{-d_{i,k}^{obstacle}(t)/\delta^{obstacle}} \cdot \vec{e}_{i,k}^{obstacle}(t) \quad [5]$$

$$\vec{e}_{i,k}^{obstacle}(t) = \frac{\vec{p}_k^i(t) - \vec{p}_i(t-1)}{\left\|\vec{p}_k^i(t) - \vec{p}_i(t-1)\right\|} \quad [6]$$

Where,

$\boldsymbol{a_i^{obstacle}(t)}$	[m/s^2]	Summation of repulsion forces from all obstacles to anticipated position of pedestrian *i* at time *t*
$\boldsymbol{K}$	[-]	Set of all obstacles
$\boldsymbol{\rho^{obstacle}}$	[m/s^2]	Interaction strength for obstacles
$\boldsymbol{d_{i,k}^{obstacle}(t)}$	[m]	Distance between pedestrian and the nearest point on obstacle edge *k*
$\boldsymbol{\delta^{obstacle}}$	[m]	Interaction distance for obstacles
$\boldsymbol{\vec{e}_{i,k}^{obstacle}(t)}$	[-]	Direction of shortest path to nearest point of obstacle *k* from anticipated position of pedestrian *i* at time *t*
$\boldsymbol{\vec{p}_k^i(t)}$	[-]	Point on obstacle *k* nearest to pedestrian *i*

The interaction strength is the magnitude of acceleration imposed on a pedestrian when the distance between the pedestrian and obstacle is zero while the interaction distance controls the strength of the force on the pedestrian in relation to the distance from the obstacle. A larger interaction distance results in greater force at larger distances but slower increase in force with decrease in distance (Sparnaaij, 2015). The interaction strength and distance are assumed to be the same for all pedestrians.

iii. Repulsion from Other Pedestrians

Similarly, repulsion forces are also used to model pedestrian interactions but instead of using Euclidean distances between pedestrians, elliptical distances are used that take into account the consideration that pedestrians approaching from the front have a greater impact than those approaching from the sides or behind (Campanella, 2016; Helbing and Molnár, 1995). This is described in Figure 11 where it can be seen the maximum influence is from pedestrians in the direction of walking and the least is from those walking behind.

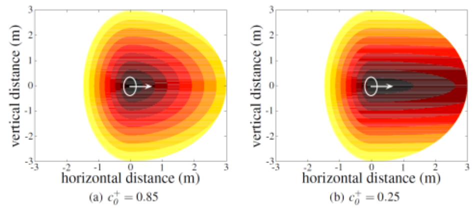

Figure 11: Isofields around a pedestrian, in white, walking left to right, indicating the influence of other pedestrians at different distances and directions. Darker colours indicate larger influence. Two figures used to demonstrate effect of parameter. From (Campanella, 2016).

The elliptical distances are calculated making use of the fact that repulsion forces increase with decreasing distance. Thus, distances in front are reduced by a factor while those at the back are increased.

$$a_i^{pedestrian}(t) = \sum_{p \in P} -\rho^{pedestrian} \cdot e^{-\breve{d}_{i,p}^{pedestrian}(t)/\delta^{pedestrian}} \cdot \vec{e}_{i,p}^{\,pedestrian}(t) \quad [7]$$

$$\vec{e}_{i,p}^{\,pedestrian}(t) = \frac{\vec{p}_p^{\,i}(t) - \vec{p}_i(t-1)}{\left\| \vec{p}_p^{\,i}(t) - \vec{p}_i(t-1) \right\|} \quad [8]$$

$$\breve{d}_{i,p}^{pedestrian}(t) = (z_{i,p}^{front})\sqrt{\left(\beta^{front} \cdot d_{i,p}^{normal}\right)^2 + \left(d_{i,p}^{lateral}\right)^2} + (1 - z_{i,p}^{front})\sqrt{(\beta^{back} \cdot d_{i,p}^{normal})^2 + (d_{i,p}^{lateral})^2} \quad [9]$$

Where,

$\boldsymbol{P}$	[-]	Set of all pedestrians
$\boldsymbol{\breve{d}_{i,p}^{pedestrian}(t)}$	[m]	Elliptical distance between anticipated positions of pedestrians *i* and *p*
$\boldsymbol{z_{i,p}^{front}}$	[1/0]	Binary variable equal to one when pedestrian *p* is in front of pedestrian *i*
$\boldsymbol{\beta^{front}}$	[-]	Multiplying factor (<=1) for anticipated distances to pedestrians in front
$\boldsymbol{\beta^{back}}$	[-]	Multiplying factor (>=1) for anticipated distances to pedestrians in back
$\boldsymbol{d_{i,p}^{normal}}$	[m]	Distance in the direction of walking of pedestrian *i*
$\boldsymbol{d_{i,p}^{lateral}}$	[m]	Distance normal to the direction of walking of pedestrian *i*

iv. Extra Lateral Force

In addition to repulsion forces from other pedestrians, an extra lateral force is induced to prevent unexpected behaviour in head-on pedestrian interactions. The lateral acceleration is fixed and acts on the right-hand side in a direction normal to the line connecting the two pedestrians. The right-hand side is chosen based on the expected cultural preferences (Versluis, 2010). The conditions of interaction under which this force is applied are based on the anticipated (i) distance and (ii) lateral distance of the pedestrians, and (iii) angle between their velocity directions. These three factors ensure that the pedestrians are indeed moving more or less head-on towards one another.

$$a_i^{lateral}(t) = \sum_{h \in H} -\rho^{lateral} \cdot \vec{e}_{i,h}^{\,lateral}(t) \quad [10]$$

$$\vec{e}_{i,h}^{\,lateral}(t) = \langle \vec{e}_{i,p}^{\,pedestrian}(t)_j, -\vec{e}_{i,p}^{\,pedestrian}(t)_i \rangle \quad [11]$$

Where,

$\boldsymbol{H}$	[-]	Set of all pedestrians on a head-on course as determined by the three conditions
$\boldsymbol{\rho}^{\boldsymbol{lateral}}$	[m/s²]	Fixed lateral force

v. Fluctuation Forces

Finally, the last force included introduces random fluctuations or disturbances in pedestrian trajectories to represent the fact that human movements do not exactly follow the model. The fluctuation forces are drawn from a Gaussian distribution.

3.3.3 Wormholes

In order to execute certain movements that either do not need to be modelled according to the model requirements or are too difficult to be modelled with the current movement model, wormholes are used. Since pedestrians changing floors have a low priority in the model requirements and it is known from (Campanella, 2016) that modelling movements on walkways such as stairs and escalators would need additional modifications to the social forces model, this becomes an example of the former. Turnstiles are an example for the second case, where the movement model, the social forces model, is unable to model movements of pedestrians walking through because of the need to pass very close to obstacles (Campanella, 2016). Wormholes allow pedestrians to move from one point to another without continuous force-based movement, that is, agents in the model would enter one wormhole and then appear at another wormhole without physically moving through space. In its simplest implementation, movements through wormhole nodes can be generated by using an adjacency matrix of such points indicating the travel time between them. The travel time can be based on the service time such as in elevators or turnstiles or on the speed of the pedestrian and the congestion conditions such as in stairs. It should be noted that wormholes model pedestrians in a mesoscopic manner, identifying pedestrians uniquely but not defining their exact position, and thus, the operational level model becomes a hybrid model where pedestrians are modelled microscopically for the majority of the time.

The wormhole locations are generated before the generation of the navigation graph and are treated as activity locations for the graph generation. Routing between activity locations, thus, takes place in two steps: first, the direct connections that do not require making a wormhole transition are made, and then those connections that need

The addition of wormholes may make some navigation points unnecessary. This is especially the case for obstacles forming a series of turnstiles as shown in Figure 12 because the navigation points that are used to guide pedestrians around corners are rendered redundant by the wormholes which perform this function for movements across wormhole edges. Such points are removed in step (ii) of the navigation graph generation. Finally, by defining points in front of turnstiles or lifts as wormhole nodes rather than navigation points, it is possible to explicitly model the formation of queues by using these nodes as reference points.

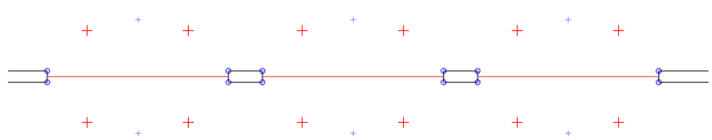

Figure 12: Example showing navigation points generated by a series of turnstiles. Black lines indicate obstacles, red lines turnstiles, red plus marks navigation points, and purple, smaller crosses wormhole nodes.

3.4 Building Representation for the Operational Level

The operational level model mainly requires the physical representation of the building. That is, the building as a composition of edges that act as obstacles and define the walking area, and nodes that act as origins and destinations. As shown by the class diagram in Figure 5, these edges and nodes are defined by their coordinates. Additional information regarding the nodes is required for two cases: first, when pedestrians have to queue at a particular node and second, to identify wormhole nodes so that a wormhole transition can be made. This information is supplied to the operational level model from the 'type' attribute of nodes.

4 Strategic + Tactical Level: Modelling Pedestrian Activities

The strategic and tactical levels of pedestrian behaviour deal with long- and medium-term decisions related to which activities pedestrians want to perform and where as well as when they would like to perform them. Typically, decisions at this level have been left out by pedestrian models assuming them to be given, instead concentrating on operational level and route choice behaviour. In this chapter: first, based on the guiding principle of flexibility, a classification of activities and a structured representation of organizations is presented to model the strategic level behaviour in a wide range of situations; and second, drawing from experiences in several fields of study including building energy performance and passenger transportation, a parsimonious Markov chain based activity scheduler, designed to function under data scarcity yet flexible enough to accept additional data when available, is selected. Following the description of the above, the applicability of the scheduling model to the context of modelling pedestrian movements in offices is studied along with analysis of the methodology to understand its behaviour and limitations.

4.1 Introduction

Strategic and tactical level pedestrian behaviour are upper levels of the hierarchy proposed by Hoogendoorn (2001), providing pedestrians with a reason to move and thereby, a destination and time of movement. The strategic level represents long-term behaviour such as the activities that pedestrians usually perform or events that are planned in advance. The tactical level generates a schedule of the activities pedestrians want to perform, and also decides their location. In addition, the tactical level also makes route choices and, in communication with the operational level, certain en-route decisions regarding location choice. The itineraries prepared and the routing decisions are used by the operational level to simulate pedestrian movements.

Unlike the operational level, few models of the strategic and tactical level pedestrian behaviour are found in literature although route choice behaviour has been explored to some extent. This is a natural consequence of the fact that most studies focus on modelling specific actions (e.g. (Moussaïd et al., 2010)) or situations such as evacuations (e.g. (Gwynne et al., 1999)) where there is no need to generate or schedule activities. Other studies that do model a broader picture under non-emergency situations assume that a movement schedule is already available as input to their model (e.g. (Campanella, 2016; Daamen, 2004)). Therefore, other fields of study such as passenger transportation and building energy performance are also reviewed in order to develop models for these behaviour levels.

This chapter is broadly divided into the strategic and tactical levels. The strategic level focuses on the representation of building occupants, the activities carried out by them, and the office organization within the model. For the tactical level the focus is on activity scheduling while location and route choice decisions are simplified. Since the aim is to model pedestrian behaviour under non-emergency conditions in offices, occupants can be assumed to have a full spatial knowledge obviating the need to explicitly model complex location choice or routing behaviour (Andresen et al., 2016; Hölscher and Brösamle, 2007).

As will be seen in this chapter, design decisions for the strategic and tactical level are often intertwined since the methodology used to schedule movements at the tactical level affects how the strategic level is modelled. Although, these decisions were made in tandem, here, to consistently divide the descriptions of the behaviour levels, the various alternatives available at the strategic and tactical levels are explored separately. Furthermore, the strategic level is discussed first since it discusses a classification of activities that is necessary for the discussion of the activity scheduler. Hence, in the following sections, first the strategic level is discussed (section 4.2), followed by a detailed description of the tactical level activity scheduler (section 4.3), and a short description on the representation of buildings for the tactical level (section 4.4) and the assumptions adopted for location choice and routing behaviour (section 4.5). The relevant parts of the literature review mentioned in the preceding paragraph are split into the strategic level and activity scheduling sections to maintain continuity in the discussions.

4.2 Strategic Level

The strategic level decides which activities occupants would like to carry out (Hoogendoorn, 2001) during their time in the office building. These activities are then scheduled and their locations are chosen at the tactical level, thus driving the movement of the occupants. In order to decide which activities they perform, each individual occupant is associated with a profile of attributes related to the different activities they perform. For this, first a classification of activity types is presented (section 4.2.1), followed by an explanation of how activities are associated with individuals (section 4.2.2). Both these descriptions are placed in the context of literature, especially that related to activity-based modelling in passenger transportation. Lastly, since many activities in offices are related to the organization, the representation of the organization in the model is described (section 4.2.3) to generate such activities.

4.2.1 Activity Classification

Movements are driven by the need to perform various activities which can vary across different situations. For example, the activities carried out in different types of organizations or in different cultures may vary. Therefore, a classification of activities is required to be able to represent all activities affecting movement patterns so that they can be generated and scheduled with a set of common methodologies within the model. In this section, first a review of the classification used in literature is presented followed by a description of the categories used in this study. This description enables model users to appropriately categorise activities performed by occupants in their respective situations.

Most activity-based models for households in transportation research first divide activities into those that occur regularly at a fixed schedule and those that are more flexible. Such activities may be called, respectively, mandatory and discretionary, primary and secondary, or skeleton and flexible (Arentze and Timmermans, 2000; Bowman and Ben-Akiva, 2001; Habib and Miller, 2008). The latter referring to the fact that the regular activities form the skeleton of a person's activity schedule. While some studies only define the primary activities (Bowman and Ben-Akiva, 2001) others also make the secondary activities explicit by categorising them according to type, duration, and frequency (Arentze and Timmermans, 2000; Habib and Miller, 2008). In the former, secondary activities may be further defined according the tours in which they are included. Tours are defined by the activities they include: primary tours start and end at the primary base, usually home, and include a primary activity, such as work or school; secondary tours may start and end at primary or secondary (work or school) bases but may not include any other bases.

Considering transportation hubs and university campuses, Danalet et al. (2013) assume that pedestrian facilities are associated with some kind of schedule whereby they divide activities into compulsory, schedule-guided activities and other activities which are conducted between the former. On the other hand, pedestrians in shopping environments generally do not have activities that can be classified into primary and secondary. They are typically assigned a sequence of points to visit (Ali and Moulin, 2006; Kłeczek and Wąs, 2014). (Borgers and Timmermans, 1986) go a step further and categorise shops by function and assign pedestrians the types of shops they should visit instead of the exact store. Moreover, they also include impulsive stops that are not directly assigned to pedestrians but depend on their destination and route choice behaviour.

For offices, given that their model is used to simulate occupancy patterns, the activities defined by Wang et al. (2011) are simply divided into arrivals, departures, lunch, meetings, and 'walking around'. These categories are used because each category requires a separate step for its generation. While the former four activities are clock time-dependent, the latter is valid throughout the day (between arrival and departure) and includes being in one's own office, visiting other offices, or any other space in the office or outside. Using the activity-based approach, Zimmermann (2007) divides activities into personal and organization related wherein the latter is further divided into (i) continuous activities that represent the base situation or the work done between all other activities, for example working at one's desk; (ii) regular activities that take place at known times for fixed durations, such as meetings; (iii) irregular activities which are induced by the current environment of the pedestrian, such as ad-hoc meetings; and (iv) secondary activities which are dependent on the occupant's current activity, for example, going to the printer while working at one's desk. A similar approach to classification of activities for the prediction of indoor movements in a university is adopted by Kolodziej et al. (2011). This classification is clearly based on the skeleton/non-skeleton division with additional categories to represent specific behaviours observed in offices. Tabak (2008) too presents a similar, albeit more detailed, taxonomy of activities based on (i) the reason for doing the activity, (ii) participation – individual or in a group, and finally (iii) whether the activity is planned or not. Activity reasons are assumed to be social, such as taking a break or chatting; physiological, activities performed periodically to fulfil needs such as drinking water; and job-related activities that depend on the organization. Based on this classification, activities are structured according to the method of generating and scheduling them. First there are organization-related skeleton activities, which can be either planned or unplanned; and second there are 'intermediate' activities, that are social and physiological activities further divided into periodic and fixed probability activities.

Here, combining the approaches by Tabak (2008) and Zimmermann (2007), activities are divided into five categories such that each category contains activities with similar properties whilst all categories together allow various activities, occurring not only in offices but also in other organizations such as schools, to be represented. Moreover, each category is such that all activities within it are generated and scheduled by the same method. In this classification, activities are divided into (i) continuous, (ii) recurrent, (iii) spontaneous, (iv) time-window, and (v) planned. Here, the planned and certain time-window activities correspond to the mandatory or skeleton activity types in that they anchor the solution and other activities are scheduled around them. Below, the properties of each of these activities are described followed by Table 2 where the five activity types are summarised in terms of the attributes that uniquely identify them.

Continuous Activities: They are performed when no other activity is being performed and thus they do not have a fixed starting time, duration, or period. These are the organization-related activities that occupants carry out on their own and are commonly associated with a spatial location such as an employee's desk or laboratory. These locations correspond to the primary base in tour-based activity models from which the series of activities to be performed in a tour starting and ending at the locations are optimized (Bowman and Ben-Akiva, 2001; Danalet et al., 2013). However, as will be discussed later (section 4.3.5), tours are neither explicitly defined nor optimized in the model. Therefore, each occupant does not need to be associated with a 'base' position. Moreover, occupants can be associated with more than one continuous activity which in turn may be associated with multiple locations or areas. For example, in a university, students may spend time between classes performing continuous activities of resting or studying. In turn, studying may take place either in the library or in the café. Finally, continuous activities are assumed to have the lowest priority. This means that they may be interrupted by any other activity and that occupants return to a continuous activity when they are unable to execute other activities.

Recurrent Activities: They are activities related to occupants' physiological processes (Tabak, 2008) and therefore can be expected to have a regular recurrence time. These include taking breaks, getting a drink, visiting the restroom, going for a smoke, etc. and are assumed to be performed alone as they are defined by one's own needs. While these activities are regular, their performance is restricted by time-window and planned activities, such as meetings, where it is assumed that individuals would not prefer to take a break. Tabak (2008) generates such activities in a similar manner to the needs-based approach in activity-based models (Arentze and Timmermans, 2009) where the utility to be achieved by performing the activity increases as time passes and is likely to be performed once it crosses a particular threshold. While Tabak (2008) also includes lunch as an intermediate activity, here, recurrent activities only include those activities that have a recurrence period smaller than the working hours of the individual.

Spontaneous Activities: They include all activities that are usually induced by other activities (which may belong to any of the above activity types) and is identical to the secondary activities proposed by Zimmermann (2007). Although such activities may be recurrent, the main difference is that they occur only in connection with another activity. Examples for such activities could be going to the printer or going to have unplanned interactions with colleagues (at their desk) while working at one's desk. However, as will be noted in the limitations of this model, unplanned interactions are currently not included.

Time-window Activities: As the name suggests, these activities have a fixed time window within which they are performed but unlike planned activities, their starting times and durations are more flexible. This class is useful to model all activities that are time constrained either due to personal habits or preferences, or external circumstances. Activities under this class include arrivals and departures to and from offices; breaks that are more or less planned by personal habits such as coffee or lunch breaks; and activities constrained by opening hours. These activities may also be performed with others, such as having lunch with a group, but the membership would not be as strict as that in team activities. Finally, these activities may be skipped if they cannot be executed in their time window due to planned activities but they have a higher priority than the above activity types.

Planned Activities: They have a fixed starting time and duration which are usually rounded to a common clock-face time such as 10 or 15 minutes. These activities are usually in the form of team activities because the dependency on others' schedules requires them to be planned in advance. However, individual planned activities can also exist, for instance, when one has to switch tasks (for example, to check on the progress of an experiment) at a fixed time or when the activity is still an interaction but through telecommunication with a person outside the organization. Such activities take precedence over all other activities and are usually organization-related as personal activities are less rigid.

Table 2: Summary of activity types in the conceptual model

Activity Type	**Start Time**	**Period**	**Participants**	**Examples**
Continuous	Random	Random	Individual	Working at desk, Roaming
Recurrent	Random	Regular	Individual	Restroom, Coffee, Breaks
Spontaneous	Random	Random/Regular	Individual/Team	Printer, Unplanned interactions
Time-window	Flexible	Random/N/A	Individual/Team	Arrival/Departures, Lunch break
Planned	Fixed	Random/Defined	Individual/Team	Meetings, Classes

4.2.2 Activity-Occupant Association

Occupants in a building perform a subset of all the activities depending on, both, their personal characteristics as well as on the characteristics of the larger organization of which they are part. This section describes how occupants are associated with different activities based on these characteristics.

In activity-based models of travel demand, it is postulated that decisions regarding the choice of activities occur first at the household level and then for the individual (Adler and Ben-Akiva, 1979; Arentze and Timmermans, 2000). This is because members of the household share resources such as income or cars, have shared needs, and often have tasks that need to be completed as a group. The choice of activities at this level depends on long-term characteristics such as household size, employment, and number of children. Further, the allocation of tasks to individuals depends on the roles that members have within the household such as gender-specific roles (Arentze and Timmermans, 2000; Timmermans and Zhang, 2009). When generating activities to be carried out in a day, individuals first consider their own mandatory activities, then the activities that need to be carried out for the household, and then their personal discretionary activities (Arentze and Timmermans, 2009).

Similarly, in organizations such as offices, activities carried out by individuals can be either organization-related or personal (Tabak, 2008; Zimmermann, 2007). Here, in order to generate and schedule activities, occupants are assumed to have two personas: organization related and non-related, which combine to form strategic level profiles for each occupant. The advantage of splitting personas is to segment the inputs required thereby making their collection easier. For instance, personal characteristics are likely to remain similar across different organizations in the same region therefore inputs for personal characteristics do not need to be collected over and over again, and the same information can be used for the non-organization related personas of different organizations. Similar to household activity-based models, allocation of organization related activities to occupants also depends on their role in the organization. Organization related activities include planned and continuous activities while recurrent and time-window activities are personal. Time-window activities such as arrival/departure may also depend on the organization-defined working hours.

Figure 13 demonstrates the relationships between different activity categories and the occupant personas through a class diagram. The strategic level profile of an occupant is given by the attributes of all the activities performed. The attributes to be given as input for a particular category depend on the methods used for its generation and scheduling. Further, spontaneous are not strictly organization- or non-organization related but are still associated with the activities as spontaneous activities are induced by the other activities. Regarding the occupants personas, the attribute 'type' is used to add further information affecting organization-related activities that are not already defined by other organization properties. The non-organization persona contains personal preferences that affect the attributes of recurrent and intermediate activities. The parent class of occupants contains all other personal characteristics such as age or gender that affect all three persona (including walking persona from section 3.3).

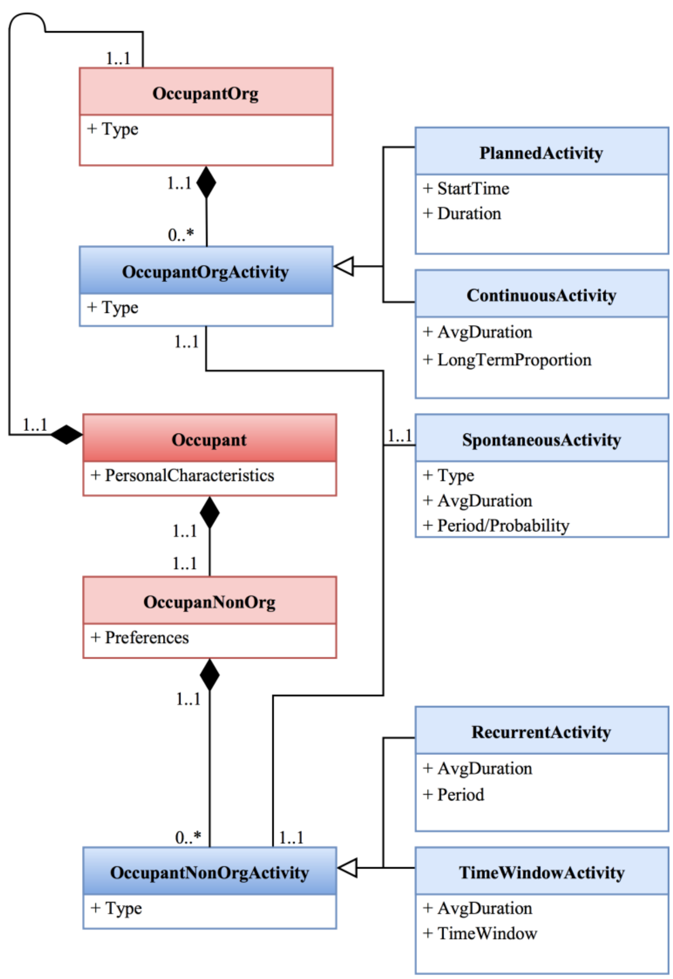

Figure 13: Class diagram of activities and association with individuals. Colours signify cohesive classes as described in section 5.2.

4.2.3 Organization Structure

Personal activities are much more generic than those related to the organization which means that personal characteristics from one observation may be applied to model personal activities in various organizations. However, in order to generate organization-related activities and allocate them to individuals based on their roles it is necessary to describe the organization in a structured manner. Moreover, it is easy to imagine that movement patterns are often guided by the type of organization where they occur; for instance, movements in a school are much more strongly guided by pre-drawn schedules than movements in an office, thus underlining the need for a flexible model that can describe different types of organizations and generate the related activities.

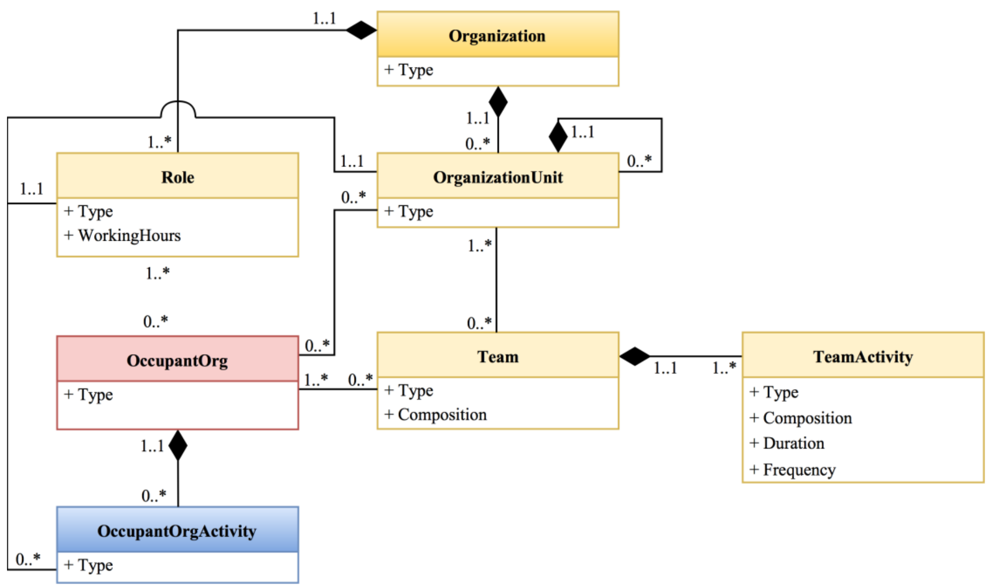

Figure 14: Class diagram of organization representation in the model. Colours signify cohesive classes as described in section 5.2.

Similar to Tabak (2008) and Zimmermann (2007), here, the organization is composed of different roles that occupants can assume and each role in turn is associated with various work-related activities. Figure 14 shows the organization structure used in the model. Roles are defined (attribute ‘type’) by their functions within the organization such as a manager, junior researcher, or visitor. While in the real world each of these roles perform very different activities, here only those activities need to be modelled that cause movements, and therefore various activities can be condensed into more abstract types such as working at one’s desk. Thus, it may be that different roles such as manager and researcher perform the same type of activities, but it is assumed that different roles have different activity properties. For example, managers may spend a larger proportion of time in their office whereas researchers may tend to have several meetings in a day. Further, it is assumed here that different roles also have typical working hours that indicate the average time spent in the organization. Organizations may also be divided according to specialised organization units to further specify work-related activities and to define people likely to work together. For example, two individuals with the role manager may behave differently if they work in different units: sales managers may spend substantial amounts of time outside the office whereas product development managers spend larger amounts of time visiting different laboratories inside the organization. Together, roles and organization units allow representing various types of real-world organization structures as they exist within the model, making the model input simpler and more intuitive. Moreover, together, they also define occupants’ organization related characteristics and thus populate the organization with individuals.

Organizations can also consist of teams that describe which occupants work together and therefore have planned interactions or meetings. Teams can also be used to represent groups of occupants following common schedules such as lecture times in universities. However, when scheduling meetings, Chen et al. (2017) and Goldstein et al. (2010) randomly select participants which is unlikely to be the case in the real world. To overcome this, other studies modelling office activities (Tabak, 2008; Zimmermann, 2007) define organization units to select meeting participants and task groups in order to match activities between individuals. The main problem with this approach is that it leads to unintuitive representations of the organization that the model user has to be aware of during the input procedure. For example, consider the university example used in (Tabak, 2008), the task “Get PhD coaching” of an individual with the role “PhD student” is matched with the task “Give PhD coaching” of an individual with the role “Professor”. However, this does not provide an explicit method to ensure that the same two persons are meeting throughout the simulation. Such consistency is important when the model is used to analyse interactions or if the meetings occur at one of the team members’ offices and thus affect movement patterns. The implicit method of implementing teams in the above models would be to specify them as separate organization units but this would create disparity in the hierarchy of units when the team consists of members from different real-world organization units. That is, if, in the above example, the PhD student belonged to the Mathematics department while one of her supervisors belonged to the Architecture department, the organization unit representing the team would have to be at the level of these departments even though it represents a much more specialised function. Therefore, in the current model, teams are defined explicitly using their composition which refers to the numbers of occupants associated with specific roles and organization units in the team.

The purpose of creating teams is to schedule planned interactions. Each team is composed of various team activities which are defined in terms of participants, meeting duration, and weekly frequency of such meetings. Redefining composition for team activity allows different meeting patterns for different fractions of the team; for example, all members of a particular team may meet only once a month but few members meet much more frequently. The attributes duration and frequency of meeting can be used by the event scheduler to schedule meetings either in a probabilistic or deterministic manner. In organizations where a predefined schedule of team activities, such as lectures in a university, is known this can be directly inserted as an input to the event scheduler.

Representing organizations in a structured manner within the model allows users to pre-define templates for various organization types from empirical observations. With such a template, one would simply need to enter the number of individuals to be modelled in order to derive the rest of the organization characteristics when data is not available for the organization to be modelled. Also, completely modelling the organization structure in parallel to individual spatial movements enables connecting the model to other organization related studies such as interaction dynamics which may be driven by team membership (Sailer and McCulloh, 2012), role in the organization, spatial proximity (Goldstein et al., 2011), spatial layout (Sailer et al., 2016) or personal characteristics, all of which can be included in the model.

Having classified activities, discussed their association with occupants, and represented the organization in the model it is possible to model at the strategic level which activities individuals wish to carry out. The strategic level profiles also enable the generation of activity episodes and their scheduling. As described in chapter 5, planned and time-window activity episodes are generated at the strategic level but for the other categories, episodes are both generated and scheduled at the tactical level. The next section describes the activity scheduler conceptual model from the tactical level.

4.3 Activity Scheduler

This section describes the methodology used to generate activity schedules of pedestrians which will be used by the operational level model as input to execute movements. First, to select an appropriate methodology, previous approaches are reviewed (section 4.3.1) to understand their pros and cons and thereby select a balanced approach that fulfils the model requirements. Next, an overview of the scheduler is presented (section 4.3.2) followed by a description of the core model used (4.3.4), its applicability to offices (section 4.3.5) and analysis of its properties (section 4.3.6).

4.3.1 Literature Review

Approaches in literature for scheduling movements can be broadly divided into (i) activity-based, (ii) location based, and (iii) random access models. In the first category, movements take place due to derived demand from activities that need to be carried out and thus, the models focus on generating realistic activity schedules; in the second, the focus is on reproducing location-based statistics such as occupancy or transition sequences using stochastic models to generate individual movement patterns; and in the last category, movements are modelled as random walks that do not consider locations or activities but may depend on the overall spatial configuration. In the latter two approaches, movements occur for movements' sake rather than being driven by the need to perform activities, although in the second category of approach zones may be linked to some activity types. Next, studies from each of these three approaches are reviewed in order to select a methodology to schedule movements in offices. Rather than explore many different studies of the same kind, here, the idea is to include, under each category, studies from different perspectives. Following this, the advantages and disadvantages of the different approaches are synthesised to assist with the design of the scheduler.

Activity-based

Arentze and Timmermans (2000) develop a comprehensive activity-based model to calculate travel demand where agents in the model learn through previous experiences in the simulation and thus continuously update their behaviour. Given the activities that have to be performed, they are scheduled by considering their priorities and applying various constraints that reduce the number of feasible solutions from which choices are made based on certain rules. Mandatory activities form the skeleton of an agent's agenda and are assumed to be fixed while discretionary activities are chosen thereafter. To schedule the flexible activities the constraints applied include: resource constraints that limit the availability of persons or transport modes such as a household car; institutional opening times; household constraints that dictate timing and chaining of certain activities; limited locations where certain activities can be performed; limited time availability and minimum duration constraints; and finally, constraints due to the space-time prism (Hägerstraand, 1970). Once all activities have been scheduled they are linked by the required travel time. Choices regarding activity location, time-of-day, duration, transportation mode, etc. are all made using heuristic rules for which an extensive and large household survey is required to deduce the required decision tree. The authors were also able to satisfactorily predict origin-destination matrices (also combined with other variables such as transportation mode) for test data thus validating the model.

Tabak (2008) presents a detailed activity-based model of pedestrians in offices that shares many characteristics with activity-based models found in transportation research. It can model various types of activities based on their many attributes such as activity type, frequency, average and minimum duration, priority, time percentage, location, and facilities required. Shared activities where agents are required to interact, such as meetings, can also be modelled. Depending on the activity classification, the activities are generated by different methods and then according to their priorities activities are scheduled using a detailed process that also considers movement time between different locations. For this, the spatial layout of the office is included in the model as a graph with links where movements occur and nodes where different locations such as printers or chairs are present. This network is used to calculate the fixed movement times between locations. The model also includes a resource handler to choose locations, prevent occupancy from exceeding capacity, and find agents for interactions. While this is the most detailed model of pedestrian behaviour in offices available in literature, because of its high level of detail, it requires extensive surveys, especially, to decide the priorities of different activities and to estimate parameters for the utility based choice models used to generate some types of activities.

Kolodziej et al. (2011) develop an activity-based, continuous time Markov model to predict pedestrian movement patterns in order to facilitate handovers in wireless networks. Using activities as the states of a time homogeneous Markov chain, they model the transitions from one activity to another as a jump process. Moreover, since the spatial location of the pedestrians are important for the objective they implement an environment model which consists of location nodes which are associated with different activities and transition nodes which are associated with movement. The environment model is thus modelled as a graph consisting of these two node types which are linked so as to reflect the real-world situation. Thus, an agent would perform a particular activity until a transition or jump occurs, then it is moved through the graph to the destination activity location. To calibrate their model, they use activity diaries where students in a university were asked to fill in where they would be (within the university) at the top of every hour. Although validation attempts returned poor results, the authors attribute this to the disproportionate impact small prediction errors have on the performance indicator used.

To study the effect of changes in retail locations, Borgers and Timmermans (1986) propose a predictive model of pedestrians in a shopping area. The model uses a time inhomogeneous Markov chain that considers, simultaneously, the transition probability between links (locations) and purposes. All pedestrians use the same Markov chain for their destination choices. In order to estimate transition matrix elements, they assume that the probability of choosing a particular street (location) depends on the distance to that street and the shopping floor space on that street, as well as the attractiveness of other streets. Other decisions modelled include route choice between two stops and impulse stops that depend on the route chosen. They calibrate the model using shopper questionnaire-based survey data and validate the model on the basis of destination choice and link flows finding that the model performs better in the prediction of the former. The number of respondents in this model would grow as the number of shop types or the number of streets in the retail location increases as each element in the transition matrix has to be estimated separately. Moreover, all shoppers are assumed to behave homogeneously and thus individual preferences would get lost if the model was used to simulate pedestrian movement patterns.

Location Based

Xia et al. (2009) model a sequence of attraction points on an island that tourists visit using a time homogeneous Markov chain whose elements are estimated from a decision tree of the sequence of points visited which is, in turn, derived from a survey. They find that the Markov chain predictions are generally a good fit to the observations but as the number of stops in the trip increases the goodness-of-fit reduces. The authors suggest that this may be because the memoryless property of first-order Markov chains may not hold as the number of stops increases. Similar to (Borgers and Timmermans, 1986), here too it can be seen that agents are assumed to be homogeneous and to predict results for a larger number of stops more respondents are required. Moreover, neither study takes into account the time spent at the destinations although an updated version of this model, Xia et al. (2011), use a semi-Markov model to explicitly include duration characteristics of each stop. Yet, movement patterns in office buildings would be quite different from the single tour type of movements modelled above.

The remaining studies in this category focus on movements within buildings which are divided into different zones. Page et al. (2008) use a 2-state time inhomogeneous Markov chains to generate stochastic patterns of binary occupancy (presence/absence) in a single-zone, such as a closed office, so that the aggregate statistics supplied as input are reproduced. Using the time inhomogeneity property, the model is able to generate long periods of absence, that is, the time between last departure and first arrival at the office, as well as shorter, intermediate periods of presence and absence indicating movements within the office building. While the model is validated, as admitted by the authors, an extensive dataset of occupancy statistics is required to calibrate the time inhomogeneous model and they propose observing the model accuracy with simpler inputs of lower temporal resolution for future studies. The model can be used for single or multiple occupants in the same zone with the assumption of independent behaviour by simply adding the occupancy patterns of each individual. Liao et al. (2011) attempt to lift the above model's single zone restrictions by developing an agent-based model which returns time-series of agents' locations which can then be summed to derive occupancy patterns. They, too, adopt a model with Markovian properties but instead of using a Markov chain where one would need to fill in all the elements of the transition matrix they develop a method wherein the user has to supply as input, probability vectors, called nominal presence profiles, for each agent, that define the probability of being in a particular zone in the office at a given time step. For scheduled events, such as meetings, that require agents to compulsorily move to another location, they use a separate scheduler module that takes as input the starting and end times of such events. Although the model performs well for single-zone scenarios, it performs poorly for multi-zone, multi-occupant ones. Moreover, defining the nominal presence profiles becomes increasingly cumbersome as the temporal resolution, the number of zones, or the number of agents becomes higher.

Aiming to model multi-zone situations with fewer and less complex data requirements, Wang et al. (2011) propose an approach to estimate time homogeneous Markov chains to model movements of agents in offices. Each agent is assigned a Markov chain, the states of which represent different zones in the office being modelled. The Markov chain is used to generate the activity of 'walking around' different zones in the office between arrival and departure, and all scheduled events of the agents. Although, by using a time homogenous chain, they assume that visiting zones by walking around does not depend on time of day, defining the transition probabilities between all the zones, for all the agents would still be quite tedious and would require large datasets. Therefore, they propose a method that uses a non-linear optimization routine, to derive transition matrix elements from profiles consisting of the proportion of time and the average continuous duration an agent occupies various zones, thus, considerably reducing the number of inputs required to model multi-occupant scenarios. Activities such as arrival, departure, and going to and coming from lunch that do not have a fixed starting time but are not completely random either, are modelled by two-state Markov chains where one of the states acts as an absorbing state generate stochastic transitions for these movements. Similar to (Liao et al., 2011), they propose generating scheduled events separately although they do not define how these events would be generated. Finally, the model is a repeated day model, so that modelling periods of absence between departure and arrival to the office is considered to be out of scope. Although, the authors do not validate the model quantitatively, they find that it is able to reproduce expected daily occupancy patterns of a sample office. Given the simplicity of the model proposed by Wang et al. (2011) it has been used by other authors to develop occupancy simulation software and web application (Chen et al., 2017; Feng et al., 2015; Luo et al., 2017). However, one disadvantages of this model that has been identified is that as the number of zones in the office increases, the time required to generate the transition matrices increases (Chen et al., 2017; Feng et al., 2015). To overcome this, Chen et al. (2017) define categories of zones such as own office, others' office, meeting rooms, auxiliary areas such as restrooms and corridors, and zones outside the building and use these as states, instead of individual zones, in the Markov chain.

Random Access

Ahn et al. (2017) hypothesize that stochastic models such as Markov chains lack an overall randomness because of their limiting behaviour to model occupancy for certain organization types which, they argue, have a more random behaviour. Analysing two university-related spaces where they don't expect to find occupancy patterns over time, they prove that a random walk model performs better than a Markovian one. However, they expect Markov chains to be suitable for at least schools and offices where they find behaviour to be more process-driven. Similar to this, Nassar and Elnahas (2007) propose an accessibility measure that depends on spatial layout rather than activity patterns to generate movement patterns. This work is similar to that proposed in (Penn and Turner, 2001) where pedestrian movements are assumed to depend solely on spatial configuration rather than attractors or destination choices. This and other space syntax studies (Gil et al., 2009; Sailer, 2007) have also concluded that spatial configuration is usually a poor predictor of movement patterns in highly programmed (Hillier, 2007) buildings, such as offices, where people walk with a specific purpose rather than in an exploratory fashion.

Summary

Table 3 summarises the general advantages and disadvantages observed in the three approaches to movement scheduling. Activity-based approaches have the advantage of reflecting the real-world purposes of movement rather than being abstract stochastic processes only. However, they are usually complex and require detailed stated preference data whereas the location based approaches are more straightforward and can be made so that the data requirements are considerably lower. Furthermore, the Markov chain models used in this approach are flexible enough so that they can also accept more information when it is available. The random access models are not applicable for offices and hence are not useful. Thus, the activity scheduling approach proposed here seeks to combine the advantages of activity and location based approaches. To this end, an activity-based framework is considered but with the less complex model proposed by (Wang et al., 2011)[1]. While the simple input needs of the Wang model fulfil the data parsimony model requirement, the activity-based framework makes the model more flexible and intuitive.

4.3.2 Scheduler Framework

In this section, the overall framework of the activity scheduler is discussed. Although the Wang model has several advantages, some issues in the model prevent it from being used directly and here it is sought to list and overcome these issues as well as add necessary functionalities to complement it. Four main issues with the Wang model have been identified:

1. *Zone based inputs*: Having zones in the building as states in the model seriously impedes its performance as either the size of the building increases. Moreover, in the current model, it is necessary to know the high resolution location of a pedestrian, for example, being at a chair or in front of a coffee machine, and not only presence in a particular area. This would cause the number of 'zones' to increase substantially. Thus, here, in line with the idea of having an activity-based framework, the states of the Markov chain are considered to be activities. Since only those activities that require movement need to be modelled separately (Zimmermann, 2007) the number of activities are limited and furthermore using activities would also make the inputs much more intuitive thus lowering the barrier to model use. However, it should be noted that this means that locations have to be associated with the activities that can be performed there as is described in section 5.3.1.

2. *Movements*: Like other location based models, the Wang model too has difficulty modelling movements from one location to another. Three reasons cause this; the first is the lack of spatial information within the model, second, the model tries to stochastically assign pedestrians time in areas such as corridors whereas in reality, usually, time is spent in these areas only when moving to a location, and third, the model assumes the move from one location to another can be completed in one time step and therefore is forced to use time steps that are sufficiently large (e.g. in their paper, Wang et al. (2011) use a time step of 5 minutes). The first problem is already solved as spatial information can be obtained from the navigation graph generated at the operational level. The second one is also solved by not using zones as states and not including 'moving' as a separate activity. For the third, instead of assuming a fixed time for movements, similar to the activity-based approaches, a separate movement scheduler is developed that extracts (static) expected movement time from the navigation graph and uses this to link consecutively scheduled activities.

3. *Running time*: As noted above, the running time of the Wang model has been found to considerably increase with the number of states. In order to mitigate this issue, the method is reformulated (section 4.3.4) for faster execution.

4. *Mathematical properties*: Neither (Wang et al., 2011) nor the authors (Chen et al., 2017; Feng et al., 2015; Luo et al., 2017) who subsequently use the Wang model consider certain mathematical properties of the methodology that (i) limit the space of valid inputs and (ii) lead to infinitely many solutions. To understand these properties better they are analysed in section 4.3.6.

[1] Called Wang model henceforth

Table 3: Pros and cons of different movement scheduling approaches

Approach	Purpose	Authors	Pros	Cons
Activity-based	Travel demand modelling	(Arentze and Timmermans, 2000)	• Comprehensive reflection of real-world purposes and thus ability to cover range of situations • Links to spatial environment explicitly modelled in terms of movement times • Quantitatively validated	• Detailed and large data requirements to estimate parameters for different choice behaviour models • Dependence on reported data rather than revealed
	Office building use simulation	(Tabak, 2008)		
	Movement predictions for wireless networks	(Kolodziej et al., 2011)		
	Pedestrian flow in retail locations	(Borgers and Timmermans, 1986)		
Location Based	Tourist attractions	(Xia et al., 2009)	• Modelled by relatively straightforward stochastic processes with combined movement generation and scheduling • Use revealed preferences from sensors usually for model calibration • Quantitatively validated	• Data requirements can be large to estimate probability profiles of occupants • Do not reflect directly that movements are derived from activities hence inputs can be unintuitive • Spatial environment not explicitly modelled
	Building occupancy simulation	(Page et al., 2008); (Liao et al., 2011)		
		(Wang et al., 2011); (Chen et al., 2017)	• Low and simple data requirements • Qualitative verification of results; quantitative validation of theoretical assumptions	• High running time for large office buildings or loss of individual zone identities
Random Access		(Ahn et al., 2017)	• Simple approach that performs better in exploratory situations and in low programmed or less process driven buildings • Consider the spatial environment of the walking facility • Quantitatively validated for some situations	• Method expected to be unsuitable for process driven buildings such as offices • Low correlation between movement patterns and spatial configuration in offices proven
	Movement patterns	(Nassar and Elnahas, 2007); (Penn and Turner, 2001)		

In addition, the following functionalities need to be added to fulfil the current model's functional requirements:

1. *Event scheduler*: The independence of time homogeneous Markov chains, that are used in the Wang model, from clock-time makes it impossible to simulate activities or transitions which are dependent on time. This problem is overcome through appropriate classification of activities which allows placing time dependent activities under categories that are scheduled by a separate event scheduler.

2. *Resource handler*: The need for resource handlers has been identified in all approach types either in the form of constraints in scheduling activities or preventing over-occupancy in zones. Resource handlers in the current model are required for (i) long- to medium-term and (ii) short-term. The former includes reserving spaces or locations for planned meetings as well as, if required, claiming certain positions as one's own base location for the long-term or for the medium-term in flexible workspaces. In the short-term, the resource handler is required to prevent over-occupancy of spaces in the building and to allow agents to dynamically revise decisions regarding location choice due to queues. The resource handler is also used to ensure that agents do not plan to be in two places at once. Although not implemented here, the resource handler can also be used to check agent availability in the short-term for unplanned interactions.

3. *Re-scheduler*: When schedules are executed at the operational level, they are not followed exactly when the operational model is not deterministic such as in network-based models where the speed on links can be fixed depending on the number of people using it. As a microscopic operational level model is used here the time from an origin to destination can change due to random fluctuations in acceleration or higher pedestrian traffic on the way. Furthermore, activity locations or even activities themselves can change if faster service is possible elsewhere, resources are not available at one location, or resources are simply not available at all to perform a particular activity. All of this requires a feedback mechanism from the operational to the tactical behaviour level to adjust the schedule according to its latest execution.

Based on the above points, Figure 15 shows the class diagram of the conceptual model used for activity scheduling. Four schedulers, three static (event, Markov chain, movement) and one dynamic (re-scheduler) based on feedback from the execution of the static schedule at the operational level interact with the resource handler to make up the activity scheduler. The static schedulers produce an initial agenda or schedule with start and end times of various activities with movements between the end and start times. The event scheduler, schedules planned and time-window activities which are dependent on clock-time and the Markov chain scheduler is the Wang model but with activities as states and is used to model continuous, recurrent, and spontaneous activities. The re-scheduler dynamically adjusts the planned schedule based on feedback from the operational level. Finally, the resource handler checks availability of locations and persons for certain activity start and end times. This working of the schedulers is given in more detail in section 5.6.

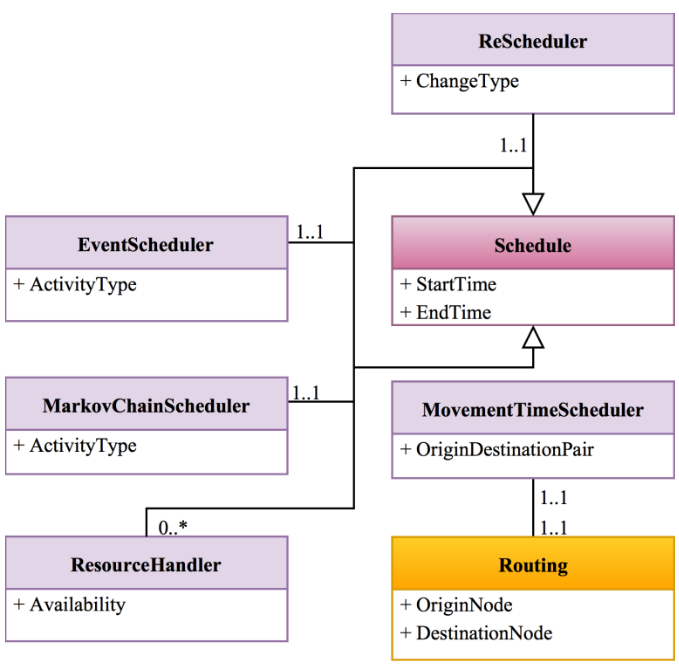

Figure 15: Class diagram of schedulers used in the model. Colours signify cohesive classes as described in section 5.2.

4.3.3 Markov Chain Theory[2]

This section gives a brief introduction to Markov chains for a better understanding of the Markov chain scheduler based on the same theory. The type of Markov chains that will be used here and its relevant properties are discussed. A Markov chain is a memoryless stochastic process defined over a countable set of states in either discrete or continuous time. Here, only the discrete time, finite state Markov chain is discussed since that is the basis of the scheduler. Such a Markov chain over a set of states S is defined by the following:

$$P\{X_t = j | X_0 = m, X_1 = k, \dots, X_{t-1} = i\} = P\{X_t = j | X_{t-1} = i\} \quad [12]$$

$$P\{X_t = j | X_{t-1} = i\} = p_{ij} \quad [13]$$

$$\forall i, j, \dots, m, n \in S$$

The first equation describes the memoryless characteristic of Markov chains indicating the that the state of the chain, *X*, at time *n* is only dependent on the state in the previous time-step. Note that it is possible to consider states further in history by changing the definition of a state to an *m*-tuple wherein the previous *m* states are stored. A Markov chain thus produced is said to be of order *m*; here only first-order Markov chains are considered. The second equation emphasises that the Markov chain does not vary with time and is therefore time-homogenous. Combining p_{ij}'s for all *i* and *j* into matrix form creates a transition matrix where each element indicates the probability of going from the row state to the column state. Below is an example of an *n*-state transition matrix.

$$P = \begin{bmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{bmatrix}$$

$$s.t. \sum_{j=1}^{n} p_{ij} = 1; \; \forall i, j \; p_{ij} \geq 0 \quad [14]$$

With this setup, given a matrix of transition probabilities and an initial probability distribution vector for X_0 the probability of any event in a Markov chain can be calculated as the product of the initial probability and all the subsequent transition probabilities.

A class of states in a Markov chain consists of all states that can communicate with one another, that is, it is possible to start from any state and reach any other state in a class. Two types of classes are possible, recurrent and transient. States in the recurrent class can be reached from all states that they themselves can access while this is not true for transient states. Thus, once a Markov chain goes out of a transient class it is unable to return to it and conversely upon entering a recurrent class it cannot exit it. This can be seen in Figure 16 where states 2 & 3 and state 5 form two recurrent classes respectively while state 1, 4 and 6 each form a separate transient class. In this example, the effect of the starting state becomes very important as it could prevent the Markov chain from ever visiting some states; for example, starting in state 6 means that states 1, 4, and 5 are never visited.

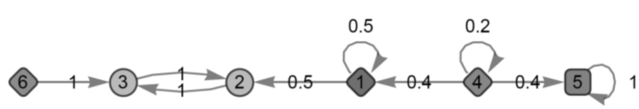

Figure 16: An example Markov chain

Another property affecting the impact of the starting state is periodicity. The periodicity of a state is calculated as the greatest common denominator of the number of steps within which the probability of returning to a particular state is greater than zero, that is, all *t* for which $p_{ii}^t > 0$. A state is aperiodic if it has a periodicity of 1 and periodic otherwise and all states in a class have the same periodicity. In the above example, it can be seen that states 2 and 3 have a periodicity of 2; if $X_0 = 3$ then it is known that $X_2, X_4, \dots, X_{2n} = 3$.

[2] This section is based on Gallager, R.G., 2011. Finite-State Markov Chains, *Discrete Stochastic Processes*. MIT OpenCourseWare. and Serfozo, R., 2009. *Basics of applied stochastic processes*. Springer Science & Business Media.

Based on the above classification of states, it is possible to define an important class of states and subsequently of Markov chains that will be used here. An ergodic class of states is both recurrent and periodic, and an irreducible Markov chain, that is, consisting of only one class, with an ergodic class is called an ergodic chain. Ergodic Markov chains have several interesting properties, especially for large values of *t* (time-steps), that are described next. While some properties may hold for other Markov chains as well, the discussion continues in the context of first-order, discrete time, finite-space ergodic Markov chains and hitherto 'Markov chain' refers to a *n*-state chain with these properties. The Markov chain scheduler will make use of such chains. Formal proofs for these properties may be found in the literature cited at the beginning of this section.

Stationary Probability Distribution: For large values of *t*, that is, $t \to \infty$, the probability of being in a state converges to its limiting value thus becoming independent of the initial state as well as the time-step. This probability distribution describes, in essence, the overall percent of time a Markov chain spends in a given state. A (typically row) vector representing such a probability distribution is called the stationary probability distribution of the chain:

$$\pi = [\pi_1, \pi_2, \dots, \pi_n]$$
$$s.t. \sum_{i=1}^{n} \pi_i = 1;\ \forall i\ \pi_i \geq 0 \tag{15}$$

By definition, the stationary probability distribution can be calculated by the following equation:

$$\pi = \pi \cdot P \tag{16}$$

Thus, the stationary probability distribution is the left eigenvector (normalized to 1) of the transition matrix. When the transition matrix is raised to a large number *k* each row converges to the stationary distribution at an exponential rate to the order of the ratio of the second and first-greatest eigenvalues of the same matrix. Further, for the type of Markov chain under discussion, each transition matrix has only one and only one stationary probability distribution. It is important to note that the reverse is not true (section 4.3.6).

Sojourn Time: This describes the time spent continuously in a state and it has a limiting mean value. The probability of a sojourn time value for a state is given below followed by its expected value:

$$P(\tau_i = t) = (p_{ii})^{t-1} \cdot (1 - p_{ii}) \tag{17}$$

$$E(\tau_i) = \sum_{t=1}^{\infty} t \times P(\tau_i = t) = \sum_{t=1}^{\infty} t \times p_{ii}^{t-1} \times (1 - p_{ii}) = \frac{1}{1 - p_{ii}} \tag{18}$$

From the above expressions, it can be seen that the sojourn times are distributed geometrically and the diagonal elements of the transition matrix can be derived from the expected values of sojourn times.

Away Time: This describes the time spent continuously outside a state and it too has a limiting value of interest. The mean recurrence time for a state is equal to the inverse of its steady state probability. However, this mean value also includes all the 'self-transitions' and is not the same as the time spent continuously outside a state. Instead the value of away time is found (by the author), through observations, to be equal to:

$$E(\alpha_i) = E(\tau_i) \cdot \left[\frac{1}{\pi_i} - 1\right] \tag{19}$$

4.3.4 Markov Chain Scheduler

In this section, the Markov chain scheduler, which is a modification of the Wang model, is presented. First, from the strategic level all the continuous, recurrent, and spontaneous activities are listed and considered as states of the Markov chain. Then, the expected sojourn time, that is the time spent continuously in an activity, and the long-term proportion of time spent in each activity (stationary probability distribution) is used as input. One can either specify these inputs for each occupant individually or create classes of common input profiles. For recurrent activities, such as going to get a drink, it is more convenient to define the away time than the long-term proportion which can be derived from the former according to equation 19 as:

$$\pi_i = \frac{E(\tau_i)}{E(\tau_i) + E(\alpha_i)} \tag{20}$$

For each occupant (or common profile) a transition matrix is derived by setting up a constrained linear least-squares problem. Unlike the non-linear problem setup by Wang et al. (2011), linearization increases the speed of the model considerably thus overcoming the problem of run-time. The system of linear equations is derived from: (i) the relation between the stationary probability distribution and the transition matrix (equation 16), (ii) the fact that the sum of columns of a row in the transition matrix sum to 1 (equation 14), (iii) the known values of the diagonal elements of the transition matrix which are derived from equation 18 as shown below in equation 21, and (iv) the constraint that all transition elements are non-negative values (equation 14).

$$p_{ii} = 1 - \frac{1}{E(\tau_i)} \tag{21}$$

Thus, for an *n*-state Markov chain, the system of equations consists of $2n$ equations, formed by the first two conditions, which are solved for *n*(*n*-1) variables that represent the elements of the transition matrix given that the diagonal elements are known from the third condition. For the 2-state case, the transition matrix is already known based on the 2 conditions. Although, the number of equations equals the number of variables for the 3-state case, a unique solution is not available since the coefficient matrix turns out to be singular and therefore has to be solved in the least-squares sense for a non-unique solution if one exists. Finally, for cases with greater than 3 states, it can be easily seen that the system of equations is under-determined and there is either no solution or an infinite number of solutions. Equation 22 shows the system of linear equations used for the constrained linear least-squares problem and equation 23 shows the setup of the problem. In the general case, constraints only represent the non-negativity of the variables. The derivation of the system of linear equations is available in appendix A.

$$A = \begin{bmatrix}
0 & \cdots & \cdots & \cdots & 0 & \pi_2 & 0 & \cdots & \cdots & 0 & \pi_3 & 0 & \cdots & \cdots & 0 & \cdots & \pi_n & 0 & \cdots & \cdots & 0 \\
\pi_1 & 0 & \cdots & \cdots & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 & \pi_3 & \cdots & \cdots & 0 & \cdots & 0 & \pi_n & \cdots & \cdots & 0 \\
0 & \pi_1 & 0 & \cdots & 0 & 0 & \pi_2 & 0 & \cdots & 0 & 0 & \cdots & \cdots & \cdots & 0 & \cdots & 0 & 0 & \pi_n & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & \cdots & \cdots & \cdots & \pi_1 & 0 & \cdots & \cdots & 0 & \pi_2 & 0 & \cdots & \cdots & 0 & \pi_3 & \cdots & 0 & \cdots & \cdots & \cdots & 0 \\
 & & & & & & & & & & & & & & & & & & & & \\
1 & 1 & 1 & 1 & 1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & \cdots & \cdots & \cdots & 0 & 1 & 1 & 1 & 1 & 1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 1 & 1 & 1 & 1 & 1 & \cdots & 0 & \cdots & \cdots & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 1 & 1 & 1 & 1 & 1
\end{bmatrix}_{2n \times n(n-1)},$$

$$b = \begin{bmatrix} (1-p_{11})\pi_1 \\ (1-p_{22})\pi_2 \\ \vdots \\ (1-p_{nn})\pi_n \\ \\ (1-p_{11}) \\ (1-p_{22}) \\ \vdots \\ (1-p_{nn}) \end{bmatrix}_{2n \times 1}, x = \begin{bmatrix} p_{12} \\ p_{13} \\ \vdots \\ p_{1n} \\ p_{21} \\ p_{23} \\ \vdots \\ p_{2n} \\ \vdots \\ p_{n1} \\ p_{n2} \\ \vdots \\ p_{nn-1} \end{bmatrix}_{n(n-1) \times 1} \tag{22}$$

$$min\|Ax - b\|_2 \; s.t. \, x \geq 0 \tag{23}$$

Since this method requires only two simple inputs it is quite parsimonious, thus fulfilling an important model requirement, but it is important to note that it does not produce a unique transition matrix. Therefore, the range of solutions possible and their impact on the simulation are analysed in section 4.3.6. Apart from being parsimonious the model is also flexible in that it can accept progressively increasing amounts of information to increase its accuracy through additional linear constraints of the form $Cx \leq d$ or upper and lower bounds for *x* in the constrained least-squares problem. An example for this can be found in section 5.6.3.

In order to generate a transition matrix, MATLAB's constrained linear least-squares solver *lsqlin* is used. As discussed above, it is also possible that there are no solutions for a particular input. Since the solver will return variable values regardless, the objective function value, which should be close to zero, is compared with a tolerance value to check whether the values obtained form a valid solution. Here, the tolerance value is chosen to be 10^{-10} to include errors caused by the fact that numerical methods are being used to solve an algebraic equation and that MATLAB number representation causes rounding errors. Once an appropriate transition matrix is available, a Monte Carlo simulation of the Markov chain is carried out by starting from a random initial state and then choosing a new state at every time step using the transition probabilities from the current state. This simulation returns a sequence of (continuous, recurrent, and spontaneous) activities for the occupant. Thus, doing both, generating activity episodes with durations and scheduling them. Figure 17 describes the process of first generating a transition matrix and then simulating it to generate a sequence of activities associated with the Markov chain scheduler.

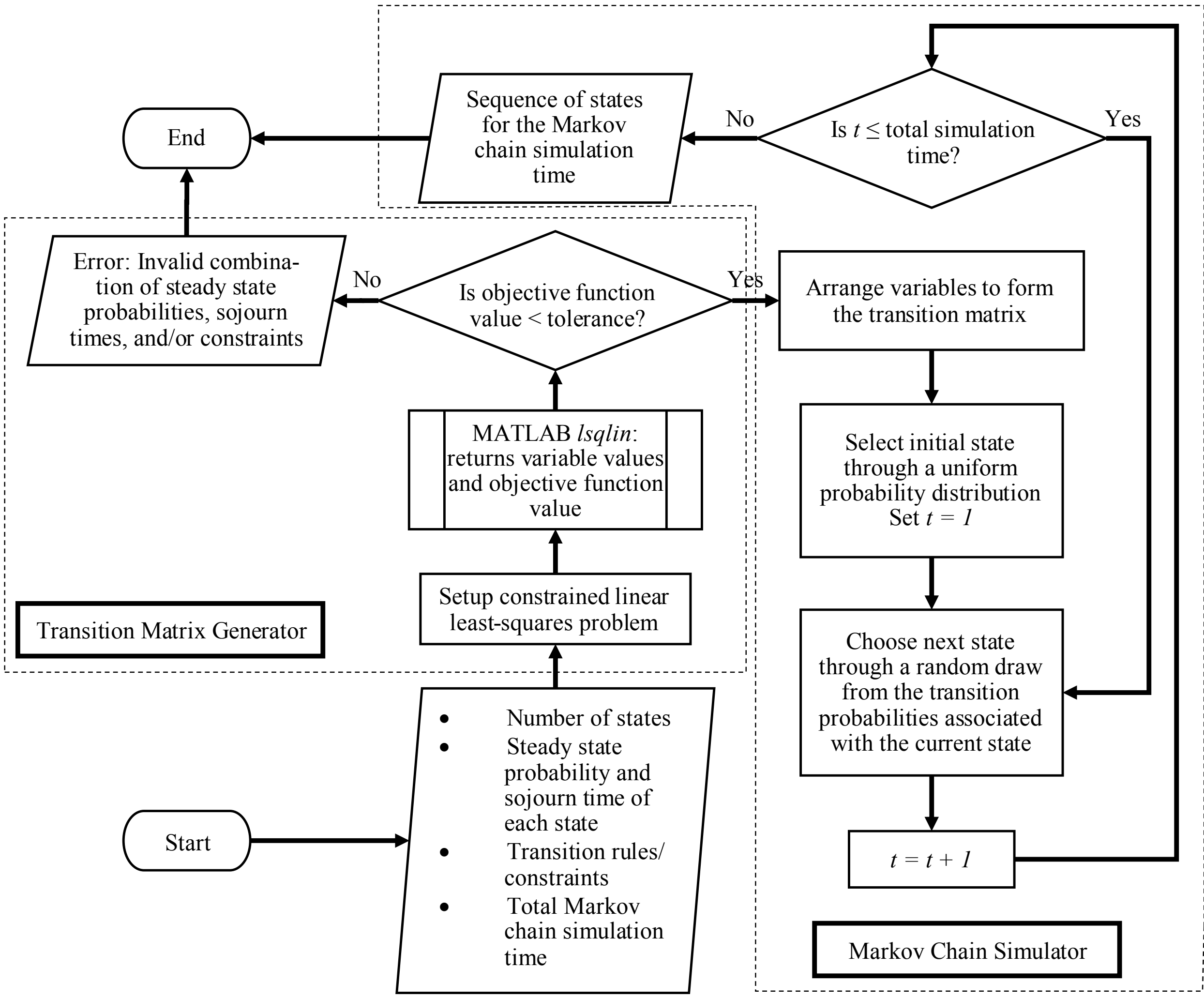

Figure 17: Workflow of Markov chain generation and simulation

4.3.5 Applicability

Having introduced the Markov chain scheduler, in this section the applicability of Markov chains is examined for the following general properties (Wang et al., 2011): (i) memoryless-ness and (ii) expected limiting behaviour in the context of simulating the previously described classification of activities in offices.

Memoryless-ness

While Markov chains have regularity in their limiting behaviours, their local behaviour is stochastic due to the characteristic of being memoryless. Studies on travel behaviour (Kitamura, 1983; Thill and Thomas, 1987) have found that this property is problematic when there is a need to simulate activity chains consisting of more than two activities. Based on analysis of a transportation and land-use dataset, Kitamura (1983) argues that a hierarchy of activities exists which a first-order Markov chain is unable to represent adequately. Although, this problem can be overcome by using higher-order Markov chains, both model complexity and data requirements rise considerably with this solution (Thill and Thomas, 1987). However, movements in offices, being confined to a much smaller, and therefore, easily accessible space, may be assumed to not have significant patterns of complex trip chains. That is, occupants in offices would not, consistently, plan their activity chains in advance because the effort required to access different locations is much lower than it is at the urban or regional scale. Moreover, given the type of activities carried out and the time for which occupants are in offices there is little benefit in optimizing trip chains consistently thus unlike (Hoogendoorn and Bovy, 2004a; Kitazawa and Batty, 2004) no optimization of activity sequencing is performed. For example, office occupants would not, consistently, plan trips to the printer so that they can get coffee on the way back. Furthermore, unlike shoppers in a supermarket or passengers at an airport, office occupants are not present in the office for a limited time within which they need to complete a set of activities in (usually) a single trip chain. Yet, when such patterns do occur, either due to intentional pre-planning or by habit, the model allows classifying such activities into appropriate categories which, as described below, are modelled by a separate event scheduler that is not Markov chain based.

Expected Limiting Behaviour

One of the advantages of using Markov chain based models is their predictable limiting behaviour, that is, properties for large values of t (time-steps). However, before using such a model to simulate real-world behaviour it is important to check whether empirical observations match with the expected distributions of Markov chains, specifically, the geometric distributions of sojourn times for continuous activities and away times for recurrent activities.

Using data from the offices of a university consisting of 25 researchers, 25 professors and 16 administrators in single occupant offices over 10-minute intervals for 3 months Luo et al. (2017) find that the average presence duration is exponentially distributed (Kolmogorov-Smirnov goodness-of-fit at 1%) for researchers and professors but not for administrators. Since for the last case too, the observed distribution is close to the exponential one, they conclude that the Markov chain model can be used. Analysing 35 single-occupant offices over 15-minute time intervals for more than a year, Wang et al. (2005) could not achieve a significant goodness-of-fit but came to the same conclusion for presence duration. Additionally, they found that dividing the observations by clock time did not improve the fit. Finally, from a preliminary analysis (by the author) carried out for an open plan office in Eindhoven, Netherlands, at-desk sojourn times were also found to be close to exponentially distributed although once again a significant goodness-of-fit was not obtained. Here, too, the aggregation interval used for smoothing out noise is 15 minutes but the observation period is only one day. While the sensors used here return the coordinates of individual occupants, the assumption used to detect whether a desk was occupied and the nature of open plan offices may have added some noise to the data aside from the intrinsic false positives and negatives. More information on the data and analysis can be found in appendix B. It should be noted that while the fitted distributions are exponential (which is indeed the distribution produced in continuous-time Markov chains), their discretized forms would be geometric as is the case with the model used. The above observations and fitted distributions are shown in Figure 18.

Regarding away-time for recurrent activities, Tabak (2008) proposes using a logistic growth rate to describe the increasing utility of performing a recurrent activity since it was last carried out. These assumptions are similar to the need-based approach to activity generation used by Arentze and Timmermans (2009). Using a stated choice experiment he finds that the utility of taking breaks and going for a drink indeed follows the expected ‘s-curve’ (cumulative distribution function of logistic distribution). However, the Markovian model would have a geometric distribution of away time whose cumulative distribution function would have only one inflection point. Since the first inflection point of the curves obtained by Tabak (2008) are at a relatively small time interval (Figure 19) it is assumed that geometric distributions of time-intervals may be used.

Since these fixed geometric distributions may not be compatible with some applications, researchers (e.g. (Kolodziej et al., 2011; Xia et al., 2011)) have chosen to use a semi-Markov or Markov renewal process models which allow the use of any probability distributions for the time-spent in a state. However, for the current model a simple discrete-time Markov chain will be used.

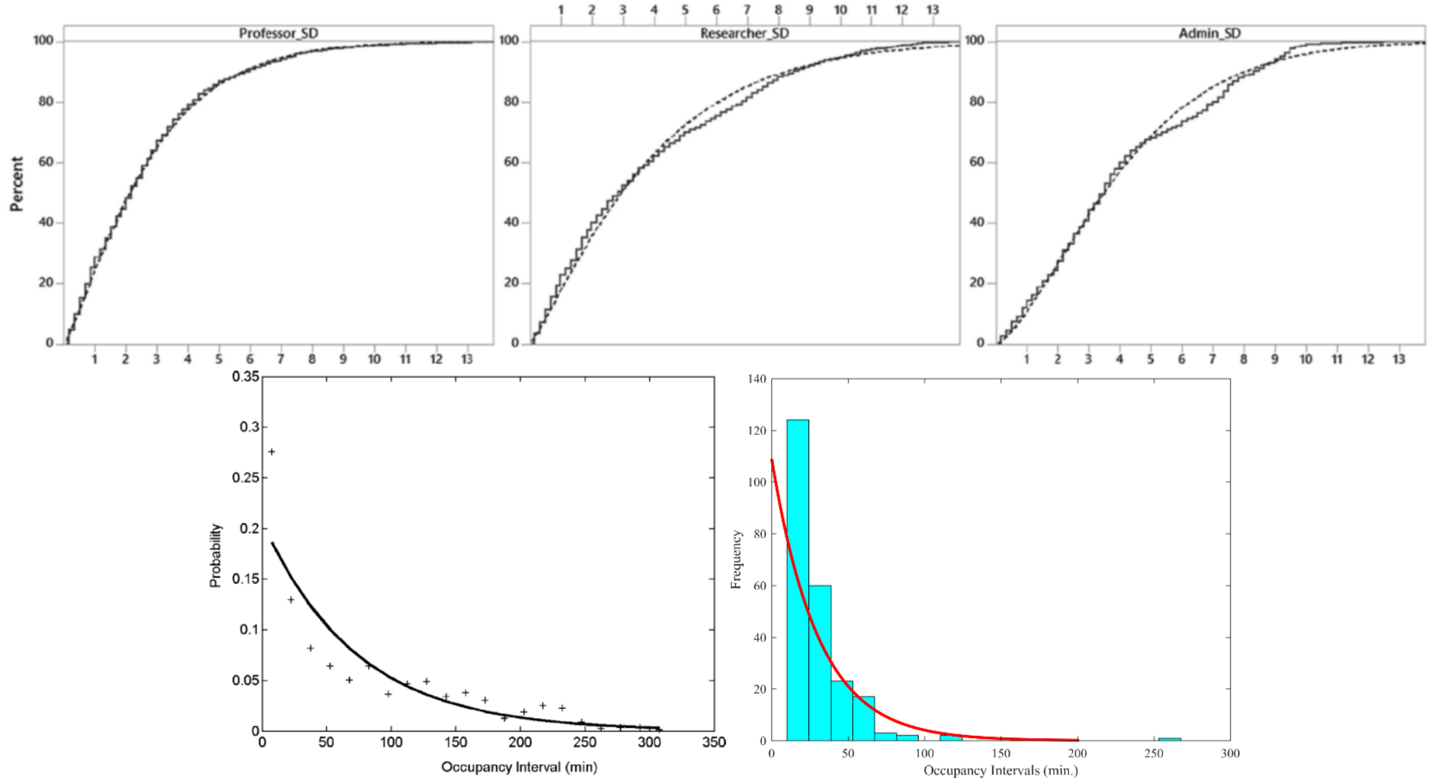

Figure 18: Occupancy intervals fitted to an exponential distribution in single-occupant offices (top, from (Luo et al., 2017) (duration in hours); bottom-left, from (Wang et al., 2005); bottom-right desks in an open plan office)

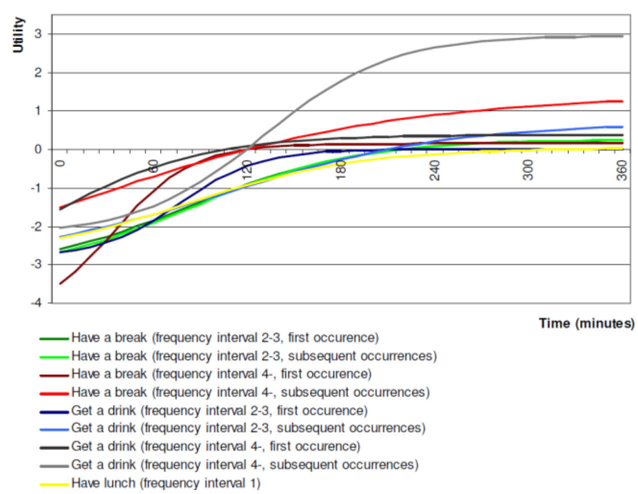

Figure 19: Logistically distributed utility growth against time interval for recurrent activities performed with different daily frequencies and occurrences (from (Tabak, 2008))

4.3.6 Analysis

Previously, it was discussed that the transition matrix for the Markov chain scheduler was derived from an under-determined system of linear equations with non-negativity constraints. For such a system of equations there may be either no solution or an infinite number of solutions leading to the questions *“which input combinations will lead to a valid solution?”* and *“which solution from the infinite solutions should be used to form the transition matrix?”*. Neither the possibility of not finding a transition matrix nor the choice of a particular transition matrix from infinitely many solutions has been discussed either in (Wang et al., 2011), where the approach is introduced, or in subsequent studies (Chen et al., 2017; Feng et al., 2015) that utilize this approach. In order to better understand these factors, in this section, the following aspects of the Markov chain scheduler are studied: (i) the range of valid input combinations, (ii) the range of solutions for a valid input, (iii) the impact of constraining certain transition matrix elements on other elements, and (iv) the behaviour of different transition matrices with the same input upon simulation.

Input Combinations

Here, the aim is to first, demonstrate that not all combinations of steady state probabilities and sojourn times can be associated with a Markov chain and second, describe the space of valid inputs. In order to understand which combinations of input are valid, transition matrices are generated using the method described in section 4.3.4 by varying the steady state probability in small increments while fixing the sojourn time input. The objective function value associated with each matrix thus generated describes the validity of the solutions. Figure 20 shows such a simulation for three different sojourn time inputs for three-state chains. Three-state chains are considered since they are the simplest form of chains on which such analysis can be carried out and for which the results can be visualized in two dimensions. The x- and y-axes in the figure give the steady state probability values of the second and third states from which the steady state probability of the first state can also be calculated as the sum of probability is equal to one (equation 14). The colours indicate the base 10 logarithm of the objective function value which should thus be as negative as possible (to get closer to zero) for a valid solution with the assumed tolerance value lying at $\log_{10}(10^{-10}) = -10$.

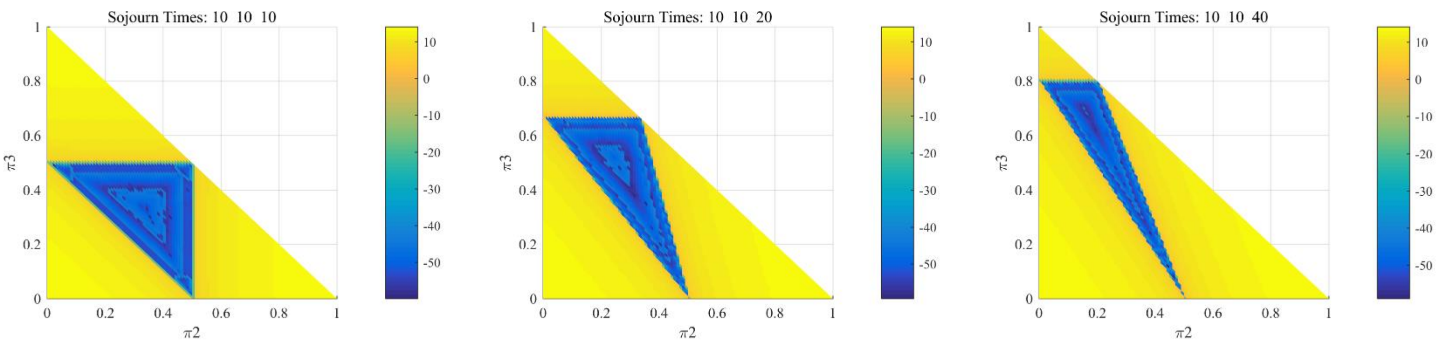

Figure 20: Objective function values for transition matrix generation with different combinations of 3-state steady state probabilities and sojourn times

It can be clearly seen that there is only a specific space (the blue triangle) of input combinations for which a corresponding Markov chain can be generated. This range of steady state probabilities is only affected by the ratio of the sojourn times of different states and not the absolute value. In order to determine the range of valid combinations, the extreme values of steady state probabilities can also be calculated from the sojourn time ratios. To do this, first it should be noted from the above figure that the corners of the valid input combinatorial space are such that one of the steady state probabilities become zero. Further, for each corner point the values of the non-zero probabilities are equal to the ratio of the respective sojourn times to the sum of the sojourn times of the non-zero stationary probability states. For example, for the rightmost diagram in Figure 20, the three corner points of the steady state probability combinations are given by:

$$\rho_1^{corner} = \begin{bmatrix} 0 & \frac{10}{10+40} & \frac{40}{10+40} \end{bmatrix} = [0 \quad 0.2 \quad 0.8]$$

$$\rho_2^{corner} = \begin{bmatrix} \frac{10}{10+40} & 0 & \frac{40}{10+40} \end{bmatrix} = [0.2 \quad 0 \quad 0.8]$$

$$\rho_3^{corner} = \begin{bmatrix} \frac{10}{10+10} & \frac{10}{10+10} & 0 \end{bmatrix} = [0.5 \quad 0.5 \quad 0] \qquad [24]$$

Therefore, the upper bound of each state's stationary probability can be determined by finding its maximum value amongst all the corner points; in the above example, the upper bound for π_1 is 0.5. Moreover, it can also be seen that as the sojourn time ratios become more skewed the space of valid steady state probabilities reduces. These observations also hold for more than three states and therefore, given the sojourn times, the upper bounds of the stationary probability of each state can be supplied to the model user in order to reduce erroneous inputs.

Finally, a small note regarding the corner points: when any steady state probability becomes zero the corresponding Markov chain is no longer an ergodic chain. This is because the state with zero stationary probability cannot be recurrent as for $t \to \infty$ the proportion of time spent in that state vanishes. Therefore, in the Markov chain scheduler used in the model, a check is made to detect and remove states with zero stationary probability when generating the transition matrices. It is assumed that such states or activities are never visited by the pedestrian. Upon generating the matrix elements for such cases, they are arranged into a transition matrix of the same number of states as other but both the row and column of the state with zeros stationary probability are explicitly assigned zeros.

Transition Elements Range

Given a valid input combination of long-term proportion and sojourn times it is implied that at least one solution exists. As suggested by the under-determined system of equations, there are an infinite number of solutions possible. Here, first, the existence of multiple solutions is shown through simulation, and then a formulation of the range of solutions, obtained through empirical observations, is presented. Since the MATLAB solver, *lsqlin*, only produces one solution per input, to produce the full range of values in each element, each element in the transition matrix is constrained to a fixed value using the additional linear constraints described in section 4.3.4. These fixed values are increased from zero to the maximum value possible in small increments per iteration per element. Since the sum of rows in the transition matrix sum to 1 (equation 14), the maximum value possible for an element is constrained by the diagonal element in its row, and is given by $1\text{-}p_{ii}$ or, as described in equation 21, $\frac{1}{\tau_i}$. Figure 21 shows the solutions possible for each element in the transition matrix with the x- and y- axes representing element and objective function (logarithmic axis) values respectively. The horizontal red line indicates the assumed tolerance value of 10^{-10} and the vertical red line represents the maximum value the element can take given the sojourn times that fix the diagonal elements. Finally, the red dot represents the solution obtained by the MATLAB solver.

It can be seen from the figure that although it has been derived using a finite number of increments, if the whole range of values under the tolerance line is considered there are infinite number of solutions. As with the valid input space, the range of solutions possible are also not affected by the absolute values but rather the ratios of the stationary probability and sojourn time values (except the maximum value constraint which depends on the absolute value of sojourn times). In fact, the formulae for the range of values presented below mainly utilise the ratio of stationary probability and sojourn times. For an intuitive understanding, the range of solutions describe a trade-off between the expected sojourn time and the steady state probability of the states. For example, although the chain is expected to spend 35% of the time in the second state, its average duration in that state is only 1/4th as compared to the others. Thus, it can be seen that transitions to the second state are biased towards higher values. Furthermore, whether the elements have a large or small range depends on the where the input lies in the input space described in Figure 20. For inputs lying near the edge of the valid space the range of solutions will be generally smaller than for those lying in the centre.

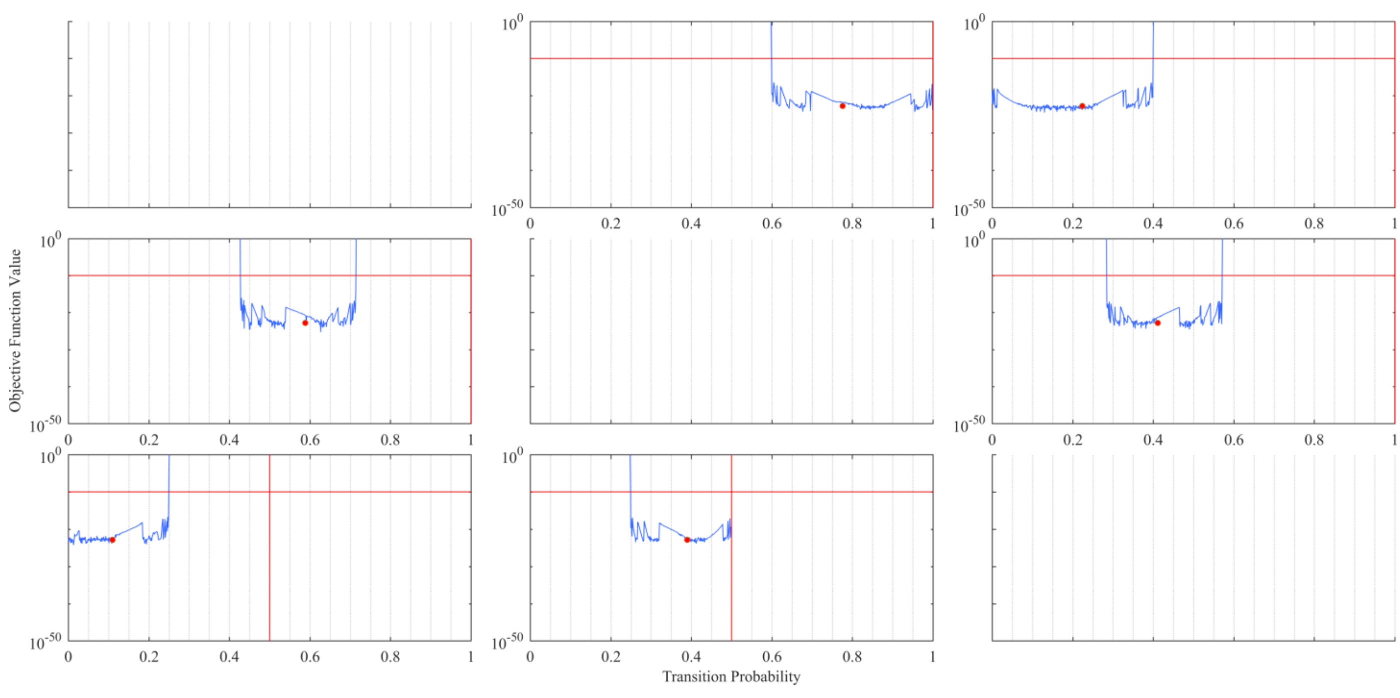

Figure 21: Transition matrix elements range for $\pi = [0.25\ 0.35\ 0.40]; \tau = [1\ 1\ 2]$. Horizontal and vertical red lines describe the threshold of acceptable objective function value (10^{-10}) and the maximum value the transition matrix element can take respectively. Red dot represents the solution found by the MATLAB solver.

Equations 25–27 present a succinct formulation to calculate the minimum and maximum transition element values possible given a valid input. It should be noted that these values have not been determined theoretically but through observations. However, they have been verified for multiple examples including the inputs used as in (Wang et al., 2011) where the existence of multiple solutions has not been considered.

$$\beta_i = \frac{\pi_i}{\tau_i} \qquad [25]$$

$$p_{ij}^{min} = \max\left(\frac{\beta_j - \sum_{\substack{k\neq i\\ k\neq j}} \beta_k}{\beta_i}, \frac{\beta_i - \sum_{\substack{k\neq i\\ k\neq j}} \beta_k}{\beta_i}\right) \times \left(\frac{1}{\tau_i}\right) \;\; s.t. \forall i, j \; p_{ij}^{min} \geq 0 \qquad [26]$$

$$p_{ij}^{max} = \begin{cases} \min\left(\frac{\beta_j}{\beta_i}, 1 + \frac{\beta_j - \sum_{\substack{k\neq i\\ k\neq j}} \beta_k}{\beta_i}\right) \times \left(\frac{1}{\tau_i}\right), & n = 3 \\ \min\left(\frac{\beta_j}{\beta_i}\right) \times \left(\frac{1}{\tau_i}\right), & n > 3 \end{cases} \;\; s.t. \forall i, j \; p_{ij}^{max} \leq \frac{1}{\tau_i} \qquad [27]$$

Solving these formulae for the above input (Figure 21) returns the following output:

$$p^{min} = \begin{bmatrix} 0 & 0.6 & 0 \\ 0.43 & 0 & 0.29 \\ 0 & 0.25 & 0.5 \end{bmatrix}, p^{max} = \begin{bmatrix} 0 & 1 & 0.4 \\ 0.71 & 0 & 0.57 \\ 0.25 & 0.5 & 0.5 \end{bmatrix}$$

The above formulae are interesting because they convert the intuitive relations of transition elements, stationary probabilities and sojourn times into verifiable mathematical relationships. In order to choose from the many solutions possible, one of the options is to add additional constraints to the model in the form of transition relationships observed in the real world. Using these formulae, model users can be made aware of the range of solutions possible for a given input and thus the possibility to add additional valid constraints to the Markov chain generation whenever more data is available.

Transition Elements Variation

When the value of one of the elements is constrained, the values of the other elements also change in order to maintain the average statistics of steady state probabilities and sojourn times. To demonstrate this, the values of p_{12} from the above considered inputs are varied from the minimum possible value to the maximum value (or less) with constant increments as shown in Figure 22. While, for the simple three-state chain the values vary from end to end in an almost-linear fashion, this is not necessarily the case for Markov chains with a higher number of states where the relationships are much more complex.

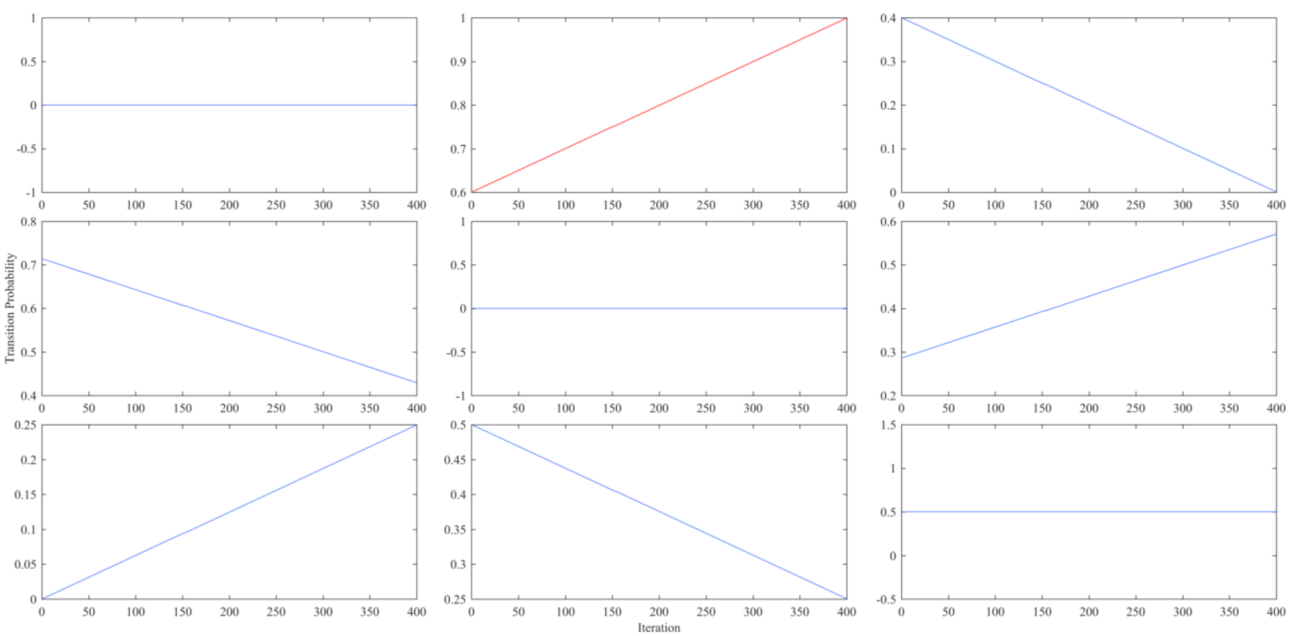

Figure 22: Variation of transition element (red) and its effect on other elements for $\pi = [0.25\ 0.35\ 0.40]; \tau = [1\ 1\ 2]$

Transition Probability Simulation

The Markov chains derived above by varying a transition element are used here to understand their behaviour upon simulation. In the model, the Markov chains are simulated repeatedly for a maximum time period of a day. Since all the results in section 4.3.3 are true for expected values at $t \to \infty$, the limited time simulations add stochasticity to the results. To check whether, then, the existence of different solutions makes any difference to the simulated transition probabilities, four different Markov chains from the various matrices obtained above are simulated 1000 times for 500 time steps (500 minutes ~ 8 hours ~ 1 working day). Figure 23 and Figure 24 shows the results of this simulation. The four colours indicate different Markov chains and the mean values of the simulations are indicated by vertical lines of the same colour. The only difference between the two figures is the sojourn times; in the second figure, the sojourn times are 10 times that of the first. Once again, the diagonal elements are more or less constant because they are fixed by the input.

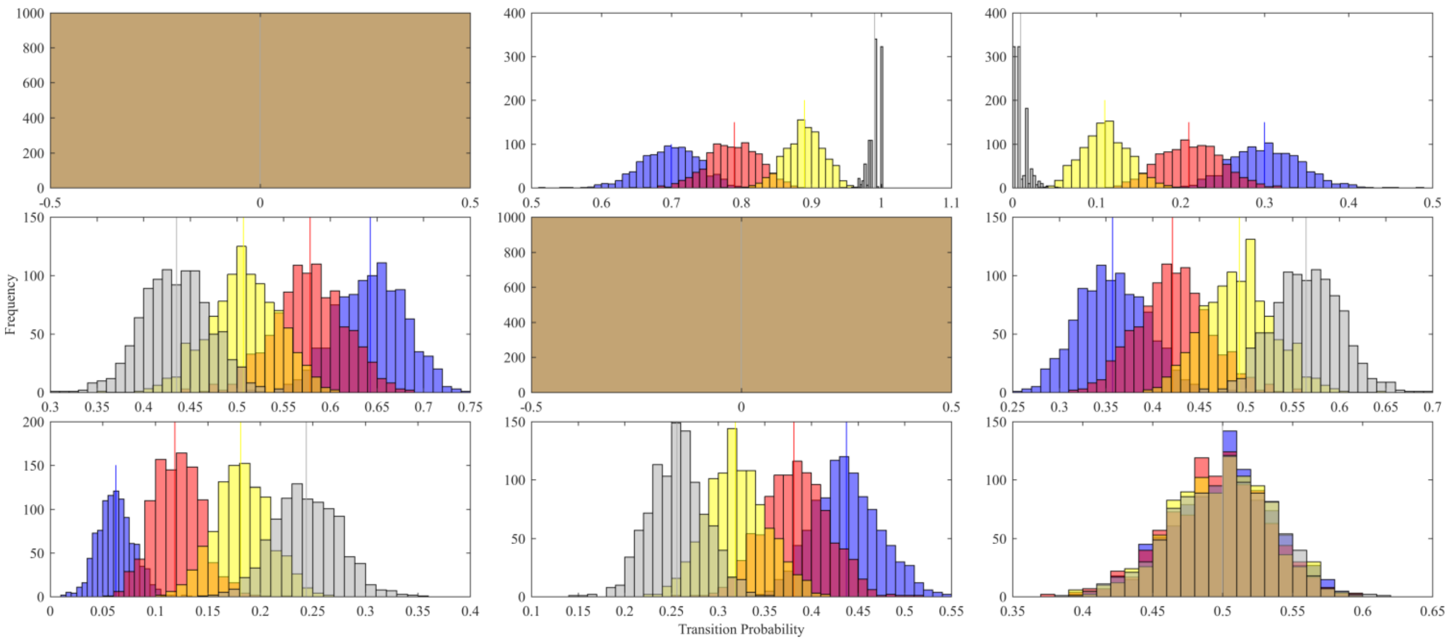

Figure 23: Simulated transition probabilities for four different Markov chains generated from $\pi = [0.25\ 0.35\ 0.40]; \tau = [1\ 1\ 2]$

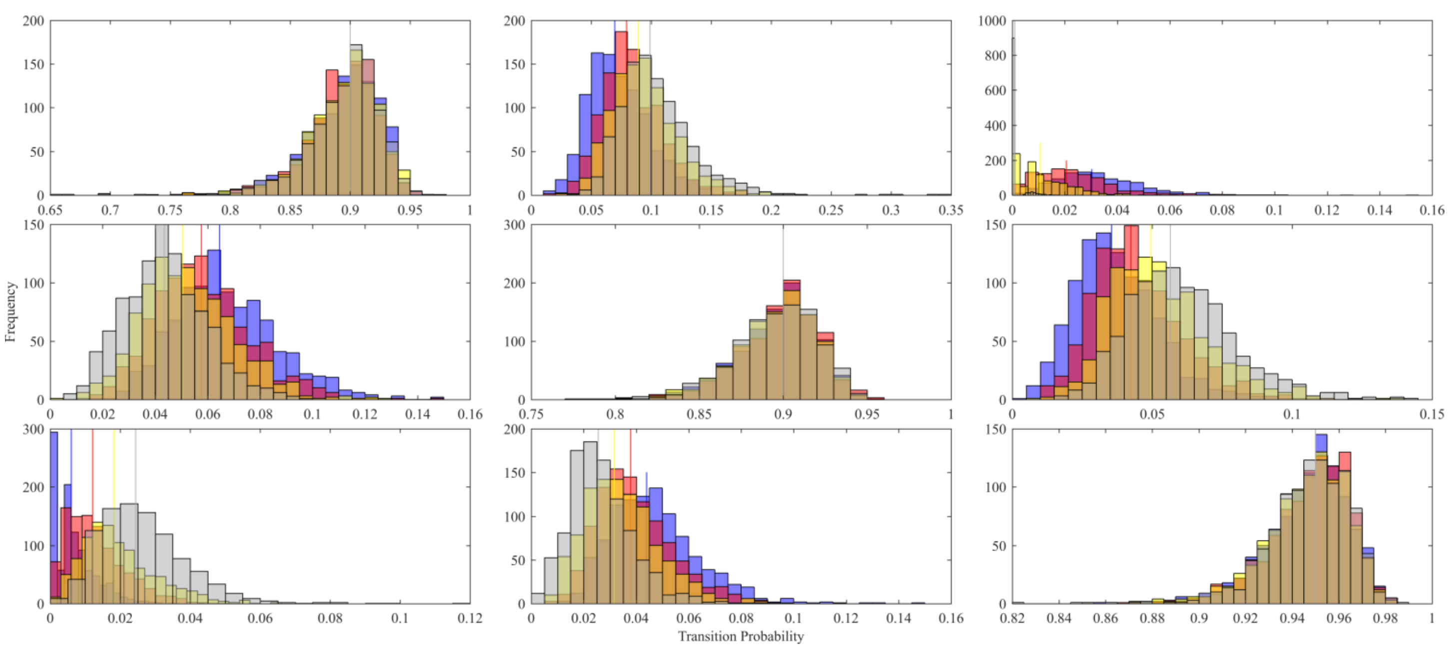

Figure 24: Simulated transition probabilities for four different Markov chains generated from $\pi = [0.25\ 0.35\ 0.40]; \tau = [10\ 10\ 20]$

It can be observed that for smaller values of sojourn times the effects of different transition matrices can be clearly observed in the simulated transition probabilities but the differences are much less for the same sojourn times multiplied by 10. This is because while the same stochasticity is added to the model, or put differently, while the standard deviations of the simulated probabilities remain more or less the same (except when constrained by the minimum limit), the mean values have been reduced by a factor of 10. Thus, for states with a high sojourn time, in absolute terms, the transitions from that state will produce similar results for different transition matrices while those with a low sojourn time will be impacted more strongly by the transition matrix choice.

4.4 Building Representation for the Tactical Level

While the operational level only makes use of the nodes in a building plan as origins and destinations, in order to decide the origins and destinations, at the tactical level they have to be associated with functions or activities. To this end, nodes in the building are divided into service and non-service locations. As shown in Figure 25, service locations are all those nodes that are associated with a service time and certain queuing characteristics; for example, nodes in front of coffee machines. Non-service locations are not associated with a service time and the time spent in the nodes is determined by the activity scheduler. Furthermore, such nodes can be allocated to occupants for certain periods of time; for example, chairs where occupants work may be assigned to them either permanently or temporarily (in flexible workspaces).

Further, as in (Tabak, 2008), abstract spaces are also defined to represent functionalities associated with certain areas in the building. Thus, abstract spaces, as the name suggests, are not associated with a particular physical object in the building but are composed of different locations and obstacles to which they add contextual information that is not available from the components themselves. For example, a collection of chairs around a table may make up a meeting room. Abstract spaces are required when it is known which space or room the agents need to be in but the exact location is unknown. For example, they are required to select rooms for meetings or select base location in a flexible workspace. Having defined abstract spaces, it is also possible to add other information such as the facilities available in a meeting room for more complex activity location decisions.

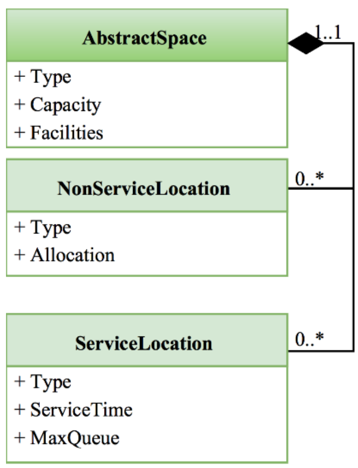

Figure 25: Class diagram of building representation for the tactical level. Colours signify cohesive classes as described in section 5.2.

4.5 Location and Route Choice

Location choice takes place during activity scheduling and the choice is made from all the locations associated with the activity to be scheduled. For meeting rooms, it is assumed that occupancy level is optimized and thus spaces with capacity closest to (but never lesser than) the number of people participating are preferred. Base locations in flexible workspaces are also chosen randomly on a first-come-first-serve basis with re-addition of spaces into the pool of choices as they become free. Finally, it is assumed that when choosing locations for all other activities, agents choose the nearest possible from their current location because in offices it is expected that locations are more or less similar and choosing farther destinations do not offer any added benefit.

In addition, activity location choice is affected by the availability of locations which is managed by the resource handler. If there is a queue in front of a service, the resource handler checks the time it would take to get the service at other locations for which, in addition to the walking times to the other locations from the current location, it requires the service time of the locations and the queue lengths there. The check for whether there is a queue in front of a service takes place when the agent is about to reach the service. However, for the second location choice, it is assumed that the agent has information about the queue lengths in front of all the alternatives including the ones that cannot be seen. This is done to represent that pedestrians make some estimation about the queues at other places after seeing the queue at one location. When a non-service location is occupied, if available, the nearest possible alternative location is chosen. Finally, some locations such as base locations need to be allocated to individuals so that they are not occupied by others. Although not implemented here, for more complex models, locations can also be allocated for activities other than continuous when it is expected that the agent will return to it soon. For instance, when a person leaves a break room chair to come back with a drink there is an understanding that the chair is allocated to that person until they have left without intention to immediately return.

As stated in the introduction of this chapter, for routing decisions, it is assumed that occupants prefer to take the shortest route between locations and are able to do so because they have full spatial knowledge of the building. Thus, Dijkstra's algorithm (Dijkstra, 1959) is used to find and store the shortest paths and their costs from the navigation graph which is then accessed for decisions regarding activity location as well as for passing routing decisions to the operational level.

5 Model Implementation and Case Study

An integrated but modular framework consisting of all the pedestrian behaviour levels: strategic, tactical, and operational is presented in the form of a simplified class diagram in UML to give an overview of the model and aid future object-oriented implementations. The current model is implemented in MATLAB as per the model requirements. This implementation of the entire model from input to output is discussed in detail with the help of flow diagrams that help verify the logical flow. In parallel, a case study of an imaginary office is used to exemplify the methods and verify the results of various parts of the model. Additionally, the pedestrian behaviour model is qualitatively validated by observing expected macroscopic patterns in aggregate measures obtained from the case study.

5.1 Introduction

In the previous two chapters, conceptual models for the three pedestrian behaviour levels were presented. In this chapter, these models are put together and the implementation of the complete model is discussed with the help of flow diagrams alongside a running example that is used as a case study. The case study comprises of an imaginary office with 18 occupants and a relatively simple building plan and organization structure. After the model implementation, some model applications and results are presented from the case study.

The chapter is structured as follows: first the framework of the integrated model is presented (section 5.2) followed by a discussion of the model input (section 5.3), implementation of the operational (section 5.4), strategic (section 5.5) and tactical (section 5.6) levels, and the model output (section 5.7). Case study results are described in section 5.8. Throughout the chapter, the parts on the case study are indicated by grey boxes. Finally, model limitations are discussed in section 5.9 as part of the model assessment.

5.2 Framework

Figure 26 integrates various models discussed in the previous two chapters into a single modular framework presented in the form of a simplified class diagram in the Unified Modelling Language (UML). The class diagram gives an overview of the model and describes relationships between various parts of the model and helps translate the conceptual model into an executable implementation. Despite the object-oriented framework described in the class diagram, the current implementation of the model is done in MATLAB in a non-object-oriented fashion but the modular structure is maintained.

The three pedestrian behaviour levels, strategic, tactical, and operational, can be clearly demarcated within the framework. Furthermore, cohesive classes are grouped into: (i) the organization (yellow), (ii) the individual (red), (iii) activities (blue), (iv) schedulers (purple), (v) functional spaces (green), and (vi) physical spaces (orange). The organization and activity classes define the activities to be performed and belong to the strategic level; the activity schedulers, the functional spaces which decide activity location, and routing classes are in the tactical level; and finally, the physical spaces which define the walking facility are used for the operational level. Individual agents are composed of three personas, two – organization-related and non-organization related – concerned with strategic level decisions, while one decides its walking characteristics. The modularity of the framework is an important asset that allows users to easily modify different parts of the model as and when required. For example, in the current model the route choice algorithm simply assigns the shortest routes to pedestrians but in case future applications require a more complex model it is easy to replace the routing module with another one. Moreover, although some features may be missing, the different models can run independent of one another. For instance, one can obtain detailed agent activity schedules by running the tactical level model without the operational level model although in this case execution level constraints such as capacity will not be considered.

The interfaces between the three behaviour levels are for their integration and describe how they interact with each other. The interactions between behaviour levels is shown in Figure 27 where it can be seen that not only does interaction occur from longer- to shorter-term decision making but there are also feedback loops. Arrow 1 indicates the passing of time-dependent activities that are to be performed and individual profiles for time-independent activities from the strategic level into the appropriate schedulers in the tactical level. As discussed earlier, the Markov chain scheduler does both, generation and scheduling, for the time-independent activities. The tactical level is connected to the operational level in three ways. First, the physical spaces, obstacles and positions, are assigned functions in the tactical level making them functional spaces for activity location choice; second, routing decisions at the tactical level are used to execute movements at the operational level; and third, the executed schedule at the operational level dynamically changes the planned one through the re-scheduler in the tactical level. Arrow 3 denotes that the tactical level passes a planned schedule of movements to the operational level while the arrow 4 describes the dynamic feedback that changes this schedule. If a location to perform activities scheduled by the Markov chain scheduler cannot be found, they are cancelled by the re-scheduler. However, since these activities are generated at the tactical level they do not represent the interaction suggested by arrow 2 where long-term decision making is changed but can also be denoted by arrow 4.

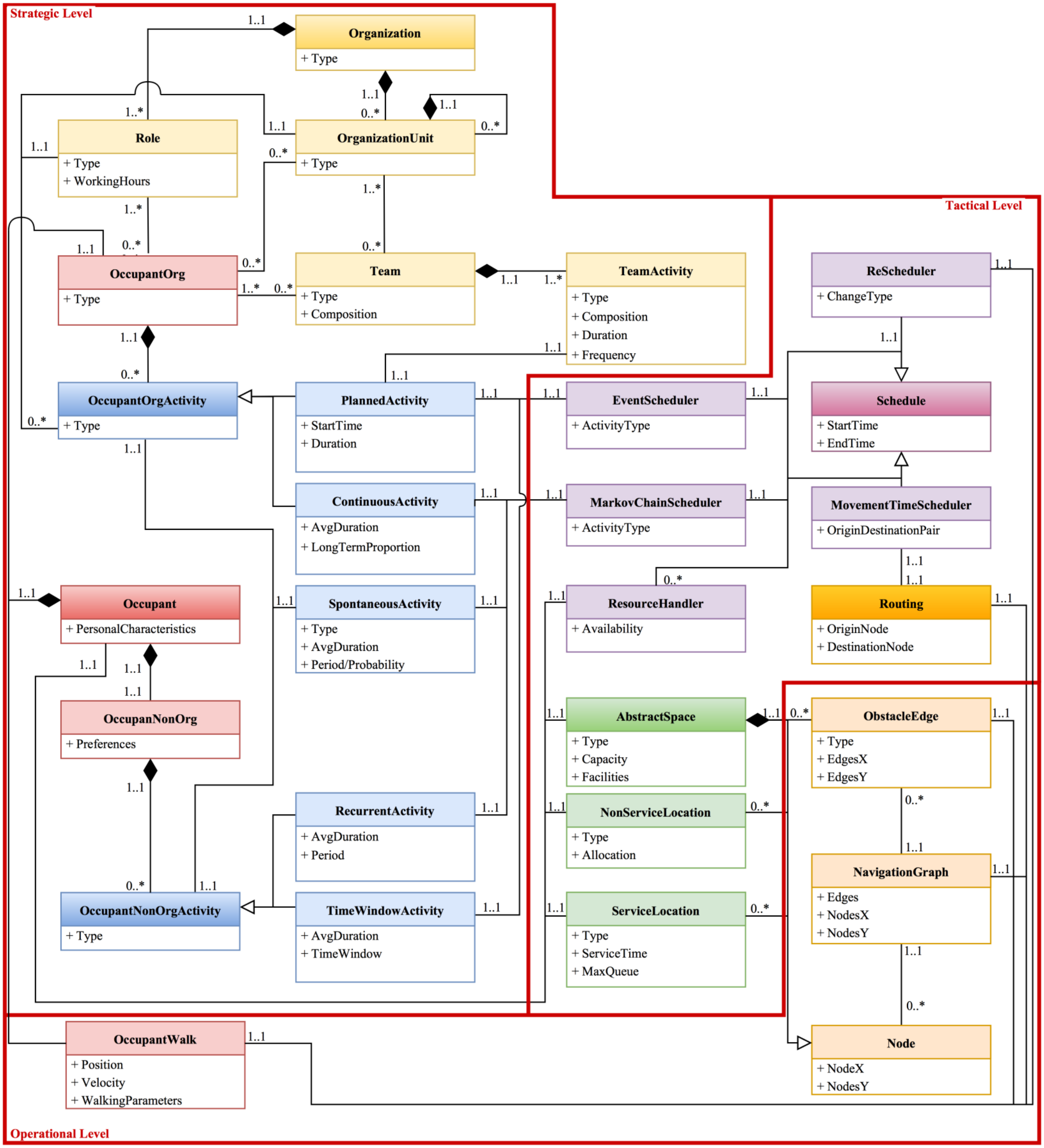

Figure 26: Class diagram of the complete pedestrian model. Similar colours indicate cohesive classes.

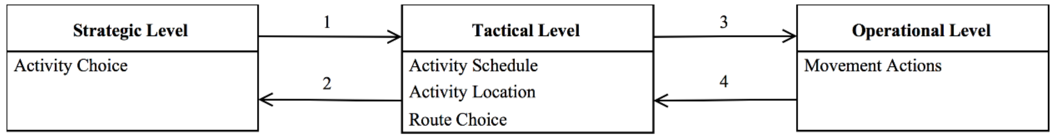

Figure 27: Pedestrian behaviour levels and their interaction based on (Hoogendoorn, 2001). Arrow 1: Time-dependent activity episodes and occupant profiles for time-independent activities; arrow 2: feedback from tactical level (does not occur in this model); arrow 3: schedule of movements; arrow 4: feedback from operational level for re-scheduling

5.3 Model Input

The model requires two types of inputs from the user; first, the plan and certain characteristics of the building within which occupant movements have to be modelled, and second, information for the strategic level such as the organization structure and occupant profiles. Default values for occupants' operational level characteristics are already made available in the model but if required they can also be changed.

5.3.1 Building Plan

The model uses building plans prepared in AutoCAD according to certain rules such that they can be correctly interpreted. While detailed instructions for this can be found in appendix C, here, the main ideas are discussed. A building plan on its own only represents the physical spaces which are indifferentiable from one another, which is why it has to be annotated appropriately to add functions. In AutoCAD, this annotation is done through drawing layers. Each plan is composed of five types of layers, each unique in its function: (i) non-service locations, (ii) services, (iii) wormholes, (iv) obstacles, and (v) doors while a sixth, purely functional, layer, abstract spaces, gives functions to areas in the building and can contain instances of the above five types within it. Non-service locations are points where activities not associated with a service time take place (see section 4.4), services identify obstacle edges as being associated with particular services (see section 4.4), wormhole edges allow the model to transport pedestrians from one point to another within a given time without using the social forces model (see section 3.3.3), obstacles are edges that act as boundaries to the walkable area (see section 3.4), and finally doors are edges denoting the separation between two such areas. For services and wormholes, the respective layers are used to mark obstacle edges in front of which service and wormhole nodes are automatically derived within the model. Although not used in the current implementation, at the operational level the doors layer can be used to explicitly model a brief waiting period, as experienced in front of automatic doors for example.

In addition to associating locations and spaces with activities, the user needs to define their capacity too. The model permits locations to have a capacity of more than one which can be used to build vertical queues. If base locations are permanently allocated, that is individuals have their own chairs, then this should also be given as input otherwise the spaces where users can sit should be defined and they will then choose their own temporary base locations.

Case Study: Building Plan

Figure 28 shows the office building plan used for the case study as both, physical and functional spaces. Although the plan contains only one floor, the lift and stairs are shown to describe how vertical movements may be modelled. Positions behind tables indicate chairs, and the location outside acts as the source or sink for people coming from or going outside the building. Chairs in the closed office, the open plan office on the right, and behind the reception desk are allocated to certain occupants but those in the left hand side open plan office which is a flexible workspace, the meeting room, and in the break area are not allocated. As described in section 4.4, where agents know the space they have to be in but not the exact location, an abstract space has to be defined which can be seen in Figure 28 (right). Also note that the plan does not contain service or wormhole nodes which are derived from the edges in the model.

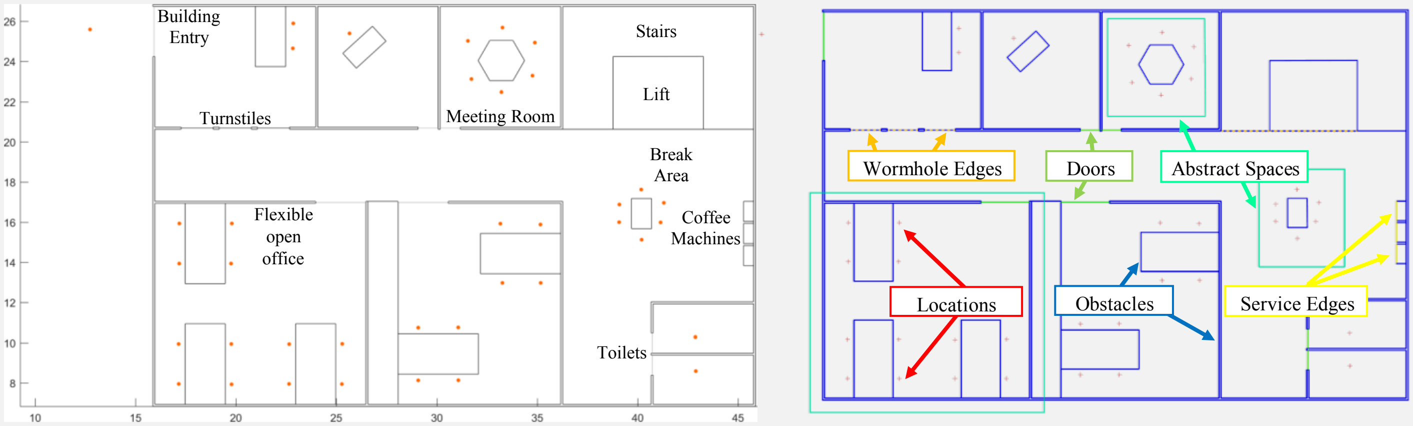

Figure 28: Case study: Building plan – physical spaces as obstacles and positions, and text annotations (left; distance marks in meters) and as functional spaces with functions embedded as layers in the AutoCAD plan (right)

5.3.2 Strategic Level Input

In the current implementation, the organization characteristics are only used to define team activities for which the number of people, their roles, teams and team activities have to be defined. For each team activity, the average weekly frequency is used as input. The roles are further kept in mind when defining the individual attributes for organization-related continuous activities. While not modelled here, if unplanned interactions are to be modelled, a matrix of individuals representing the likelihood of interactions can be input in order to decide who goes to meet who. Next, in order to define individual probability profiles, first the activities to be considered have to be listed. For each activity scheduled by the Markov chain scheduler, the long-term proportion of time spent and the average duration should be given. As mentioned previously, the long-term proportion of time can also be calculated from the away time and the fact that the proportions should add up to 1. For time-window activities the start time, duration and their uniform variation is used as input.

Case Study: Strategic Level Input

The office considered consists of 18 occupants with 4 different roles: Manager (1), Senior Engineer (4), Junior Engineer (12), and Receptionist (1). 3 teams exist each with one team activity each as shown below. All occupants do not have to be part of a team as can be seen here.

Table 4: Case study: Team definitions

Team ID	Manager	Sr. Eng.	Jr. Eng.	Receptionist
1	1	4	0	0
2	1	1	3	0
3	0	2	4	0

Table 5: Case study: Team activity definitions

Team Activity ID	Weekly Frequency	Duration (minutes)	Team ID	Manager	Sr. Eng.	Jr. Eng.	Receptionist
1	1	60	1	1	4	0	0
2	1	60	2	1	1	3	0
3	1	60	3	0	1	2	0

The activities modelled are shown below. For the four Markov chain activities, ‘At Desk’, ‘Toilet’, ‘Coffee’, and ‘Break’, the average away time and duration is used as input while for the time-window activities the start time and duration is required. For the Markov chain activities, it can be seen that the long-term proportion is much more intuitively obtained from the away time than directly estimating it. The time-window for time-window activities is defined by the variation allowed around the start time. The occupant profiles used here are based on the roles, for example, it is assumed that the receptionist does not take breaks and arrives and has lunch earlier than the others. On the other hand, the manager is modelled to arrive and have lunch later than others.

Table 6: Case study: Activities modelled

Continuous	Recurrent	Spontaneous	Time-window	Planned
At Desk	Toilet	-	Arrival	Meeting
	Coffee		Departure	
	Break		Lunch	

Table 7: Case study Attributes for Markov chain activities

MC Activities	Away Time (minutes)	Average Duration (minutes)	Long-term Proportion
At Desk	-	75,120,135	1-(0.04+0.0164+0.04) = 0.9036
Toilet	120	5	5/(5+120) = 0.04
Coffee	120	2	2/(2+120) = 0.0164
Break	240	10	10/(10+240) = 0.04

Table 8: Case study: Attributes for time-window activities

Time-window Activities	Start Time	Max/min variation (minutes)	Duration (minutes)
Arrival	0730h,0900h,1030h	5,15,30	-
Departure	1630h,1700h,1730h,1830h	5,15,30	-
Lunch	1100h,1200h,1400h	5,30	20

5.4 Operational Level

After the building plan has been imported and parsed, wormhole and service nodes are added and the navigation graph is created using the method based on (Kneidl et al., 2012) as described in section 3.3.1. This navigation graph is used by the tactical level to extract shortest routes between locations and subsequently by the operational level to execute movements from one location to another.

Case Study: Navigation Graph

The figure below shows the navigation graph for the simple office. It can be seen that wormhole and service nodes have been added by the model and that each location is directly connected to at least one navigation node. Since the navigation nodes are also interconnected all locations can be accessed by agents in the model. Currently the stairs and lift are not used.

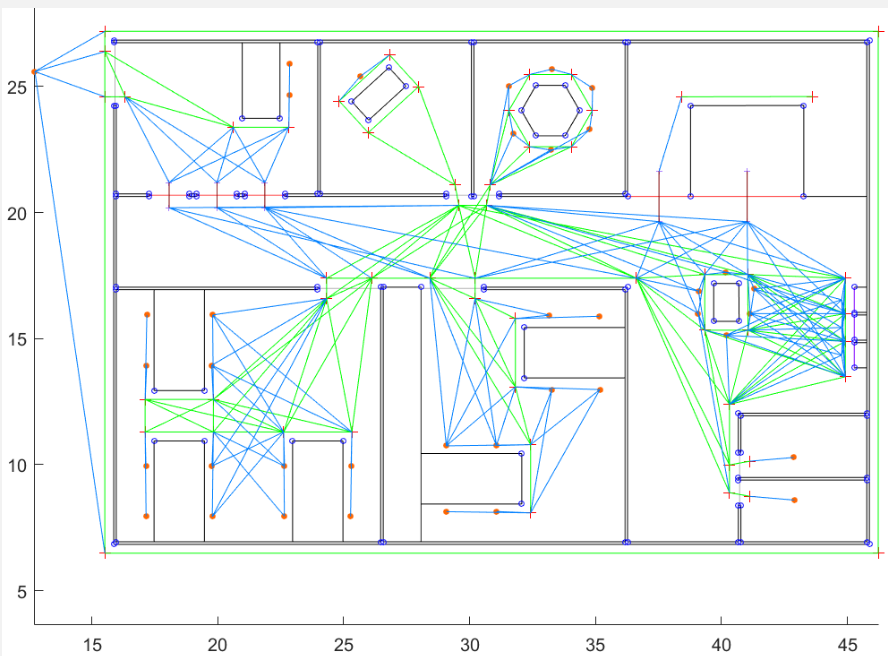

Figure 29: Case study: Navigation graph of the office; blue lines are connections between locations (orange dots) to navigation points (red crosses); green lines are connections between navigation points

As input, the operational level uses the planned schedule obtained from the tactical level (section 5.6) and to execute occupant movements, the social forces model (section 3.3.2) is used. The movement model uses Euler's explicit integration method for the execution as it is straightforward and simple to implement even though it is slow because of the need to simulate very small time-steps and thus perform calculations a large number of times. The model begins at the time the first occupant starts moving and then the algorithm is run in time-step increments. If at some time no occupants are moving, the model either skips to the first moment an occupant begins to move or if no more movements are scheduled the program ends. In order to keep track of the movement status of occupants, the model assigns them two properties. The first indicates whether they are active, that is, whether they are moving, and the second indicates whether they are waiting. Occupants wait if they are in a queue, in transition through a wormhole, or yielding to let other pedestrians pass.

Both inactive and waiting pedestrians are not affected by the movements of other pedestrians and hence do not move but the difference between these states is that inactive pedestrians continue to repel other pedestrians while waiting pedestrians don't. In the current implementation of the model, queues are modelled as vertical queues following the first-in-first-out principle; therefore, it is assumed that pedestrians in such queues do not have a repulsive effect so that other pedestrians are able to approach the location of vertical queuing. Further, since transition movements are assumed to overrule the underlying movement model and describe the pedestrians in a mesoscopic manner they do not have a repulsive effect. Thus, when a pedestrian's route consists of a wormhole transition, the pedestrian approaches the origin wormhole, transitions according to the time in the wormhole adjacency matrix, and then continues following the sequence of nodes in its route. Finally, in a few extreme cases, since contact forces are not modelled, pedestrians may 'jump' over walls due to large forces from several pedestrians in close proximity. Although such crowded situations are not expected during non-emergency situations in office building, to prevent destabilisation of the model it randomly selects one pedestrian out of all those whose velocity gets updated to a value greater than the maximum speed, and makes that pedestrian wait at their location for a given time without repelling other pedestrians. This represents yielding behaviour that one can expect to see in non-emergency situations. If, after forcing one of the pedestrians to yield, the total force on others in the next time step is still too high, the model forces another pedestrian to yield. This goes on until the pedestrians are able to pass, thus ensuring a solution to the blockade.

The operational level model also interacts dynamically with the tactical level model through the re-scheduler (section 5.6) in order to dynamically update the planned schedule.

5.5 Strategic Level

As discussed previously, the strategic level is responsible for deciding which activities occupants perform as well as generating planned and time-window activity episodes. From the strategic level inputs of occupant profiles it is already possible to tell which activities they perform and the number of time-window activity episodes that are generated. For planned activities which are usually team activities in the form of meetings, first teams have to be defined and then team activity episodes can be generated. Using the team definitions in terms of the number of individuals per role, occupants are assigned to a team. Further, using the team activity definition, which may be different from that of the team, occupants are selected for the meeting. Individual membership to teams and team activities is held constant for all the days modelled. The occurrence of a team activity episode is then decided based on its daily probability which is calculated from its weekly frequency. For example, a meeting occurring once a week has a daily probability 1/5 = 0.2. The generation of continuous, recurrent, and time-window activities takes place at the tactical level and the habitual activities are taken directly from the user.

Case Study: Strategic level

The member agents of each team are selected so that team participation is distributed, as far as possible, equally over all agents according to their roles. The table below shows the occupants belonging to the team activities (which remain constant) and the occurrence of the activities for a given day. Here occupant participation in a team activity is derived from the team activity and team composition, and the occurrence is simulated using the input of weekly frequency. It is assumed that the activities can only occur once a day at maximum.

Table 9: Case study: Member agents of team activities

Team Activity ID	Event Probability	Occurs Today?	Team ID	Agent ID			
				Manager	Sr. Eng.	Jr. Eng.	Receptionist
1	0.2	No	1	1	2,3,4,5	-	-
2	0.2	No	2	1	2	6,7,8	-
3	0.2	Yes	3	-	3	10,12	-

5.6 Tactical Level

In order to schedule an individual's activities over a day, first the planned and time-window activities are scheduled by the event scheduler followed by the other activities and movement times filling up the gaps left in the agenda with the Markov chain and movement schedulers. Furthermore, agents' planned schedules are updated dynamically during execution by the re-scheduler. Activity location choices are also carried out by the schedulers.

5.6.1 Event Scheduler

The event scheduler first schedules planned activities and then the lower priority time-window activities. The event scheduler uses a coarse resolution of 1 minute since greater resolution would not be considered when planning activities or time-windows for activities. For planned meetings (Figure 30), the agents involved, duration, and meeting spaces are used as the input. Feasible times for the meetings are the times between the latest expected arrival and earliest expected departure times. Among these times, those that are not occupied by an activity are considered as available time. Similarly, spaces or rooms are also considered to have available time when they are not planned to be occupied. The resource handler is used to find all the common available times for the meeting agents and the lowest capacity room that can hold all members. If among the set of all contiguous available times found by the resource handler there are those of equal or greater duration than the planned meeting duration, then one is randomly selected for the meeting. Once a suitable time is found for the meeting it is blocked in the agendas of the member agents and the selected room. When a contiguous time of required duration is unavailable, another room is selected until all rooms have been checked after which it is assumed that the meeting is cancelled. Furthermore, it should be noted that the scheduler does not optimize time allocation to different meetings so that all generated meetings will be scheduled. Such optimization or rescheduling to fit in more activities may be done by individuals in real life when they expect to have many meetings but this is not considered here. This may lead to occupants spending lesser time in meetings and meeting room occupancy being lower than expected. Finally, the implementation allows starting times to be forced to occur at rounded clock times as has been observed in the real world ((Chen et al., 2017), appendix B) and also allows agents to keep a preferred gap between scheduled meetings.

To schedule time-window activities (Figure 31), their time-window, duration, and associated locations as well as the partly filled agendas of agents from above are used as input. First, the arrival and departure times are selected from the preferred time windows such that they are, respectively, before and after all planned activities. For other habitual activities, the resource handler checks for available times within the time window as well as, in this implementation, after and before the window. Allowing time-window activities to be performed outside their time-window is an optional aspect that indicates that there is a preferred time-window to perform an activity but it is not strict. If contiguous time equal to or greater than the activity duration is available in the time window, one of the periods meeting the conditions is selected. Otherwise, the first or last available period before or after the time window is selected and if the activity cannot be scheduled it is skipped. The activity location is assigned by randomly selecting one of the associated locations.

After running the event scheduler, incomplete agendas for each agent are returned. These agendas or schedules consist of event episodes with gaps of available, unscheduled time between them.

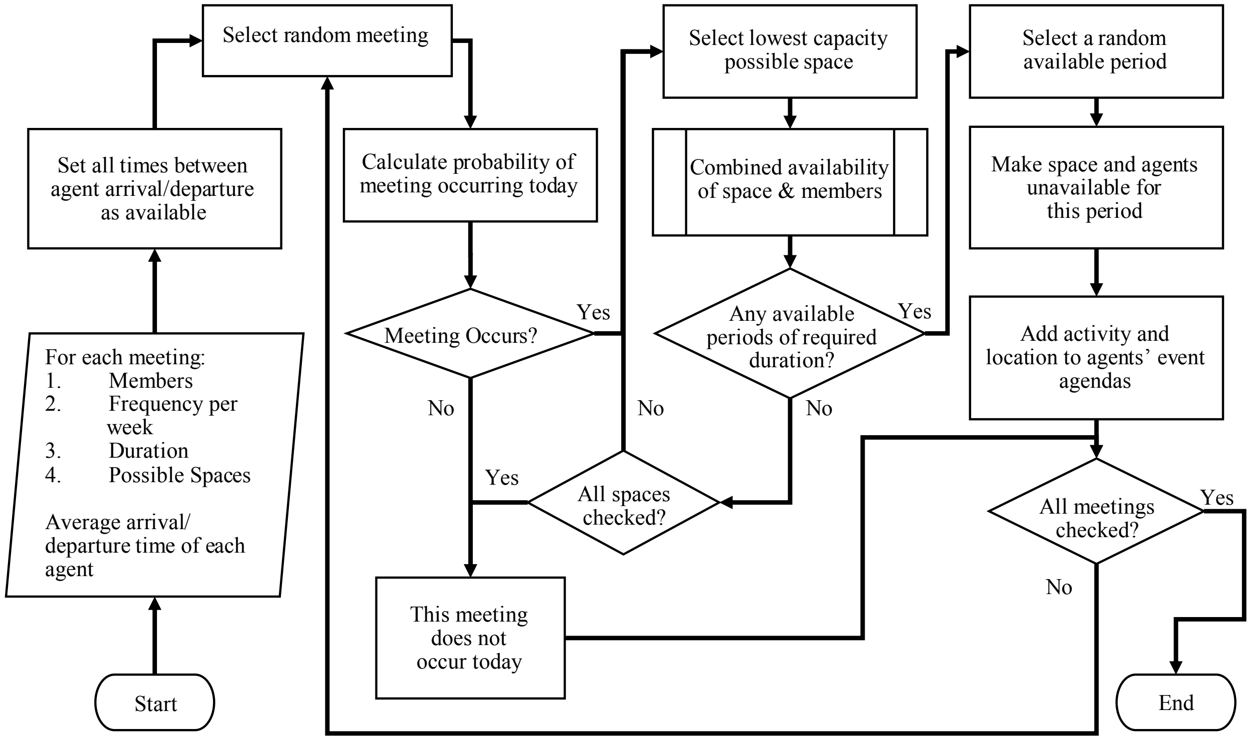

Figure 30: Flow diagram of the event scheduler for planned activities in a day

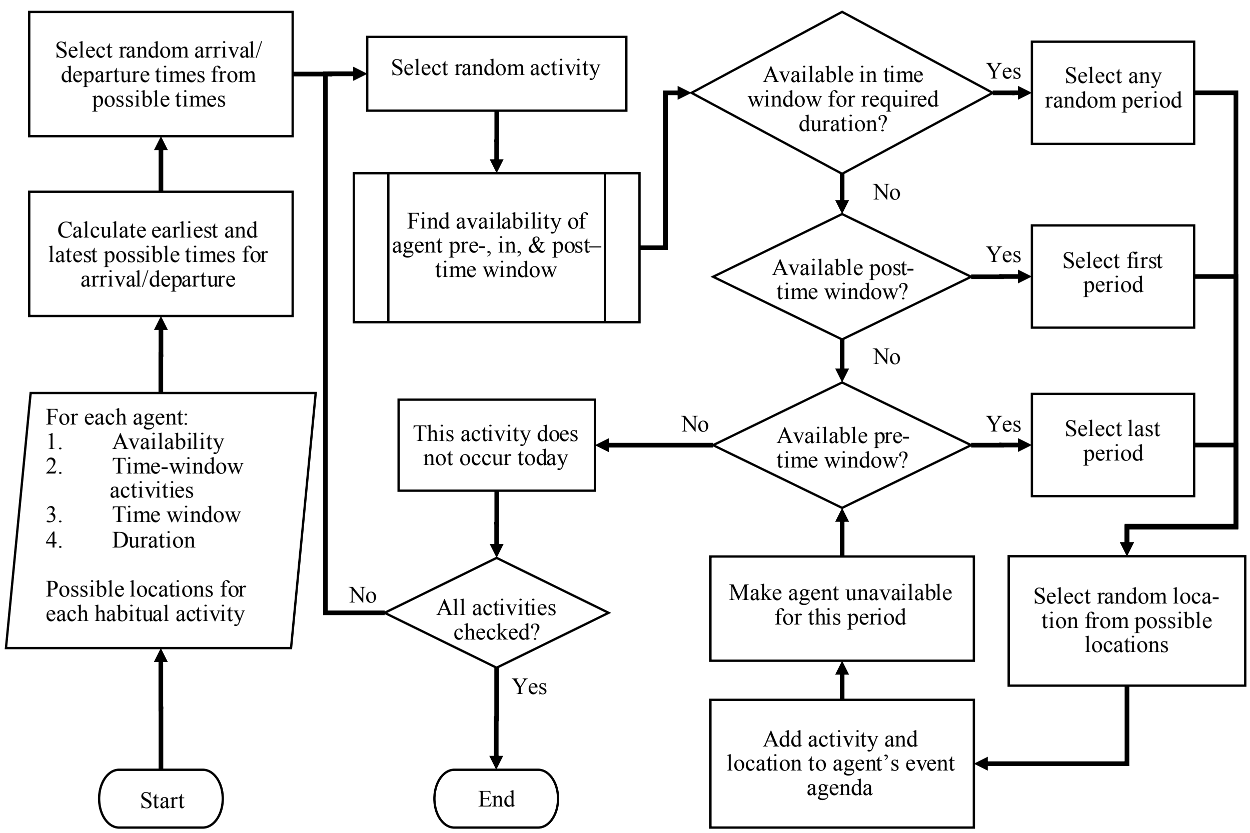

Figure 31: Flow diagram of the event scheduler for time-window activities in a day

Case Study: Event Scheduler

The figure below shows the event schedule for agent 12, a junior engineer, belonging to team 3 which has a team activity scheduled, as can be observed. Events scheduled are the planned meeting and the time-window activities: arrival, lunch, and departure. Between these events gaps available time can be seen with no activities scheduled.

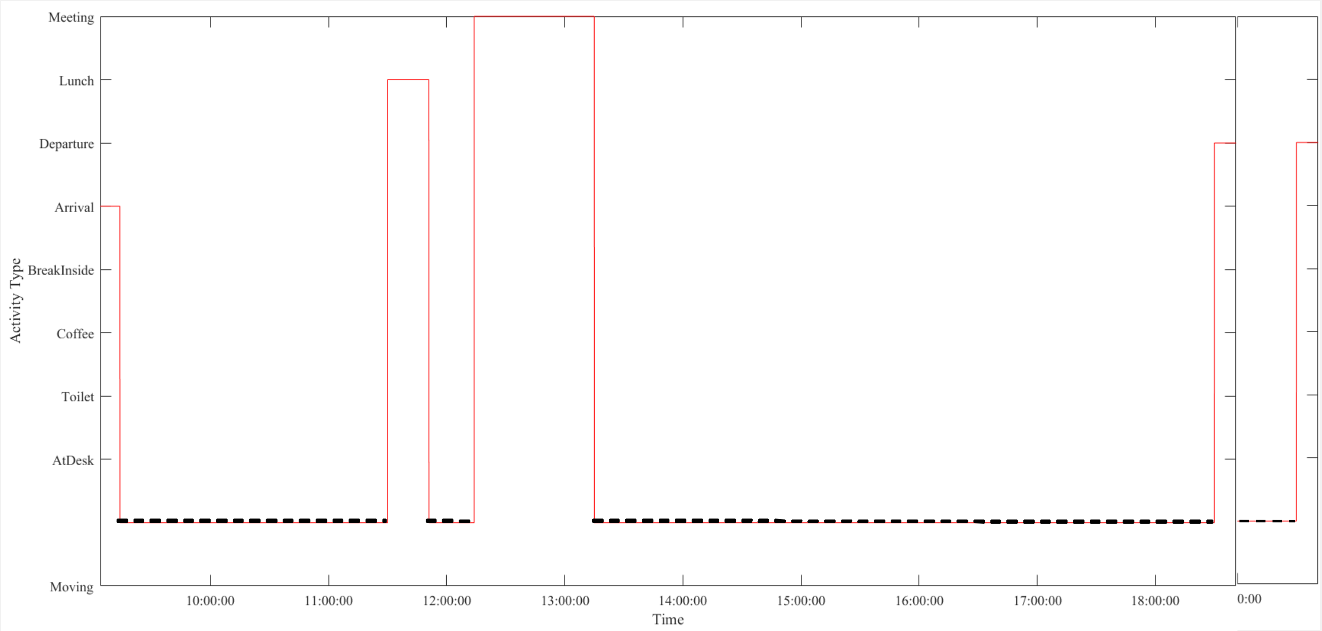

Figure 32: Case study: Event schedule of agent 12. Dashed lines represent gaps of available time.

5.6.2 Movement Scheduler

In the current implementation of the model only the shortest paths are used. For this, the shortest paths and their distances between all locations are calculated and stored. Later, this data is accessed by the movement scheduler to return the movement time which it calculates using the highest desired speed an agent may have. This is done to ensure that during execution of the planned schedule, agents never reach a location before time. If this would happen, agents in the model could, potentially, interrupt ongoing meetings as there are no specific arrangements to represent waiting outside rooms leading to unexpected behaviour. For more complex routing decisions, the movement scheduler would also play the role of choosing a route from several options and storing it so that it can be executed at the operational level. The movement scheduler works together with the Markov chain scheduler to fill the gaps between events in individuals' agendas.

5.6.3 Markov Chain Scheduler

Since the Markov chain scheduler and movement scheduler together form the final planned schedule that the operational level will use for execution it requires a finer time resolution so that the exact movement times are considered. For this, the individual agendas which contain only events are expanded to a resolution of 1 second. The activities to be scheduled by the Markov chain scheduler, individual profiles, and locations associated to different activities are used as input. Assuming independence of schedules for the time-independent activities, the Markov chain scheduler (Figure 33) is run for each individual.

The first step is to generate the agent's transition matrix for which extra constraints in the form of external rules and spontaneous activities have to be defined. Since spontaneous activities take place only during another activity, transition probabilities to these activities should be zeros from all activities except the inducing ones. External rules allow users to embed known relationships between activities into the model. These rules are added as extra linear constraints as described in section 4.3.4. Using the transition matrix, activities are simulated (Figure 17) for a total time equal to twice the total available time the agent has between scheduled events. This sequence of activities is referred to as the Markov chain agenda. The Markov chain agenda contains a sequence of activity episodes with their durations. For service activities, such as getting coffee from a machine, the duration is fixed to the service time. Since, fixing the duration for service activities may reduce the duration obtained from the simulation itself, the Markov chain is simulated for a longer time than the agent's total available time. Thus, the factor two is arbitrarily chosen to ensure that enough activities have been generated to fill the gaps in agents' agendas. It should be noted that simulating the Markov chain for the total available time is better than running it for each gap individually as a longer simulation allows the input properties to be reflected better in the results than several shorter ones.

The gaps in the agendas are filled one by one. Selecting the first gap, its duration (gap time) as well as the locations the agent is expected to be in, in the beginning and at the end of the gap, are stored. These 'event' locations are already known as the gaps are formed in between events for which the event scheduler has already assigned a location. Next, the walking time between the two event locations is obtained from the movement scheduler and compared against the gap time; if it is lower, it means that there is time to insert another activity in the gap, otherwise the gap is filled with only movement as 'activity'. When there is time for another activity, the first activity episode from the Markov chain agenda is selected. The nearest possible location for this activity from the location at the beginning of the gap is the activity location choice. Once again, the movement scheduler is used to calculate the time to go from the previous location to this activity and then to the event location at the end of the gap. The total movement time (from beginning event location to intermediate Markov chain activity and from that to end event location) plus the duration of the Markov chain activity episode is considered as the fill time. If the fill time is still lower than the gap time, the first activity in the Markov chain agenda is deleted, and the new first activity in the Markov chain agenda is chosen. The fill time now includes movement time from the beginning event to the first chosen Markov chain activity, the duration of both Markov chain activities, and the movement time from the last chosen Markov chain activity to the end event. This process is repeated until the fill time is greater than or equal to the gap time. When this happens the duration of the last added Markov chain activity is reduced so that the fill time equals the gap time. If this causes the duration to be negative then it is reduced to zero in the current implementation. Thus, a minimum duration of activities is not considered. Having reduced the duration of the last Markov chain activity it is removed from the beginning of the Markov chain agenda. The gap in the agent agenda is then filled in by the movements activities from the beginning event to the first added Markov chain activity and from the last activity to the end event with all the activity durations in the middle. The entire process is repeated for the next gap in the agent's agenda until all gaps have been filled up with activities and movements.

After running the above three schedulers, a planned agenda of activities, their locations, and movements between these locations is available at the resolution of 1 second for each agent.

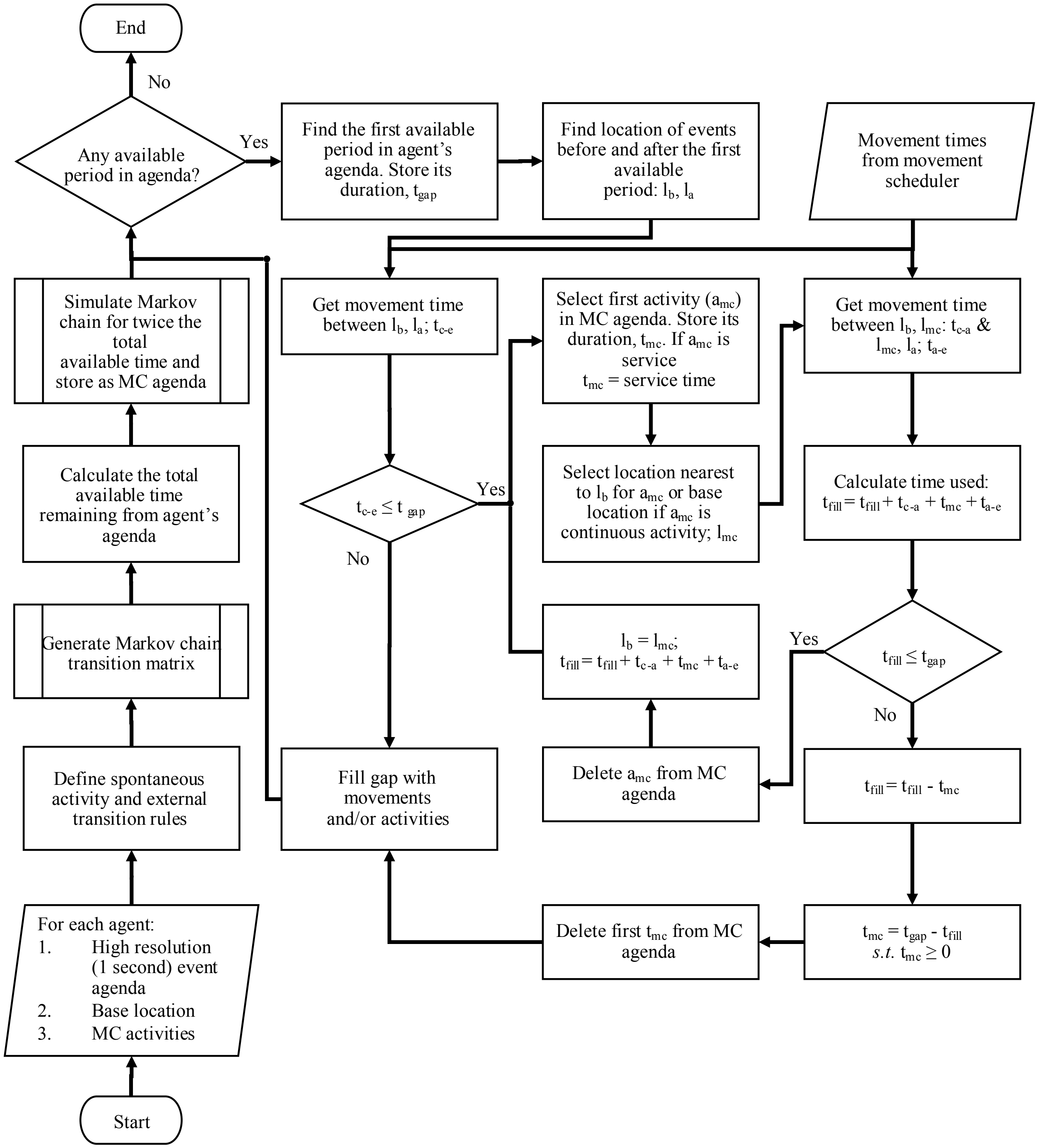

Figure 33: Flow diagram of the Markov chain scheduler

Case Study: Markov Chain Scheduler

While no spontaneous activities occur in the case study an external rule representing the tendency of people to take a break after getting coffee is considered. The rule says that after getting coffee there is a 1/3rd probability that agent's will take a break. It should be noted that the transition probability from coffee to break will be 1/3rd of the sum of transition probabilities from coffee to other activities to exclude self-transitions. The table below shows the transition matrix generated for agent 12, a junior engineer.

Table 10: Case study: Transition matrix for agent 12

	At Desk	Toilet	Coffee	Break
At Desk	0.98667	0.006179	0.006051	0.001104
Toilet	0.14924	0.8	0.041597	0.009162
Coffee	0.25715	0.082032	0.5	0.160818
Break	0.04657	0.026791	0.026637	0.9

Using this transition matrix, the Markov chain agenda is simulated for twice the available time in the event schedule, which as can be seen in Figure 32 is about 8 hours. The figure below shows the Markov chain agenda for about 16 hours which is used to fill in the gaps of the event scheduler.

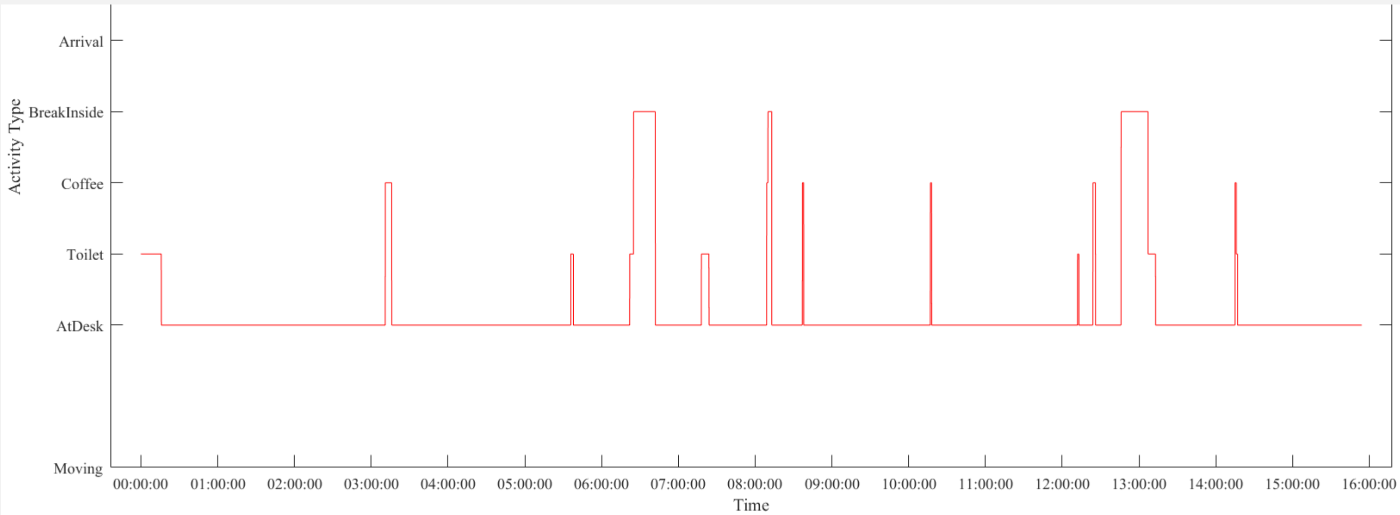

Figure 34: Case study: Markov chain agenda for agent 12 from simulation of transition matrix

After filling the gaps in the event scheduler with the Markov chain agenda and movements the final planned schedule is obtained as shown in Figure 35.

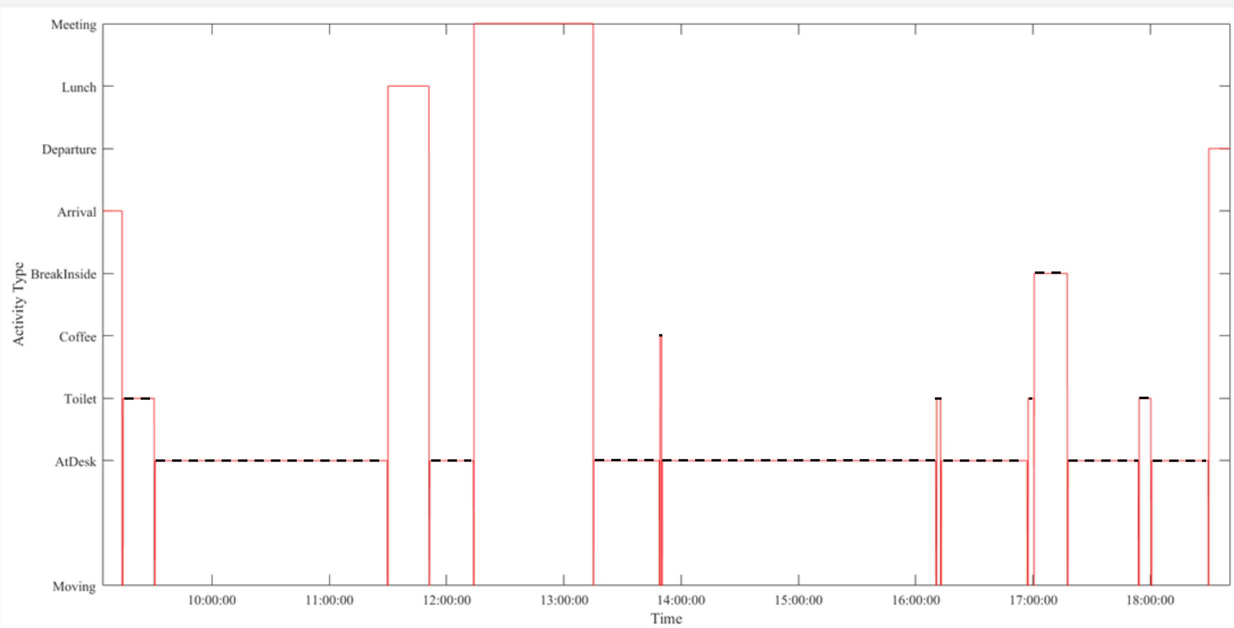

Figure 35: Case study: Full planned schedule for agent 12. Dashed lines show the parts of the final schedule that come from the Markov chain

agenda.

5.6.4 Re-scheduler

As described previously (section 4.3.2), the planned schedule cannot be followed exactly due to accumulation of sub-second errors in movement time, random fluctuations in acceleration, pedestrian traffic, etc. Further, it should be noted that the Markov chain scheduler does not make use of the resource handler to allocate locations based on their occupancy. This reflects the fact that individuals would normally not know the occupancy status of activity locations such as toilets until they actually go there and check. For these reasons a separate scheduling module, the re-scheduler, is required to run in tandem with the operational level model to dynamically update the planned schedule. This scheduler interacts with the resource handler, which during the execution of the operational level, keeps track of location occupancy.

The occupancy of the activity location is checked when the next node in an agent's route is the destination location. Thus, it is assumed that the agent is able to assess the occupancy from the second last node in its route which would usually be true as there has to be a direct line of sight between consecutive points in the route. Thus, at this point, the agent checks its activity location choice once again. When the destination is a service, the time to get the service is calculated from the queue length and service time of the planned location. This is compared against the queue lengths and service times of other possible locations and the time it would take to reach there. As discussed in section 4.5, the assumption that the queue lengths at other locations, that are possibly not visible, are known to the agent, is made to represent that occupants make estimations about queue lengths. If another location is more attractive in terms of time to obtain the service, the location choice is changed and a route is chosen to reach the new location. Similarly, if, at the second last node in the agent's route, it is discovered that the destination non-service location is occupied, the nearest possible alternative is chosen and the agent is re-routed. It is assumed that agents will not change their location choice more than once, so if on reaching an alternative service location, they find that there is an even longer time to obtain service they will stick with the choice. For locations, if the alternative is also occupied, or if there is no alternative then the activity to be performed is skipped (called activity 'rejection') and the agent would return to its base location where it remains until it has to go to the next activity. The location choice for the next activity remains the same as it is in the planned schedule.

Various factors, including on-the-fly location choices and activity rejections, will require the planned schedules to be updated dynamically. To do this, it is assumed that the planned durations for all activity episodes except planned and continuous activities cannot be reduced due to delays in the execution. This means that when the executed schedule is delayed and agents reach late to planned and continuous activities they will still end these activities according to the planned schedule. There are two reasons for doing this. First, planned activities have characteristically fixed start and end times so it would not make sense to shift planned activities due to operational level delays, that is, delays which are not caused due to the extension of another activity but due to delays experienced when moving within the building. Keeping the end times of planned activities fixed would reflect reality better than (i) shifting the start and end times of one participant individually possibly leading to that individual remaining in the meeting location alone for some time after everyone else leaves, or (ii) shifting the planned activity for everyone. In addition, these alternatives would require rescheduling other activities as in the same room, meetings may be scheduled immediately next to one another resulting in a chain of rescheduling that would make the model unnecessarily complex. Second, planned and continuous activities are expected to be, generally, of a reasonably large duration so that the delays can be absorbed by them without having to cancel the activity altogether or very often reduce their duration to an unreasonably small time. By also letting continuous activities absorb delays, the accumulation of delays may be reduced as it can be expected that occupants return to activities such as working at desk quite often during the day.

To update the schedule, whenever an agent reaches an activity location, the re-scheduler compares the time it was supposed to reach with the time it actually did. The start and end times for all activities, until the first absorbing (planned or continuous) activity, after the current time in the planned schedule are shifted forward by this difference. The start time of the absorbing activity is also shifted forward. For activity rejections, the agent is said to have reached the activity location when it returns to a continuous activity which absorbs delays and thus, the activity end time remains unchanged unless it occurs before the agent reaches the base location. In this case, the agent is at the base location only momentarily before leaving for the next activity. On reaching the next activity location the planned times can be changed as discussed above. Apart from updating schedule times, locations and activities are also updated in the planned schedules in case of re-routing or activity rejections.

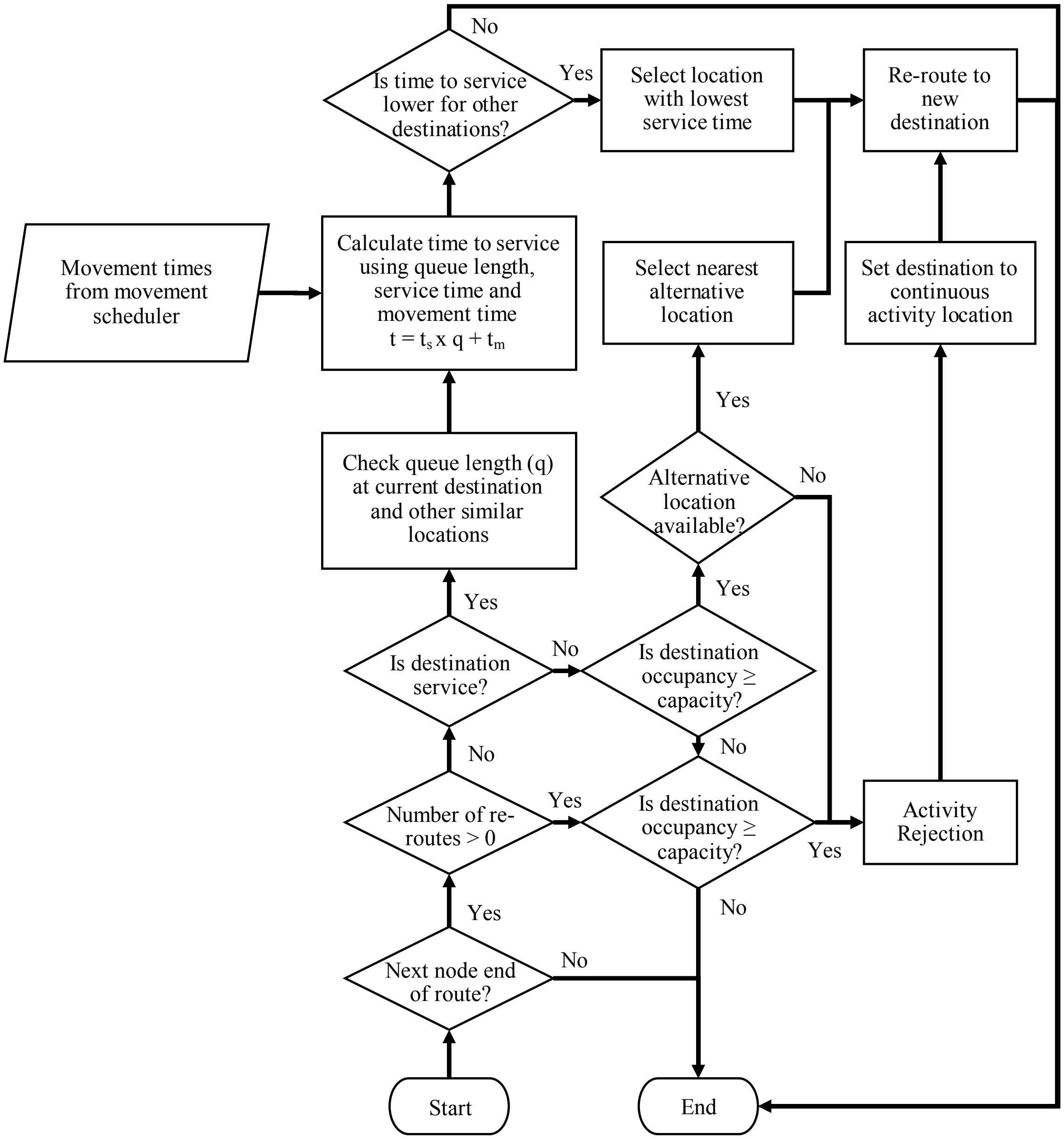

Figure 36: Flow diagram for on-the-fly location and activity choice by the re-scheduler

Case Study: Re-scheduler

Figure 37 (left) shows the result of rescheduling due to operational delay, that is when it takes more time for the movement than planned, and due to activity rejection. The activity ‘toilet’ is shifted to the right without reducing its duration but, although the next activity ‘at desk’ is reached late, it is not shifted (cannot be seen in figure) and the activity’s end time will be the same for planned and executed schedules.

Figure 37 (right) shows a case of activity rejection due to unavailability of location. The activity ‘rejection’ represents that the agent has returned to the continuous activity. Since after returning, the same continuous activity is taken up there is no movement to another activity and hence the figure shows the activity rejection continuing until the next movement to the toilet. This does not mean the continuous activity is rejected but it is an artefact caused by the fact that there is no real cut-off felt by the occupant between being at the desk due to an activity being rejected and being at the desk because it is scheduled to be there. The movement time to the next activity, ‘toilet’, is the same as planned.

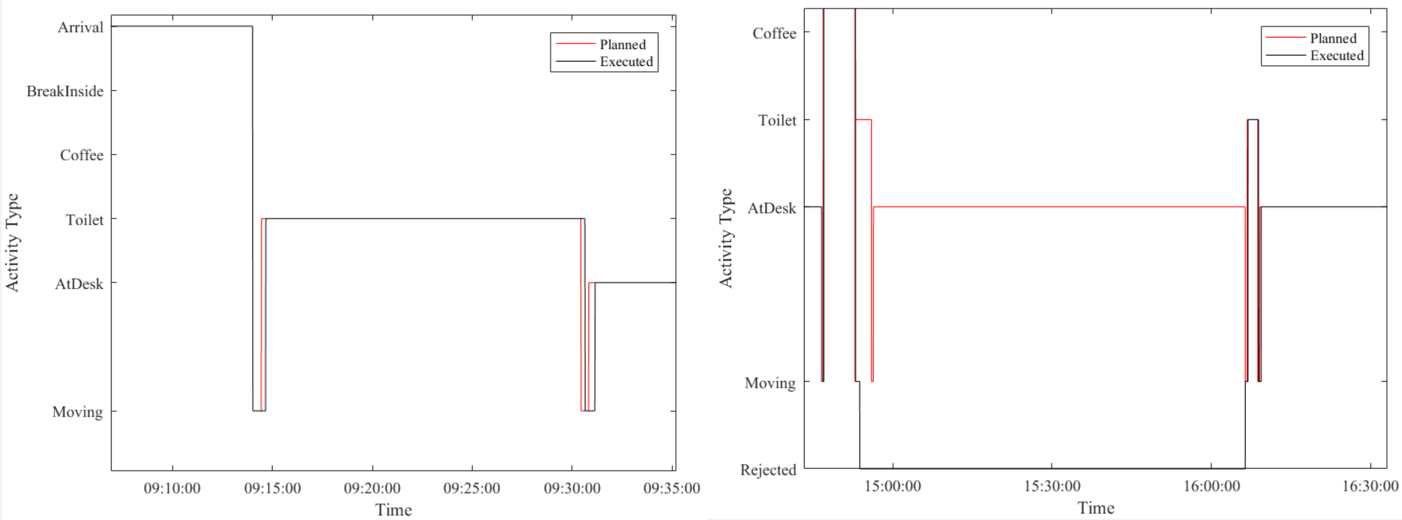

Figure 37: Case study: Rescheduling due to operational delay (agent 12; left) and activity rejection (agent 17; right). Activity rejection continues as after returning to the desk there is no specific movement to the next activity which is also at the desk.

5.7 Model Output

At every time-step of the operational level, the model saves the coordinates and velocity of pedestrians for later use. Although the implementation offers an online visualization of pedestrian trajectories, as one would expect, this results in considerably slow execution time. Therefore, the stored data is used in a separate, offline visualization of pedestrian movements where the animation speed and the start and end times can be controlled. The stored data can be used by other applications such as detecting triggers for sensors in the building modelled. If required activity patterns can also be stored.

5.8 Case Study: Applications

In this section, some results from the case study are presented in order to qualitatively verify and validate the model as well as put forward applications of the model. Four measures are presented: (i) break durations, (ii) coffee away times, (iii) desk occupancy, and (iv) sensor activations. While the first two measures describe the difference between the input and output, the latter give an aggregated sense of the model performance and its ability to produce qualitatively believable outputs. Due to the unavailability of time and appropriate data this is done instead of comparing simulated behaviour with revealed behaviour in the real-world. Except for the last one, all measures are calculated the aggregated output of 10 simulations of the model, that is, 10 days. The running time for each simulation (on a 64-bit Windows 10 PC; Intel Core i5-6300U; 8GB RAM) with the given inputs was observed to be ~120-140s.

Figure 38 shows the distribution of break times (left) and away times for coffee (right) for all agents over 10 simulations of the model. As expected, both distributions show features of exponential distributions and with a greater number of simulations the distribution will become exponential. For these 10 simulations, the mean break duration is found to be 9.3 minutes while the mean away time for coffee is 84.6 minutes. While the former is close, both values underestimate the true inputs of 10 minutes and 120 minutes respectively. A possible reason for lower duration may be the splitting of activities from the Markov chain agenda while filling the gaps in the event schedule. Splitting activities would reduce their average duration. Another possible reason for the underestimated values may simply be that a greater number of simulations are required to approach the mean.

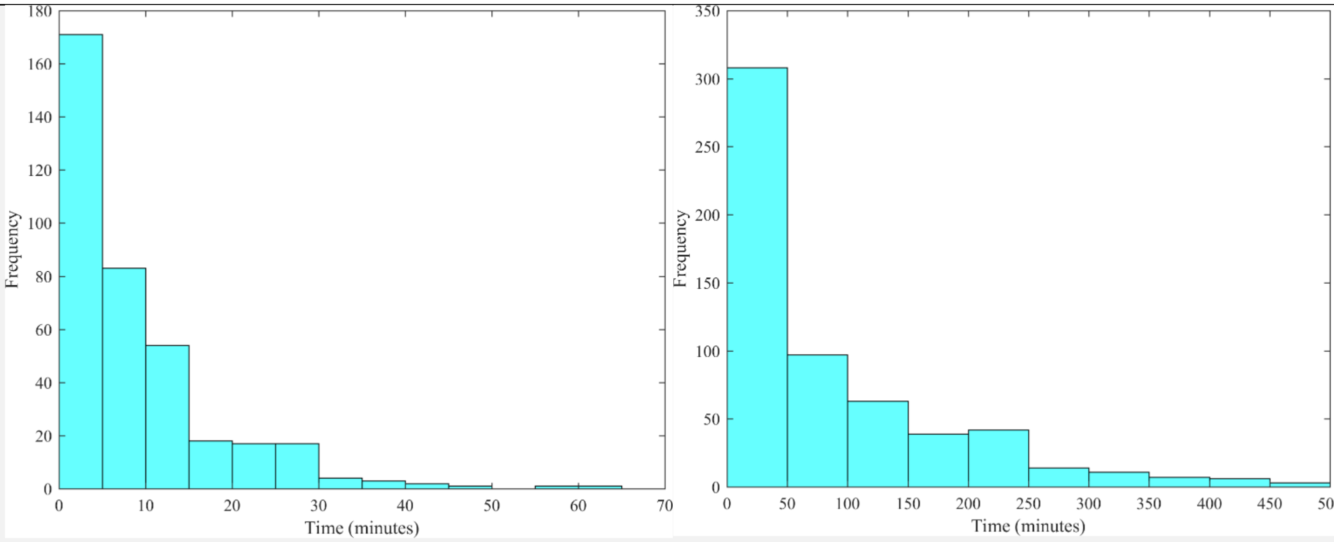

Figure 38: Case study: Distribution of break durations (left) and coffee away times (right)

Desk occupancy describes the number of office occupants who are at their desk at a given moment. It is an important parameter for lighting and HVAC controls in office buildings because these building services are based on the presence of people in certain areas. Figure 39 shows the maximum, minimum, and average desk occupancy in a day over 10 simulations. A steep rise at 09:00 can be seen as occupants arrive to the office while during the lunch hours the number of occupants at their desks clearly drops. Further from 16:30 onwards the occupants begin to reduce as people leave the office for the day.

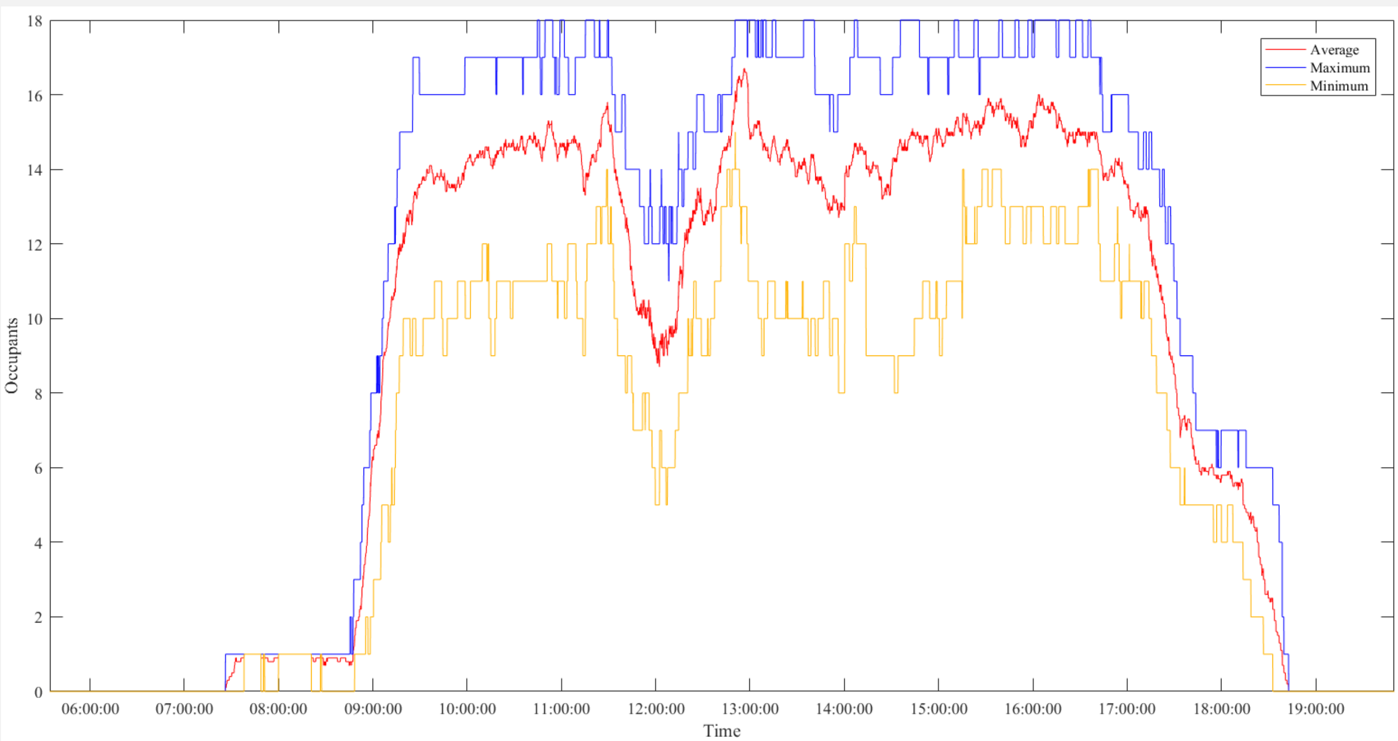

Figure 39: Case study: Desk occupancy over 10 model runs

Figure 40 shows a sensor field setup over the set of turnstiles in the office building. The saved model output can be used to generate activations of any sensor field drawn on the building plan. Here a simple sensor model is considered wherein if a pedestrian is within the field, the sensor is activated. More complex models with probabilities of false positives and negatives can also be used. Figure 41 plots the sensor activations over the period of one day in a single simulation. Clear patterns of arrival, lunch, and departure can be identified in these activations. Since lunch takes place outside the building, occupants cross the turnstiles while going out and coming back as well. From the inputs, it is known that while the receptionist arrives early, the manager arrives and has lunch later than others; this is reflected in the activations. Moreover, while the 09:00 average arrival time is common for 16 occupants (everyone except the manager and receptionist), the departure time is more spread out from 16:30 to 18:30 as can also be observed in the activations.

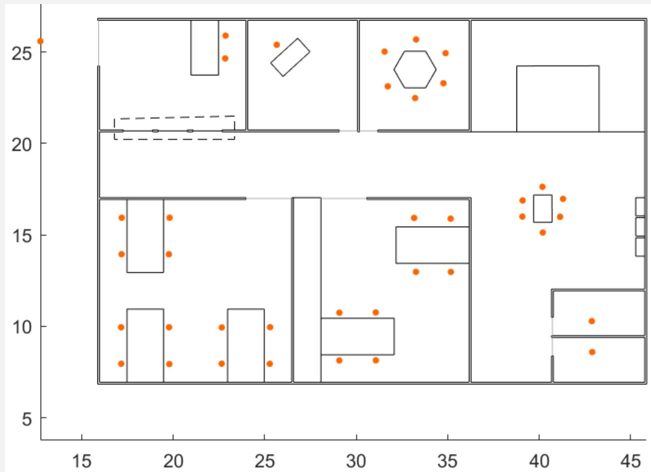

Figure 40: Case study: Sensor field (dashed lines) over turnstiles

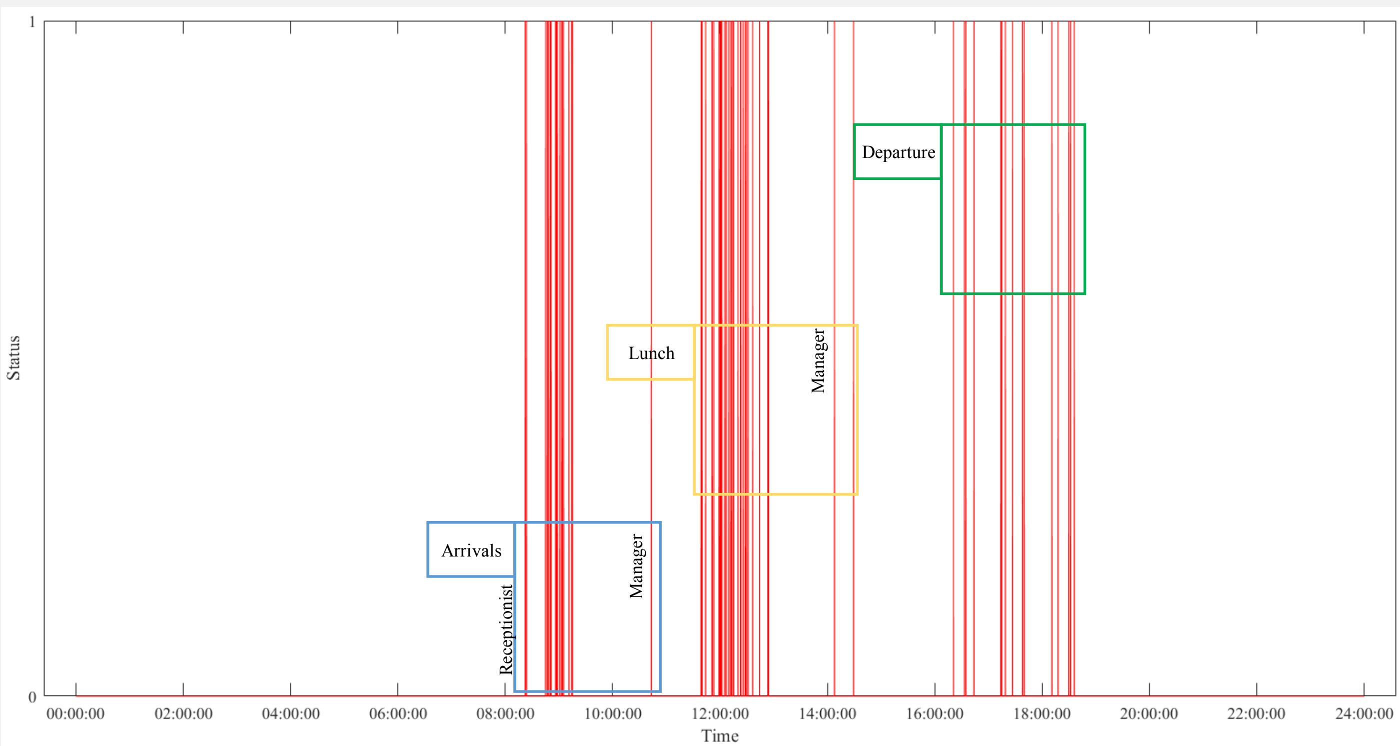

Figure 41: Case Study: Sensor activations and possible reasons for turnstile sensors

5.9 Model Limitations

Understanding the limitations of the model is an important step in its assessment. Model limitations are listed below first for the operational level and then for the strategic and tactical levels. Alongside the limitations, where possible, suggestions for overcoming them are also made.

Operational Level:

First, different behaviours represented in the model are compared against the model requirements. As modelling waiting behaviour is a medium priority requirement, it is modelled using vertical queues to save time. However, this means that high pedestrian density situations that may occur where people queue are not observed. The model can be made more realistic by modelling horizontal queues and then different queue organizations explicitly. For this, algorithms suggested in (Kneidl, 2016) for linear queues or in (Campanella, 2016; Davidich et al., 2013) for queues in an area can be used on top of the current model. Modelling social groups at the operational level is also a medium priority requirement. However it is not included in the model due to time constraints and problem at higher levels of the model in the formation of such groups.

Since the navigation graph generation is to work with the social forces model, an extra step is added that removes connections that lie very close to obstacle corner. Although unlikely, this may result in some locations getting disconnected from the navigation graph making them inaccessible. This can be corrected by running a check at the end of the graph generation to see if the navigation graph remains connected. Further a method should be provided so that additional nodes can be added as required so that the graph can be connected. Also regarding the navigation graph is the calibration of its parameters such as half-cone angle which are currently chosen arbitrarily. Since the navigation points heavily influence where pedestrians walk this is important.

Wormhole transitions between two nodes from either side take place simultaneously in the current implementation, whereas in reality, when two persons want to pass through the same turnstile or exit and enter a lift, one of the persons will yield. To represent this, one possibility is to check who enters the transition first, let that pedestrian transition to the other side first whilst making the other pedestrian wait longer in transition.

Finally, there is no way, in the current implementation, to have pedestrians wait outside busy meeting rooms when they arrive earlier than scheduled. While measures are taken at the tactical level to prevent occupants from reaching meetings early, since this can also happen in reality and cause high densities of waiting pedestrians, it would be better to add waiting points or areas outside such rooms and allow meetings to run late (causing others to wait) or occupants to arrive early.

Strategic + Tactical Level:

While spontaneous activities model activity-induced activities, environment-induced activities are not modelled. Environment-induced activities are those activities that are induced by the current position of the occupant. For example, returning from a meeting, an occupant may find itself in sight of a colleague and decide to have an unplanned interaction. Such activities need to be scheduled instantaneously possibly by the re-scheduler using feedback on position of pedestrian from the operational level. By using information from the resource handler on the availability of other occupants within a particular distance it would also be possible then to model environment induced unplanned interactions which is a topic many space syntax studies have focussed on.

Related to the above point, while planned interpersonal interactions can be simulated, unplanned interactions are not modelled in the current implementation. One of the main reasons why unplanned interactions are difficult to model is that the Markov chain scheduler assumes independence of activity schedules. For example, assume that unplanned interactions are modelled as a spontaneous activity. Say, occupant A has an unplanned interaction scheduled with occupant B. While this may sound oxymoronic, it is not, as the interaction does not exist in B's schedule. Three cases may then occur: (i) A finds B at its seat and is able to complete the interaction before moving to the next activity, (ii) A finds B at its seat but is not able to complete the interaction as B goes off to another activity, and (iii) A does not find B and is unable to even start the interaction. While cases (i) and (iii) are realistic, case (ii) is often not because in reality B would either consider delaying lower priority activities such as getting coffee or A would go along (social group walking). To model either of these real-world behaviours a link to one another's schedules would have to be made so that activities can be rescheduled appropriately.

The assumption of geometric distributions of away times for recurrent activities could not be verified from the single study reporting results in literature. Data should be collected for these activities and analysed to verify or dismiss the assumption.

The Markov chain and event schedulers being completely separated are not able to mix events. That is, it is not possible for agents to, for example, get coffee during a meeting. While it can be reasonably assumed that this does not take place usually, it would be nice to have the option of simulating such behaviour. For this reason among others, the scheduler framework developed here is more suitable to model individuals whose activity agendas are not dominated by meetings. This is because with less gaps between events, the Markov chain is simulated for a smaller amount of time resulting in higher stochasticity, and the simulated activities are also more likely to be split up due to the many events. These factors would lead to the simulated pedestrian not showing expected behaviour over time.

6 Conclusion

6.1 Summary

In this thesis, an integrated pedestrian behaviour model is developed to simulate human movements in office buildings. The model is to be used as a research platform that can be connected to - , and generate movement data for external applications such as the testing of lighting control strategies. In order to generate pedestrian movements in buildings all three levels of pedestrian behaviour, strategic, tactical, and operational, are modelled so that given the office building layout and some organization and individual characteristics, the locations of each occupant throughout the day can be returned. Furthermore, the model is designed so that it is parsimonious in terms of input data requirements, it is flexible in accommodating various situations, and it is extensible so that it can be connected to other studies and applications as and when required in the model's function as a research platform.

At the strategic level, a classification of activities that individuals perform and a structured representation of organizations in the model are defined. The former is created so that the generation and scheduling methodology for each activity type is similar whereas the latter is necessary to generate organization-related activities. The generation of other activities, is carried out at the tactical level which also schedules the activities, selects activity location and routing from origin to destination. Clock-time dependent activities are scheduled by the event scheduler, time-independent activities are generated and scheduled by the Markov chain scheduler, and the movement scheduler is used to plan the time required to move from origin to destination. Further, in order to update planned schedules and location and activity choices on-the-fly at the operational level the re-scheduler is used. Finally, the operational level executes the planned schedule from the tactical level using a social forces model in continuous space contextualised by a network of navigation points.

This chapter concludes the thesis by answering the research questions in section 6.2, discussing the main scientific and practical contributions of this thesis in section 6.3, and finally suggesting next steps of research in section 6.4.

6.2 Answers to Research Questions

Section 1.4 presented several research sub-questions, corresponding to steps in the model development process, to aid in answering the main research question. Furthermore, each sub-question is associated with a number of supporting sub-questions. Here, the four sub-questions are answered first followed by the main research question. While the supporting sub-questions are not answered individually they guide the flow of answers for each sub-question.

1(1) What are the stakeholder requirements for the pedestrian behaviour model so that it can generate data for present and potential external applications?

Functional and non-functional requirements from the model are elicited, both, directly from the stakeholders at Philips Lighting and indirectly from the expected applications. Pedestrian movement data in offices has to be generated for applications ranging from testing the layouts of occupancy sensors in a building to developing building automation strategies using individual tracking. Thus, the output level-of-detail requires individual pedestrian locations which can be aggregated for applications with lower spatial resolution needs. Horizontal movements with collision avoidance are a strict functional requirement for the operational level model. Furthermore, waiting for services such as coffee machines and elevators, walking in social groups, and vertical movement across floors are also required since these behaviours are commonly observed in office buildings. The strategic and tactical levels should simulate individual activities as well as interpersonal interactions such as meetings in order to produce movements representative of those that occur in offices. In addition, the non-functional requirements form three important guides for the model design: (i) low and simple model data requirements, (ii) flexibility to work with a wide range of building plans and organizations, and (iii) extensibility to connect this research platform to other workplace related studies.

1(2) Which features, peripheral to the pedestrian behaviour model are within the scope of this project?

In addition to the core assignment of developing the pedestrian behaviour model, some peripheral components have to be designed to some extent. For inputs, the model should be able to make use of CAD files for the building plan input. For this, development of a methodology to import CAD files is required. For the output, a simple visualization of the model output – occupant positions – should be developed for demonstration purposes as well as to verify the functioning of the operational level model. Additionally, a simple data storage system must be set up so that simulated data can be reused. Extensive GUI for input/output operations and the design of applications are excluded.

2 *Based on the model requirements, which approaches can be brought together to produce a conceptual model for each of the three pedestrian behaviour levels?*

Operational Level: Extensive literature is available for the operational level although most studies focus on situations with large crowds such as in evacuations or movements in public and semi-public spaces rather than the non-emergency, low density situations in office buildings that are the focus of this thesis. However, it is still possible to select suitable methodology from previous studies and therefore a thorough literature review is conducted. Operational level models can be broken down into their model features, methodological approaches, actions modelled, and non-functional features, each of which in turn offer different options. Through comparison of the model requirements against these options suitable model components are selected.

A microscopic model in continuous space is needed to produce smooth and natural-looking trajectories of individual pedestrians. The continuous space representation is contextualised by a network of navigation points. This method of contextualisation is widely used and is chosen over the potential and visual field methods because of its execution efficiency during initialization. The methodology adopted makes use of corner points in obstacles to generate navigation points and connect them to form a navigation graph based on a number of rules. Pedestrians can navigate around obstacles by following a sequence of points in the navigation graph from origin to destination. The methodology is flexible in that it can generate navigation graphs given most building plans and it enables the use of simple graph-based routing algorithms.

The social forces model is adopted as the methodological approach despite its poorer execution efficiency because it has been found suitable to model a wide range of actions even beyond those required here which can be useful if the model is to be further extended. Since it does not require discretizing space several drawbacks of cellular automata models are overcome. Furthermore, several studies calibrate and validate this methodology allowing the use of previously obtained parameter values directly for the current model. The actions modelled are horizontal movement with collision avoidance, the formation of vertical queues for waiting, and a setup for vertical movements and passages through turnstiles. This last setup is a mesoscopic model wherein pedestrians are identified individually but their individual locations are not available. It uses the concept of wormholes where pedestrians disappear into a wormhole node and after some amount of time, which can be fixed or dependent on individual properties and congestion factors, appear at another. This setup permits modelling movements that either do not need to be modelled according to the model requirements or are too difficult to be modelled with the current movement model.

Strategic Level: The strategic level decides which activities office building occupants carry out in order to drive their movements throughout the day. To do this, each individual is assumed to have two personas, organization-related and non-organization related, which are linked to different activities and can be defined separately. For example, attributes for activities not related to organizations can be expected to hold across different organizations and hence will not need to be estimated for each application. To decide which activities are performed first a classification of activities is necessary so that each activity can be assigned to a methodology for its generation and scheduling. Five types of activities, continuous, recurrent, spontaneous, planned, and time-window, are derived so that they are flexible enough to accommodate various activities occurring in office buildings as well as other types of organizations such as schools. The first three types are independent of time and usually not specifically guided by the organization, whereas the last two depend on clock time and can be organization related.

Organization related planned and time-window activities are generated at the strategic level for which it is also necessary to represent the organization in the model in a structured manner. The organization is divided into role and organization units which define individual activities whereas the concept of teams and team activities is used to generate planned meetings between consistent sets of individuals as is usually observed in different organizations. Representing the organization structure within the model also enables extending it to other workplace related studies.

Tactical Level: The tactical level focuses on scheduling activities to be performed throughout the day. For activity location choices and routing decisions simple assumptions of choosing the nearest location and the shortest route are made consistent with what is expected in offices where occupants are likely to have full spatial knowledge about the building. Unlike the operational level, only a handful of studies exist on scheduling pedestrian movements especially in office buildings. Therefore, reviewing a broader range of studies in fields including passenger transportation and building energy performance it is found that three main approaches exist: activity-based, location-based, and random-access models, the last of which do not perform well for programmed spaces such as office buildings. While activity-based models reflect real-world purposes of movements they are usually complex and have large data requirements. On the other hand, location-based approaches are more straightforward but are unintuitive because they do not reflect the purpose behind movements.

Therefore, the activity scheduling approach proposed here combines the advantages of activity and location based approaches by using an activity-based framework with the less complex location-based, Markov chain model proposed by (Wang et al., 2011). While the simple input needs of the Wang model fulfil the data parsimoniousness model requirement, the activity-based framework makes the model more flexible and intuitive. In this model, the time-independent activities are considered to be the states of an ergodic Markov chain and the activities are generated and scheduled by simulating the Markov chain. To derive the related transition matrices for each individual, the inputs only require the percentage of time spent in an activity and the average duration of each activity episode. The modified Wang model or the Markov chain scheduler is supplemented by three other schedulers. A separate scheduler for clock-time dependent activities such as meetings or lunches is used called the event scheduler. It uses information from the strategic level on duration and time or preferred time-window of activities for planned and time-window activities respectively. A movement scheduler is used to plan time for moving between activity locations by considering distances from the navigation graph. Since this is an integrated model, in order to dynamically update the planned schedule during its execution and carry out, on-the-fly, location and activity choices, the re-scheduler is used. Lastly, a resource handler is used that interacts with the above schedulers. The event scheduler uses it to book rooms for meetings and check the availability of the people involved and the re-scheduler uses it to check the occupancy of activity locations.

3 What should be the framework to integrate the conceptual models of all three pedestrian behaviour levels and implement them in a computer language?

For the integration of the different models the interfaces between them have to be designed properly so that they can function together. The three pedestrian behaviour levels interact not only hierarchically from strategic to tactical to operational but also the other way around as a feedback mechanism. In the hierarchical interactions, the upper levels provide input to the lower levels: the strategic level decides the type of activities to be performed and generates team activities; the tactical level schedules (and, for some, generates) these activities, assigns them a location and makes routing decisions; the operational level uses the planned schedules and routing decisions to execute pedestrian movements. Using feedback from the operational level regarding time of arrival and occupancy of activity locations, the re-scheduler at the tactical level update the planned schedule and makes decisions on changes in activity location and activity choice. Feedback to the strategic level does not occur in the current framework.

A conscious effort is made to keep the integrated framework as modular as possible so that, if required by a particular application, different model components can either be replaced or extended. Six modules are identified in the framework: organization, individual, activities, schedulers, functional spaces, and physical spaces. Physical spaces are used by the operational level to model movements from origin to destination locations through a sequence of navigation points obtained from the routing decisions at the tactical level. To create functional spaces, the physical spaces are assigned functions so that they become activity locations for the tactical level. Schedulers schedule different activities which in turn are associated with individuals. The organization is used to assign activities and their attributes to the organization-related persona of individuals. For the implementation of these modules a simplified UML class diagram is presented in Figure 26.

However, as per model requirements, the pedestrian behaviour model is currently modelled in MATLAB in a non-object-oriented manner. The implementation of the different modules is not discussed here to avoid lengthy repetition but the reader is directed to sections 5.4, 5.5, and 5.6. In order to verify the working of the model flow diagrams are prepared for different modules to check the flow of logic. Furthermore, an imaginary office is used as a case study and results from different modules for this are obtained in a step-wise manner to verify the model's functioning.

4(1) Is the pedestrian behaviour model able to produce sufficiently realistic outputs as per the model requirements?

In order to assess model performance, (Naylor and Finger, 1967) propose a three-stage validation procedure. The first step is based on rationalism, that is, choosing model assumptions rationally based on general knowledge, knowledge from similar systems, and other observations. During the model development process, a number of assumptions are made on the basis of expected behaviour in offices and observations in literature:

- At the operational level, several design decisions are based on the assumption that in non-emergency situations in office buildings, pedestrian density remains low. Thus, for example, physical contact is not observed and does not need to be modelled.
- At the tactical level, assumptions of nearest activity location choice and shortest route selection are based on the hypotheses that occupants do not have specific location choices, they have full spatial knowledge of the building, and are purpose-driven as opposed to leisurely loitering.
- For scheduling activities, to justify the memoryless property of the Markov chain scheduler, it is assumed that occupants do not intentionally optimize activity sequencing within tours (starting and ending at their base location) because the effort required to access different locations within an office is quite low and there is little benefit to consistently optimize tours given the activity types and the time spent in the office.
- Another Markovian property is its geometric (exponential for continuous time) distribution of sojourn time at and away times from activities.

The second step, based on empiricism, requires comparing model postulations statistically against empirical observations, wherever possible, in order to verify them. From the above assumptions, the only empirically verifiable assumption is the last one of geometric distributions. In section 4.3.5 it is shown that sojourn times of being at desks in a variety of situations are indeed either distributed exponentially, or have a distribution that is very similar (albeit subjectively) to being exponential.

Lastly, the third step deals with the model's ability to predict behaviour by testing input-output relationships against observed behaviour in the real-world system. In order to conduct this validation, ideally, a real-world study should be undertaken. Apart from the building plan, individual and organization characteristics can be obtained from simple survey of occupants to make up the model input. Next, model results should be obtained and compared against observations from sensors in the same building to check the model's prediction power. The comparison methodology would depend on the type of sensors available. However, due to lack of time and availability of both stated and revealed observations from the same office building, the model is instead judged qualitatively using a case study of an imaginary office. Four measures are presented: duration distribution of breaks, away time distribution of the activity getting coffee, average desk occupancy over a day, and activations of a sensor placed at the entrance of the main part of the building. The first two measures return expected exponential distributions but underestimate the mean values which may be due to the limited number of simulations used. The lower duration value may also be due to the fact that Markov chain activities may be split into two episodes during scheduling. Both average desk occupancy and sensor activations show expected macroscopic patterns of arrivals, lunch breaks, and departures, thus, qualitatively validating the model output.

4(2) What are the limitations of the pedestrian behaviour model?

Limitations in the conceptual designs of the individual behaviour level models and the integrated framework are discussed in section 5.9 along with possible solutions to overcome them.

Having answered the research sub-questions it is now also possible to answer the main research question:

What is an appropriate pedestrian behaviour model of occupants in office buildings that integrates the three pedestrian behaviour levels (Hoogendoorn, 2001) so that it can describe movement patterns of different organizations and generate high spatial and temporal resolution data on individual occupant positions?

The answer to the main research question is the pedestrian behaviour model developed after following the four model development steps which the above research sub-questions addressed. First, to find out what would be an appropriate model, model requirements were elicited from the stakeholders at Philips Lighting. Based on the possible applications for this research platform and direct input from the stakeholders, possible inputs, required outputs, pedestrian behaviours to be represented in the model and other guidelines were derived. In accordance with these requirements and guidelines, conceptual models for the three pedestrian behaviour levels were developed. The operational level simulates movements and consists of a microscopic social forces model in continuous space contextualised by a navigation graph alongside a mesoscopic model that is used for only certain movements. The strategic level models which activities individuals perform, for which a classification of activities is developed and associated with occupants, and the organization structure is represented within the model. The tactical level schedules the activities associated with different individuals and also makes activity location and routing decisions. It consists of a scheduler based on Markov chains supplemented by event and movement schedulers, as well as a re-scheduler that interacts with the operational level to dynamically adjust the schedule. The three levels are brought together under a modular framework and implemented in MATLAB so that given a building plan and the required occupant and organization characteristics, individual occupant positions can be generated at a resolution of a fraction of a second. To test the appropriateness of the model, the model implementation is verified and the results qualitatively validated through a case study of an imaginary office.

6.3 Main Contributions

This section presents the main scientific and practical contributions of the thesis. From the literature reviews of the different pedestrian behaviour levels it is found that while the operational level has been studied extensively, the strategic and tactical levels have received far less attention. For office buildings specifically, there are only a handful of studies available even after including studies on building occupancy. Furthermore, there are almost no studies that integrate all three levels of pedestrian behaviour under one framework and none, to the best of the author's knowledge, in the context of office buildings. Only one work, (Tabak, 2008), comprehensively covers strategic and tactical level behaviour in office buildings, but even then, the framework developed in it does not integrate microscopic operational level behaviour and the activity-based model developed in this work has complex input requirements which are not easy to obtain.

Therefore, this thesis makes two scientific contributions. The first one is the integration of all the pedestrian behaviour models into a single framework which can generate pedestrian locations throughout a day in office buildings using as input the building plan and individual and organization characteristics. Secondly, a Markov chain based model is designed to make the generation and scheduling of activities much more straightforward whilst reducing the data requirements. This scheduler uses the methodology proposed in (Wang et al., 2011) modifying it and supplementing it with other schedulers to overcome its limitations. Moreover, the approach is thoroughly analysed for valid input range, solution range, and simulation results, something that is missing in previous studies using this approach.

Even though an increasing number of people are working in offices and detection techniques in buildings have improved greatly the unavailability of data to researchers and designers impedes the development and testing of building service controls. The practical contribution of this thesis is towards the solution of this problem through the development of a simulation model, the pedestrian behaviour model, that can generate detailed occupant movement data in offices. The model does not address a specific issue but rather becomes a research platform to which different applications can be connected. Furthermore, its data requirements are few and simple, it is flexible in terms of the building layouts and organizations it can be applied to, and lastly, it can be easily extended due to its modularity and the structured representation of the modelled organization in it.

6.4 Next Steps

In this section, possible future directions to take this research forward in are discussed. Next steps of research from the practical perspective depend on the applications stakeholders have for this research platform but some general recommendations are made here. Firstly, the model needs to be quantitatively validated across different organizations using real-world stated and revealed behaviour patterns as described under question 4(1) of section 6.2. In addition to this, as suggested previously under model limitations (section 5.9), there is need for more empirical verification of assumptions made for the Markov chain model such as the assumption that occupants do not optimize tours or the geometric distribution of away times. Thus, better verification of assumptions and validation of results is important future research work.

Currently stated preference input is required for the schedulers as it is expected that revealed preference data from sensors are usually unavailable. However, when more information is available, additional constraints in the form of known transition probabilities can also be added to make use of it. Methodology for converting information from sensors, such as presence detection from PIR sensors, to transition probability constraints is missing. In order to be able to make full use of sensor data, when it is available, alongside the pedestrian behaviour model suitable methodology for the estimation of transition probabilities should be developed.

The pedestrian behaviour model takes inputs from the user and produces occupant location data as output which can then be used for external applications. At this moment there is no mechanism to allow feedback from applications to influence model simulations. It would be interesting to add such a feedback mechanisms so that the application, for example a control system being tested, can also affect the behaviour of the occupants. This two way interaction would help study the performance of control systems better.

In addition to the practical recommendations, future paths from the academic perspective are also important. In the field of space syntax much attention is being paid to interactions in offices (Sailer et al., 2016; Sailer and McCulloh, 2012). These studies usually use indicators of spatial configuration, such as integration, to derive relationships with interactions in offices. However, movements in office buildings have been found to be more programmed, that is dependent on functional spaces, than configurational (Sailer, 2007). Using the pedestrian behaviour model, it is possible to obtain relationships between interactions and the functions of various spaces when environment-induced activities of unplanned interactions can be simulated. For this two things are required, first the hard independence between individual schedules must be removed in order to be able to model unplanned interactions. Secondly, the re-scheduler should be further developed so that environment induced activities can be modelled.

The second possible direction of further research is to improve evacuation time estimations in office buildings. Evacuation times in buildings are affected by the spatial distribution of its occupants. Since the pedestrian behaviour model simulates locations of occupants throughout the day, it can generate this information. If the operational level model is properly calibrated and more features are added to it, the model may be used to evaluate evacuation performances at different times of the day more accurately.

7 References

Abdelghany, A., Abdelghany, K., Mahmassani, H., Al-Zahrani, A., 2012. Dynamic Simulation Assignment Model for Pedestrian Movements in Crowded Networks. *Transportation Research Record: Journal of the Transportation Research Board* 2316, 95-105.

Adler, T., Ben-Akiva, M., 1979. A theoretical and empirical model of trip chaining behavior. *Transportation Research Part B: Methodological* 13, 243-257.

Ahn, K.-U., Kim, D.-W., Park, C.-S., de Wilde, P., 2017. Predictability of occupant presence and performance gap in building energy simulation. *Applied Energy* 208, 1639-1652.

Al Horr, Y., Arif, M., Kaushik, A., Mazroei, A., Katafygiotou, M., Elsarrag, E., 2016. Occupant productivity and office indoor environment quality: A review of the literature. *Building and Environment* 105, 369-389.

Ali, W., Moulin, B., 2006. How Artificial Intelligent Agents Do Shopping in a Virtual Mall: A 'Believable' and 'Usable' Multiagent-Based Simulation of Customers' Shopping Behavior in a Mall BT - Advances in Artificial Intelligence: 19th Conference of the Canadian Society for Comput, in: Lamontagne, L., Marchand, M. (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 73-85.

Andresen, E., Haensel, D., Chraibi, M., Seyfried, A., 2016. Wayfinding and Cognitive Maps for Pedestrian Models, in: Knoop, V.L., Daamen, W. (Eds.), *Traffic and Granular Flow '15*. Springer International Publishing, Cham, pp. 249-256.

Arentze, T., Timmermans, H., 2000. *Albatross: a learning based transportation oriented simulation system*. Eirass Eindhoven.

Arentze, T.A., Timmermans, H.J.P., 2009. A need-based model of multi-day, multi-person activity generation. *Transportation Research Part B: Methodological* 43, 251-265.

Asano, M., Sumalee, A., Kuwahara, M., Tanaka, S., 2007. Dynamic Cell Transmission-Based Pedestrian Model with Multidirectional Flows and Strategic Route Choices. *Transportation Research Record: Journal of the Transportation Research Board* 2039, 42-49.

Aurum, A., Wohlin, C., 2005. Requirements Engineering: Setting the Context, in: Aurum, A., Wohlin, C. (Eds.), *Engineering and Managing Software Requirements*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1-15.

Bandini, S., Manzoni, S., Vizzari, G., 2005. Crowd Modeling and Simulation. *Recent Advances in Design and Decision Support Systems in Architecture and Urban Planning*, 161-175.

Bergs, J., 2002. Effect of healthy workplaces on well-being and productivity of office workers, *Proceedings of International Plants for People Symposium*.

Biedermann, D.H., Borrmann, A., 2016. A generic and hybrid approach for pedestrian dynamics to couple cellular automata with network flow models, *8th International Conference on Pedestrian and Evacuation Dynamics*, Hefei, China.

Bierlaire, M., Antonini, G., Weber, M., 2003. Behavioral dynamics for pedestrians, in: Axhausen, K. (Ed.), *Moving through nets: the physical and social dimensions of travel*. Elsevier.

Blue, V., Adler, J., 1999. Cellular Automata Microsimulation of Bidirectional Pedestrian Flows. *Transportation Research Record: Journal of the Transportation Research Board* 1678, 135-141.

Borgers, A., Timmermans, H., 1986. A Model of Pedestrian Route Choice and Demand for Retail Facilities within Inner-City Shopping Areas. *Geographical Analysis* 18, 115-128.

Bosina, E., Meeder, M., Büchel, B., Weidmann, U., 2016. Avoiding Walls: What Distance Do Pedestrians Keep from Walls and Obstacles?, in: Knoop, V.L., Daamen, W. (Eds.), *Traffic and Granular Flow '15*. Springer International Publishing, Cham, pp. 19-26.

Bowman, J.L., Ben-Akiva, M.E., 2001. Activity-based disaggregate travel demand model system with activity schedules. *Transportation Research Part A: Policy and Practice* 35, 1-28.

Burstedde, C., Klauck, K., Schadschneider, A., Zittartz, J., 2001. Simulation of pedestrian dynamics using a two-dimensional cellular automaton. *Physica A: Statistical Mechanics and its Applications* 295, 507-525.

Campanella, M., 2016. Microscopic modelling of walking behaviour, *Transport and Planning*. Delft University of Technology, Delft.

Chen, Y., Hong, T., Luo, X., 2017. An agent-based stochastic Occupancy Simulator. *Building Simulation*, 1-13.

Daamen, W., 2004. Modelling Passenger Flows in Public Transport Facilities, *Transport & Planning*. Delft University of Technology, Delft.

Danalet, A., Farooq, B., Bierlaire, M., 2013. Towards an activity-based model for pedestrian facilities, *12th Swiss Transport Research Conference (STRC)*, pp. 1-33.

Davidich, M., Geiss, F., Mayer, H.G., Pfaffinger, A., Royer, C., 2013. Waiting zones for realistic modelling of pedestrian dynamics: A case study using two major German railway stations as examples. *Transportation Research Part C: Emerging Technologies* 37, 210-222.

de Bakker, C., Aries, M., Kort, H., Rosemann, A., 2017. Occupancy-based lighting control in open-plan office spaces: A state-of-the-art review. *Building and Environment* 112, 308-321.

Dijkstra, E.W., 1959. A note on two problems in connexion with graphs. *Numerische mathematik* 1, 269-271.

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

Feng, X., Yan, D., Hong, T., 2015. Simulation of occupancy in buildings. *Energy and Buildings* 87, 348-359.

Gallager, R.G., 2011. Finite-State Markov Chains, *Discrete Stochastic Processes*. MIT OpenCourseWare.

Geraerts, R., Overmars, M.H., 2004. A Comparative Study of Probabilistic Roadmap Planners, in: Boissonnat, J.-D., Burdick, J., Goldberg, K., Hutchinson, S. (Eds.), *Algorithmic Foundations of Robotics V*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 43-57.

Gil, J., Tobari, E., Lemlij, M., Rose, A., Penn, A., 2009. The Differentiating Behaviour of Shoppers, in: Koch, D., Marcus, L., Steen, J. (Eds.), *7th International Space Syntax Symposium*. KTH, Stockholm.

Gloor, C., Stucki, P., Nagel, K., 2004. Hybrid techniques for pedestrian simulations, *International Conference on Cellular Automata*. Springer, pp. 581-590.

Goldstein, R., Tessier, A., Khan, A., 2010. Customizing the behavior of interacting occupants using personas, *Proceedings of the National IBPSA-USA Conference, New York, USA*.

Goldstein, R., Tessier, A., Khan, A., 2011. Space layout in occupant behavior simulation, *Proceedings of Building Simulation 2011: 12th Conference of International Building Performance Simulation Association*, pp. 1074-1080.

Gwynne, S., Galea, E.R., Owen, M., Lawrence, P.J., Filippidis, L., 1999. A review of the methodologies used in the computer simulation of evacuation from the built environment. *Building and Environment* 34, 741-749.

Habib, K.M.N., Miller, E.J., 2008. Modelling daily activity program generation considering within-day and day-to-day dynamics in activity-travel behaviour. *Transportation* 35, 467.

Hägerstraand, T., 1970. What about people in regional science? *Papers in regional science* 24, 7-24.

Helbing, D., Farkas, I., Vicsek, T., 2000. Simulating Dynamical Features of Escape Panic. *arXiv preprint cond-mat/0009448*.

Helbing, D., Johansson, A., 2011. Pedestrian, Crowd and Evacuation Dynamics, in: Meyers, R.A. (Ed.), *Extreme Environmental Events: Complexity in Forecasting and Early Warning*. Springer New York, New York, NY, pp. 697-716.

Helbing, D., Johansson, A., Al-Abideen, H.Z., 2007. Dynamics of crowd disasters: An empirical study. *Physical Review E* 75, 046109.

Helbing, D., Molnar, P., 1998. Self-organization phenomena in pedestrian crowds. *arXiv preprint cond-mat/9806152*.

Helbing, D., Molnár, P., 1995. Social force model for pedestrian dynamics. *Physical Review E* 51, 4282-4286.

Hillier, B., 2007. *Space is the machine: a configurational theory of architecture*. Space Syntax.

Hölscher, C., Brösamle, M., 2007. Capturing indoor wayfinding strategies and differences in spatial knowledge with space syntax, *6th International Space Syntax Symposium*.

Hoogendoorn, S.P., 2001. Normative Pedestrian Flow Behavior: Theory and Applications. Transport and Planning, Delft University of Technology.

Hoogendoorn, S.P., Bovy, P.H.L., 2004a. Dynamic user-optimal assignment in continuous time and space. *Transportation Research Part B: Methodological* 38, 571-592.

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

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

IEEE, 2011. ISO/IEC/IEEE 29148:2011 Systems and software engineering - Life cycle processes - Requirements engineering.

Ijaz, K., Sohail, S., Hashish, S., 2015. A Survey of Latest Approaches for Crowd Simulation and Modeling Using Hybrid Techniques, *Proceedings of the 2015 17th UKSIM-AMSS International Conference on Modelling and Simulation*. IEEE Computer Society, Washington, DC, USA, pp. 111-116.

Karamouzas, I., Geraerts, R., Overmars, M., 2009. Indicative Routes for Path Planning and Crowd Simulation, *Proceedings of the 4th International Conference on Foundations of Digital Games*. ACM, New York, NY, USA, pp. 113-120.

Kitamura, R., 1983. Sequential, history-dependent approach to trip-chaining behavior. *Transportation Research Record*.

Kitazawa, K., Batty, M., 2004. Pedestrian behaviour modelling. *An Application to Retail Movements using a Genetic Algorithm. Centre for Advanced Spatial Analysis, University College London*.

Kłeczek, P., Wąs, J., 2014. Simulation of Pedestrians Behavior in a Shopping Mall BT - Cellular Automata: 11th International Conference on Cellular Automata for Research and Industry, ACRI 2014, Krakow, Poland, September 22-25, 2014. Proceedings, in: Wąs, J., Sirakoulis, G.C., Bandini, S. (Eds.). Springer International Publishing, Cham, pp. 650-659.

Klepeis, N.E., Tsang, A.M., Behar, J.V., 1996. Analysis of the National Human Activity Pattern Survey (NHAPS) respondents from a standpoint of exposure assessment. *Washington, DC*.

Kneidl, A., 2016. How Do People Queue? A Study of Different Queuing Models, in: Knoop, V.L., Daamen, W. (Eds.), *Traffic and Granular Flow '15*. Springer International Publishing, Cham, pp. 201-208.

Kneidl, A., Borrmann, A., Hartmann, D., 2012. Generation and use of sparse navigation graphs for microscopic pedestrian simulation models. *Advanced Engineering Informatics* 26, 669-680.

Kolodziej, J., Khan, S.U., Wang, L., Min-Allah, N., Madani, S.A., Ghani, N., Li, H., 2011. An application of markov jump process model for activity-based indoor mobility prediction in wireless networks, *Frontiers of Information Technology (FIT), 2011*. IEEE, pp. 51-56.

Koster, G., Treml, F., Godel, M., 2013. Avoiding numerical pitfalls in social force models. *Physical review. E, Statistical, nonlinear, and soft matter physics* 87, 063305.

Köster, G., Zönnchen, B., 2016. A Queuing Model Based on Social Attitudes, in: Knoop, V.L., Daamen, W. (Eds.), *Traffic and Granular Flow '15*. Springer International Publishing, Cham, pp. 193-200.

Kuligowski, E., 2004. Review of 28 egress models. *NIST SP* 1032.

Lakoba, T.I., Kaup, D.J., Finkelstein, N.M., 2005. Modifications of the Helbing-Molnár-Farkas-Vicsek Social Force Model for Pedestrian Evolution. *SIMULATION* 81, 339-352.

Liao, C., Lin, Y., Barooah, P., 2011. Agent-based and graphical modelling of building occupancy. *Journal of Building Performance Simulation* 5, 5-25.

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

Luo, X., Lam, K.P., Chen, Y., Hong, T., 2017. Performance evaluation of an agent-based occupancy simulation model. *Building and Environment* 115, 42-53.

Martin, S., Aurum, A., Jeffery, R., Paech, B., 2002. Requirements engineering process models in practice, *7th Australian workshop on requirements engineering. Deakin University, Melbourne, Australia*, pp. 41-47.

Moussaïd, M., Helbing, D., Theraulaz, G., 2011. How simple rules determine pedestrian behavior and crowd disasters. *Proceedings of the National Academy of Sciences* 108, 6884-6888.

Moussaïd, M., Perozo, N., Garnier, S., Helbing, D., Theraulaz, G., 2010. The Walking Behaviour of Pedestrian Social Groups and Its Impact on Crowd Dynamics. *PLOS ONE* 5, e10047.

Nasir, M., Lim, C.P., Nahavandi, S., Creighton, D., 2014. Prediction of pedestrians routes within a built environment in normal conditions. *Expert Systems with Applications* 41, 4975-4988.

Nassar, K., Elnahas, M., 2007. Occupant Dynamics: Towards a New Design Performance Measure. *Architectural Science Review* 50, 100-105.

Naylor, T.H., Finger, J.M., 1967. Verification of computer simulation models. *Management Science* 14, B-92-B-101.

Okazaki, S., Matsushita, S., 1993. A study of simulation model for pedestrian movement with evacuation and queuing, *International Conference on Engineering for Crowd Safety*.

Page, J., Robinson, D., Morel, N., Scartezzini, J.L., 2008. A generalised stochastic model for the simulation of occupant presence. *Energy and Buildings* 40, 83-98.

Penn, A., Turner, A., 2001. Space syntax based agent simulation, *1st International Conference on Pedestrian and Evacuation Dynamics*, University of Duisburg, Germany.

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

Russell, S., Norvig, P., 1995. Intelligent agents. *Artificial intelligence: A modern approach*, 31-52.

Sailer, K., 2007. Movement in workplace environments – configurational or programmed?, *6th International Space Syntax Symposium*.

Sailer, K., Koutsolampros, P., Zaltz Austwick, M., Varoudis, T., Hudson-Smith, A., 2016. Measuring Interaction in Workplaces, in: Dalton, N.S., Schnädelbach, H., Wiberg, M., Varoudis, T. (Eds.), *Architecture and Interaction: Human Computer Interaction in Space and Place*. Springer International Publishing, Cham, pp. 137-161.

Sailer, K., McCulloh, I., 2012. Social networks and spatial configuration—How office layouts drive social interaction. *Social Networks* 34, 47-58.

Schadschneider, A., Klüpfel, H., Kretz, T., Rogsch, C., Seyfried, A., 2009. Fundamentals of pedestrian and evacuation dynamics, *Multi-Agent Systems for Traffic and Transportation Engineering*. IGI Global, pp. 124-154.

Seitz, M., Köster, G., Pfaffinger, A., 2014. Pedestrian Group Behavior in a Cellular Automaton, in: Weidmann, U., Kirsch, U., Schreckenberg, M. (Eds.), *Pedestrian and Evacuation Dynamics 2012*. Springer International Publishing, Cham, pp. 807-814.

Seitz, M.J., Köster, G., 2012. Natural discretization of pedestrian movement in continuous space. *Physical Review E* 86, 46108.

Serfozo, R., 2009. *Basics of applied stochastic processes*. Springer Science & Business Media.

Song, W., Xu, X., Wang, B.-H., Ni, S., 2006. Simulation of evacuation processes using a multi-grid model for pedestrian dynamics. *Physica A: Statistical Mechanics and its Applications* 363, 492-500.

Sparnaaij, M., 2015. Social Forces Model: Predictive vs. Reactive: A comparative research, *Transport & Planning*. Delft University of Technology.

Sud, A., Andersen, E., Curtis, S., Lin, M., Manocha, D., 2008. Real-time Path Planning for Virtual Agents in Dynamic Environments, *ACM SIGGRAPH 2008 Classes*. ACM, New York, NY, USA, pp. 55:51--55:59.

Tabak, V., 2008. User simulation of space utilisation. Eindhoven University Press.

Thill, J.-C., Thomas, I., 1987. Toward Conceptualizing Trip-Chaining Behavior: A Review. *Geographical Analysis* 19, 1-17.

Timmermans, H.J.P., Zhang, J., 2009. Modeling household activity travel behavior: Examples of state of the art modeling approaches and research agenda. *Transportation Research Part B: Methodological* 43, 187-190.

Turner, A., Penn, A., 2002. Encoding Natural Movement as an Agent-Based System: An Investigation into Human Pedestrian Behaviour in the Built Environment. *Environment and Planning B: Planning and Design* 29, 473-490.

U.S. E.I.A., 2012. Commercial Buildings Energy Consumption Survey.

U.S. E.I.A., 2017a. Energy Consumption by Sector.

U.S. E.I.A., 2017b. Energy Use in Commercial Buildings.

Versluis, D., 2010. Microscopic interaction behaviour between individual pedestrians, *Transport and Planning*. Delft University of Technology, Delft.

Wang, C., Yan, D., Jiang, Y., 2011. A novel approach for building occupancy simulation. *Building Simulation* 4, 149-167.

Wang, D., Federspiel, C.C., Rubinstein, F., 2005. Modeling occupancy in single person offices. *Energy and Buildings* 37, 121-126.

Wei-Guo, S., Yan-Fei, Y., Bing-Hong, W., Wei-Cheng, F., 2006. Evacuation behaviors at exit in CA model with force essentials: A comparison with social force model. *Physica A: Statistical Mechanics and its Applications* 371, 658-666.

Xia, J., Zeephongsekul, P., Arrowsmith, C., 2009. Modelling spatio-temporal movement of tourists using finite Markov chains. *Mathematics and Computers in Simulation* 79, 1544-1553.

Xia, J., Zeephongsekul, P., Packer, D., 2011. Spatial and temporal modelling of tourist movements using Semi-Markov processes. *Tourism Management* 32, 844-851.

Xiong, M., Lees, M., Cai, W., Zhou, S., Low, M.Y.H., 2010. Hybrid modelling of crowd simulation. *Procedia Computer Science* 1, 57-65.

Zheng, X., Zhong, T., Liu, M., 2009. Modeling crowd evacuation of a building based on seven methodological approaches. *Building and Environment* 44, 437-445.

Zhou, S., Chen, D., Cai, W., Luo, L., Low, M.Y.H., Tian, F., Tay, V.S.-H., Ong, D.W.S., Hamilton, B.D., 2010. Crowd Modeling and Simulation Technologies. *ACM Trans. Model. Comput. Simul.* 20, 20:21--20:35.

Zimmermann, G., 2007. Modeling and Simulation of Individual User Behavior for Building Performance Predictions, *Proceedings of the 2007 Summer Computer Simulation Conference*. Society for Computer Simulation International, San Diego, California, pp. 913-920.

Zowghi, D., Coulin, C., 2005. Requirements Elicitation: A Survey of Techniques, Approaches, and Tools, in: Aurum, A., Wohlin, C. (Eds.), *Engineering and Managing Software Requirements*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 19-46.

A Appendix: Deriving System of Linear Equations for Transition Matrix Elements

This appendix shows how the generalized system of linear equations (equation 22) to be used for the constrained linear least-squares problem, which is solved to obtain transition matrix, elements is derived. Equation 22 is shown again below for the purpose of reading flow:

$$A = \left[\begin{array}{ccccccccccccccccccccc}
0 & \cdots & \cdots & \cdots & 0 & \pi_2 & 0 & \cdots & \cdots & 0 & \pi_3 & 0 & \cdots & \cdots & 0 & \cdots & \pi_n & 0 & \cdots & \cdots & 0 \\
\pi_1 & 0 & \cdots & \cdots & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 & \pi_3 & \cdots & \cdots & 0 & \cdots & 0 & \pi_n & \cdots & \cdots & 0 \\
0 & \pi_1 & 0 & \cdots & 0 & 0 & \pi_2 & 0 & \cdots & 0 & 0 & \cdots & \cdots & \cdots & 0 & \cdots & 0 & 0 & \pi_n & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & \cdots & \cdots & \cdots & \pi_1 & 0 & \cdots & \cdots & 0 & \pi_2 & 0 & \cdots & \cdots & 0 & \pi_3 & \cdots & 0 & \cdots & \cdots & \cdots & 0 \\
 & & & & & & & & & & & & & & & & & & & & \\
1 & 1 & 1 & 1 & 1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & \cdots & \cdots & \cdots & 0 & 1 & 1 & 1 & 1 & 1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 1 & 1 & 1 & 1 & 1 & \cdots & 0 & \cdots & \cdots & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 1 & 1 & 1 & 1 & 1
\end{array}\right]_{2n\times n(n-1)},$$

$$b = \begin{bmatrix} (1-p_{11})\pi_1 \\ (1-p_{22})\pi_2 \\ \vdots \\ (1-p_{nn})\pi_n \\ \\ (1-p_{11}) \\ (1-p_{22}) \\ \vdots \\ (1-p_{nn}) \end{bmatrix}_{2n\times 1}, x = \begin{bmatrix} p_{12} \\ p_{13} \\ \vdots \\ p_{1n} \\ p_{21} \\ p_{23} \\ \vdots \\ p_{2n} \\ \vdots \\ p_{n1} \\ p_{n2} \\ \vdots \\ p_{nn-1} \end{bmatrix}_{n(n-1)\times 1}$$

The derivation will use a 4-state Markov chain as an example which can be easily generalized for *n*-states. From equation 16 it is known that:

$$\pi = \pi \cdot P$$

Writing this in full for 4 states:

$$[\pi_1 \quad \pi_2 \quad \pi_3 \quad \pi_4] = [\pi_1 \quad \pi_2 \quad \pi_3 \quad \pi_4]\begin{bmatrix} p_{11} & p_{12} & \cdots & \cdots \\ p_{21} & p_{22} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ p_{41} & \cdots & \cdots & p_{44} \end{bmatrix}$$

Multiplying the right hand side:

$$[\pi_1 \quad \pi_2 \quad \pi_3 \quad \pi_4] = [p_{11}\pi_1 + p_{21}\pi_2 + p_{31}\pi_3 + p_{41}\pi_4 \quad p_{12}\pi_1 + p_{22}\pi_2 + p_{32}\pi_3 + p_{42}\pi_4 \quad \cdots \quad \cdots]$$

Thus we have 4 equations (only first two shown):

$$\begin{aligned} &p_{11}\pi_1 + p_{21}\pi_2 + p_{31}\pi_3 + p_{41}\pi_4 = \pi_1 \\ &p_{12}\pi_1 + p_{22}\pi_2 + p_{32}\pi_3 + p_{42}\pi_4 = \pi_2 \\ &\cdots \\ &\cdots \end{aligned}$$

Rearranging and inserting all unknowns in each equation:

$$0p_{12} + 0p_{13} + 0p_{14} + \pi_2 p_{21} + 0p_{23} + 0p_{24} + \pi_3 p_{31} + 0p_{32} + 0p_{34} + \pi_4 p_{41} + 0p_{42} + 0p_{43} = (1 - p_{11})\pi_1$$
$$\pi_1 p_{12} + 0p_{13} + 0p_{14} + 0p_{21} + 0p_{23} + 0p_{24} + \pi_3 p_{31} + \pi_3 p_{32} + 0p_{34} + 0p_{41} + \pi_4 p_{42} + 0p_{43} = (1 - p_{22})\pi_2$$
$$\dots$$
$$\dots$$

This creates the first *n* (here 4) equations represented in matrix form below:

$$\begin{bmatrix} 0 & 0 & 0 & \pi_2 & 0 & 0 & \pi_3 & 0 & 0 & \pi_4 & 0 & 0 \\ \pi_1 & 0 & 0 & 0 & 0 & 0 & 0 & \pi_3 & 0 & 0 & \pi_4 & 0 \\ 0 & \pi_1 & 0 & 0 & \pi_2 & 0 & 0 & 0 & 0 & 0 & 0 & \pi_4 \\ 0 & 0 & \pi_1 & 0 & 0 & \pi_2 & 0 & 0 & \pi_3 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} p_{12} \\ p_{13} \\ p_{14} \\ p_{21} \\ p_{23} \\ p_{24} \\ p_{31} \\ p_{32} \\ p_{34} \\ p_{41} \\ p_{42} \\ p_{43} \end{bmatrix} = \begin{bmatrix} (1 - p_{11})\pi_1 \\ (1 - p_{22})\pi_2 \\ (1 - p_{33})\pi_3 \\ (1 - p_{44})\pi_4 \end{bmatrix}$$

The next *n* (here 4) equations come from equation 14 which is:

$$\sum_{j=1}^{n} p_{ij} = 1$$

Expanding this equation with all unknowns:

$$1p_{12} + 1p_{13} + 1p_{14} + 0p_{21} + 0p_{23} + 0p_{24} + 0p_{31} + 0p_{32} + 0p_{34} + 0p_{41} + 0p_{42} + 0p_{43} = (1 - p_{11})$$
$$0p_{12} + 0p_{13} + 0p_{14} + 1p_{21} + 1p_{23} + 1p_{24} + 0p_{31} + 0p_{32} + 0p_{34} + 0p_{41} + 0p_{42} + 0p_{43} = (1 - p_{22})$$
$$\dots$$
$$\dots$$

This creates the next *n* (here 4) equations represented in matrix form below:

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} p_{12} \\ p_{13} \\ p_{14} \\ p_{21} \\ p_{23} \\ p_{24} \\ p_{31} \\ p_{32} \\ p_{34} \\ p_{41} \\ p_{42} \\ p_{43} \end{bmatrix} = \begin{bmatrix} (1 - p_{11}) \\ (1 - p_{22}) \\ (1 - p_{33}) \\ (1 - p_{44}) \end{bmatrix}$$

Combining these 2*n* (here 8) equations we get the generalized form:

$$\begin{bmatrix} 0 & 0 & 0 & \pi_2 & 0 & 0 & \pi_3 & 0 & 0 & \pi_4 & 0 & 0 \\ \pi_1 & 0 & 0 & 0 & 0 & 0 & 0 & \pi_3 & 0 & 0 & \pi_4 & 0 \\ 0 & \pi_1 & 0 & 0 & \pi_2 & 0 & 0 & 0 & 0 & 0 & 0 & \pi_4 \\ 0 & 0 & \pi_1 & 0 & 0 & \pi_2 & 0 & 0 & \pi_3 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} p_{12} \\ p_{13} \\ p_{14} \\ p_{21} \\ p_{23} \\ p_{24} \\ p_{31} \\ p_{32} \\ p_{34} \\ p_{41} \\ p_{42} \\ p_{43} \end{bmatrix} = \begin{bmatrix} (1-p_{11})\pi_1 \\ (1-p_{22})\pi_2 \\ (1-p_{33})\pi_3 \\ (1-p_{44})\pi_4 \\ (1-p_{11}) \\ (1-p_{22}) \\ (1-p_{33}) \\ (1-p_{44}) \end{bmatrix}$$

B Appendix: Office Data Analysis

Two datasets are analysed from a commercial campus in Eindhoven, Netherlands. The first is obtained from sensors that record positions of persons around desks in an open plan office and the second from an online meeting room reservation service. These datasets have been sufficiently anonymized before receipt such that it would not be possible to identify persons or groups without further information.

B.1 Position Data

This dataset contains positions of persons recorded per second to a temporal and spatial resolution of approximately 1 second and 1 centimetre respectively. Due to confidentiality reasons, the field of view of the sensors cannot be shown here but it consists of about 40 desks, of which not all are occupied, in an open plan office. The sensor observations are from one working day.

Since areas other than the desks and the adjoining corridors are not observed, activities other than ‘working at one’s desk’ are difficult to identify. Therefore, this dataset is used only to verify the theoretical assumption that the sojourn time of the activity ‘working at desk’ is geometrically distributed. To identify whether a desk is occupied or not each desk is assigned an area; if this area is occupied at a given time then the desk is assumed to be occupied. The area assigned includes the desk and the space halfway to the next desk up to a maximum of 1 meter from the desk (Figure 42).

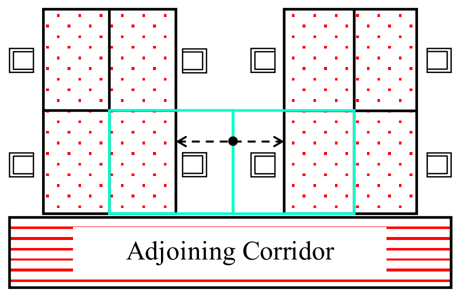

Figure 42: Example of defining areas (cyan) around desks to check occupancy status

However, the occupancy status thus determined is not entirely correct due to false negatives and positives in the sensor observations. Moreover, due to the nature of open plan offices, people passing through to their own desks may cause other desks to be incorrectly labelled as occupied for small periods of time. To overcome these errors that lead to noisy data, the occupancy status is smoothed using a fixed time window of 15 minutes, that is, starting from the beginning of the time period under analysis (0000h), whenever a change in occupancy occurs for a particular desk it is accepted only if the new change lasts for at least 15 minutes. At the beginning, the desks are assumed to be empty. Errors may also be present if the locations of furniture in the building plan and in reality do not match exactly. However, since the desks are fixed in place it is assumed that the building plan gives a correct representation of the situation. To derive the distribution of ‘at desk’ sojourn times, data from all the desks are combined.

B.2 Meeting Rooms Data

Meeting room data from more than 30 meeting rooms in a multi-storey office building occupied by a single company collected for nearly 2 months is used to understand meeting patterns. It is assumed that all meetings are pre-planned. The information available from this dataset includes meeting start time, duration, and capacity of the meeting room booked. The following observations can be made from the analysis:

- Meetings tend to start at times rounded to 30 minutes (Figure 43 (a)). The non-rounded start times may be explained by the fact that it is possible to book rooms instantly at the door.
- Meeting start times have a typical pattern with most meetings being held in the periods of 0900-1100h and 1300-1500h. Hence time is left for lunch hours and fewer meetings are held as people depart for the day (Figure 43 (b)).
- Meeting durations are typically multiples of 15 minutes with most meetings lasting for an hour (Figure 43 (c)).
- Most meetings are held in rooms with capacity of 6 persons with fewer meetings in larger and smaller rooms. However, this observation does not necessarily reflect the number of persons meeting or a preference for this room size as this may, simply, be a result of a higher number of such rooms in the building (Figure 43 (d)).

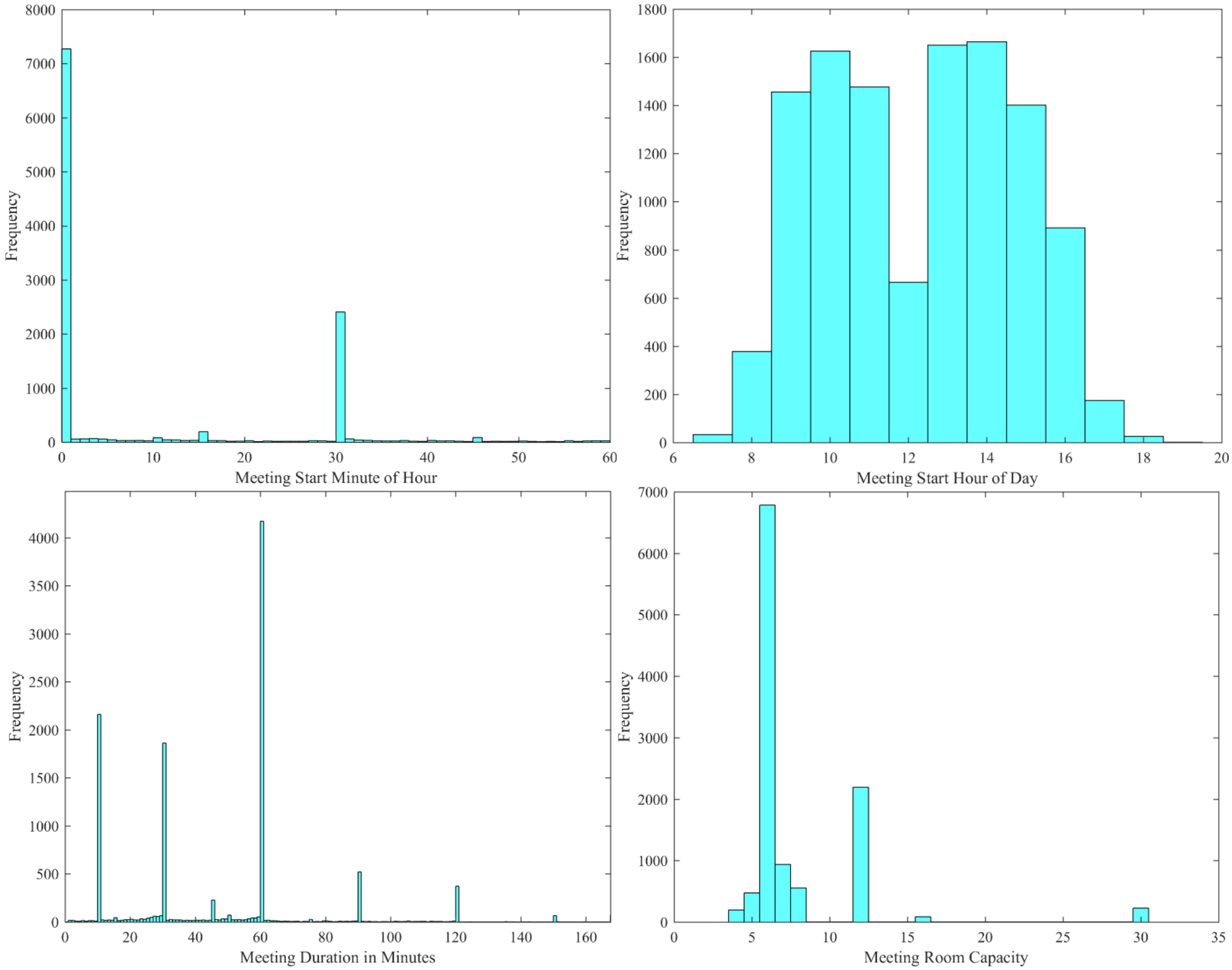

Figure 43: Meeting room data analysis results (row-wise): (a) meeting start time minute, (b) meeting start time hour, (c) meeting duration in minutes, (d) meeting room capacity

C Appendix: Preparing Building Plans

For the model to be able to understand and derive required information from the building plan they have to prepared suitably. Each drawing, prepared in AutoCAD, should consists of six types of layers that add functional context to the physical spaces represented by the building plan: (i) location, (ii) obstacle, (iii) service, (iv) wormhole, (v) door, and (vi) abstract.

- Naming convention for the layers is of the form *typeSubtype*. For example: *locationChair* represents the layer consisting of locations representing chairs in the building plan.
- Table describes the drawing commands in AutoCAD, the definition, and purpose of the different layers.

Table 11: Building plan AutoCAD layers' properties

Layer	**Drawing Command**	**Definition**	**Notes**
Location	POINT	Points associated with an activity but not with a service or wormhole	- Should be approximately 10*x* metres away from obstacles edges. The value of *x* is given by settings.indiWalk.distInteractionWall. Otherwise repulsion forces will make it difficult for pedestrians to approach the locations.
Obstacle	LINE	Static edges acting as boundaries of walkable area	- Currently no subtypes of obstacle type exist so this layer is simply named as obstacle. - Service and wormhole edges are also obstacles. The service and wormhole edges respectively should be drawn on top of the obstacle layer.
Door	LINE	Edges acting as boundaries of walkable area but can be moved by pedestrians	- Currently they have no function as doors are simply treated as gaps by the operational level model. Future implementations may include complex behaviour with respect to interactions with doors.
Wormhole	LINE	Edges across which movements occur using wormholes	- Drawn over corresponding obstacle edge - Each wormhole is automatically associated with a location point on either side. For this a distance must be given in the 'getSettings.m' file. The variable for this should be named as "settings.wormholeLocations.wormhole*Subtype"* (e.g. settings.wormholeLocations.wormholeTurnstile) and the distance value in meters should be assigned. - The values should be greater than 10*x* metres as explained for the obstacle layer. - It should be made sure that there is indeed space available on either side of the wormhole edge for the assigned distance. - The throughput time in seconds should be assigned in the same file under the variable "settings.serviceTime.wormhole*Subtype*" - Currently the model only connects wormholes of the same edge.
Service	LINE	Edges associated with one activity location which follows a known duration distribution	- Drawn over corresponding obstacle edge - Each service is automatically associated with a location point on one side. For this a distance must be given in the 'getSettings.m' file. The variable for this should be named as "settings.serviceLocations.service*Subtype"* (e.g. settings.serviceLocations.serviceCoffee) and the distance value in meters should be assigned. - The values should be greater than 10*x* metres as explained for the obstacle layer. - It should be made sure that there is indeed space available in front of the service edge for the assigned distance. The depth of the obstacle whose one edge is the service edge should be greater than the associated distance so that the model can identify which side is in the 'front'. - The service time in seconds should be assigned in the same file under the variable "settings.serviceTime.wormhole*Subtype*"

Abstract	RECTANG	Rectangle defining the function of a specific area in the building	- Drawn around group of obstacles/service/wormholes edges and locations that have a coherent function - Does not need to be perfectly coincident with the boundaries of the room as long as the related locations and edges are included.

- It should be made sure that all obstacle edges are constructed using one and only one line for every two points that need to be connected. The command EXPLODE will help separate polylines and blocks into such lines. The command OVERKILL can be used to remove overlapping and duplicate lines. Wormhole/service edges should remain overlapping with the obstacle edges.
- The building plan must only consist of straight lines and no splines. Curved obstacles should be suitably divided into lines. For example, a circular table may be converted into a hexagon of the same centre to vertex distance as the radius of the circular form.
- In order to import AutoCAD drawings, currently the following steps are followed:
    - Save drawing in AutoCAD in the .dxf format
    - Import this format into a GIS program (QGIS: available freely from www.qgis.org). QGIS maintains the layer information while other GISs have not been tested.
    - Use EPSG 32633 as the coordinate reference system (CRS). This is chosen as it has distance in metres. Since office building areas are much smaller (compared to the earth) the associated projection method will cause almost no distortion.
    - The edges and locations form two separate layers. Save as shapefiles (.shp) and note file locations under 'getFiles.m' file.