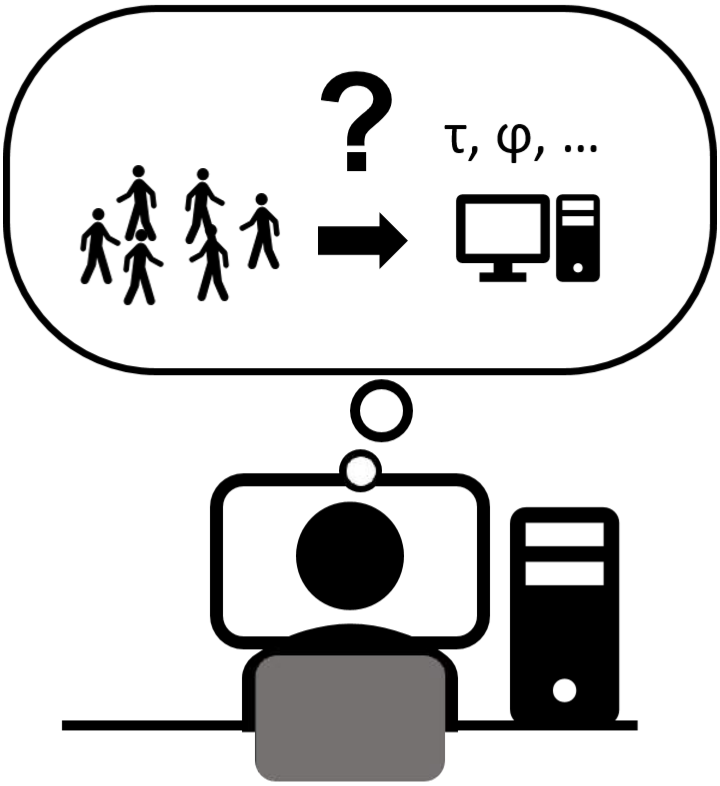
How to calibrate a pedestrian simulation model

An investigation into how the choices of scenarios and metrics influence the calibration



Martijn Sparnaaij
September 2017

Master Thesis
Delft University of Technology
Faculty of Civil Engineering
Transport & Planning

How to calibrate a pedestrian simulation model

An investigation into how the choices of scenarios and metrics influence the calibration

By

Martijn Sparnaaij

in partial fulfilment of the requirements for the degree of
Master of Science
in Civil Engineering
at the Delft University of Technology,
to be defended publicly on Monday September 4, 2017 at 01:00 PM.

Student number:	1518690	
Daily supervisor:	Dr. ir. D.C. Duives	TU Delft, Transport & Planning
Thesis committee:	Prof. dr. ir. S.P. Hoogendoorn	TU Delft, Transport & Planning (chair)
	Dr. V.L. Knoop	TU Delft, Transport & Planning
	Prof. dr. ir. C. Vuik	TU Delft, Numerical Analysis

Preface

This thesis marks the end of my graduation project on the question of how to calibrate a pedestrian simulation model. This thesis has been produced in cooperation with Incontrol simulation solutions and is the final part of my master's in Civil Engineering at the Delft University of Technology.

The thesis has been created with the help of a number of people who contributed in various ways and to whom I would like to express my gratitude. Firstly, I would like to thank Incontrol for providing me with access to their pedestrian simulation model and especially Jeroen Steenbakkers for making this possible and Jeroen Bijsterbosch for his expert help on getting to know the ins and outs of the model. Secondly, I thank all those involved in collecting the data for the HERMES project and especially the Jülich Supercomputing Centre for making this data publicly available for data is a vital element in a calibration study. Furthermore, I would also like to express my gratitude to the members of my graduation committee - Serge Hoogendoorn, Victor Knoop and Kees Vuik for their feedback and support.

Most importantly, I would like to thank Dorine Duives for her role as my daily supervisor during the whole project. The many fruitful discussions, the detailed feedback, the encouragements and support have been invaluable in bringing this project to a good end.

The participants of the monthly PET meetings also deserve a word of gratitude for giving me the opportunity to present the first results halfway through project and providing feedback on those results and for suggesting solutions to some of the problems I encountered during the project. Furthermore, I would like to thank all fellow students at the hok at Civil Engineering for all the nice discussions and good laughs during all the coffee and lunch breaks and the generally nice time at the hok. And lastly, dear family, friends and flat mates, thank you for all the support and encouragements I received from you during this project.

Martijn Sparnaaij

Delft, September 2017

Summary

Today, pedestrian simulation models are used for many different applications. These applications range from evacuation studies and risk analysis to assisting in the design process of pedestrian facilities. One of the important steps in developing a model is calibration. By estimating the values of the model's parameters based on data from actual pedestrian flows, the goal of calibration is to increase the accuracy of the model's predictions. And although this is an important step in model development, calibration has received relatively little attention in the field of pedestrian modelling. Most calibration efforts were found to have a limited focus whereby the focus is limited because only one flow situation or metric is used. Meanwhile, previous research by Campanella (2016) and Duives (2016), both performed at the Transport & Planning (T&P) department of the Delft University of Technology, has shown that using different (combinations of) flow situations and/or metrics during the calibration leads to different optimal parameter sets. Hence, it is questionable how valid it is to use a model, that has been calibrated using such a limited focus, to make predictions in other flow situations or for other metrics. To overcome this problem it has been proposed to use multiple objectives whilst calibrating a pedestrian model. However, the question what these objectives should be is largely unanswered. So, the goal of this research is to *"gain improved insights into how the choice of objectives influences the calibration results thereby improving the applicability of the multiple-objective framework"*. In order to reach this goal the following main research question is answered in this thesis:

> ***How can a microscopic pedestrian model be calibrated, using a multiple-objective approach, given its stochastic nature and differences in behaviour in different flow situations?***

So, which flow situations and metrics does one need to adequately represent the differences in pedestrian behaviour in different flow situations given that the model is stochastic in nature?

Objectives

In this research an objective, that is used to determine how well the model results fit to the data, is defined as a unique combination of a flow situation and a metric. The literature review in this research identified five properties of flow situations which possibly influence the flow and are hence relevant when choosing the objectives to use during the calibration. These five properties are the geometry of the infrastructure, the demand patterns, the population composition, the movement base case and the density level. In this research the implementation of a flow situation in the model is called a scenario. The literature review also unravelled many different metrics which can be categorized based on three properties. Namely, whether they are: 1) Macroscopic, mesoscopic or microscopic, 2) quantitative or qualitative, and 3) generally usable or linked to a specific flow situation.

The availability of reference data limited the number of scenario properties which could be researched. So, only the influence of the choice of movement base cases and density levels have been studied. Table 1 and Figure 1 illustrate which scenarios were used within this research.

Table 1: Overview of the scenarios used in this research

Scenario name	**Movement base case(s)**	**Density level**	
		Low	**High**
Bidirectional	Bidirectional straight	x	x
Bottleneck	Unidirectional entering and exiting		x
Corner	Unidirectional corner	x	x
T-junction	Merging unidirectional flows	x	x

On top of the scenarios described above, four metrics were chosen. These four metrics are all quantitative and generally usable and cover both the mesoscopic and macroscopic levels. On the macroscopic

level the flow and the spatial distribution were used and on the mesoscopic level the travel time distribution and the distribution of the effort were used. The four chosen metrics were chosen such that at both aggregation levels they measure different aspects of the flow. The flow and travel time give insight into the efficiency of the flow whilst the spatial distribution and effort give more insight into the underlying behaviour.

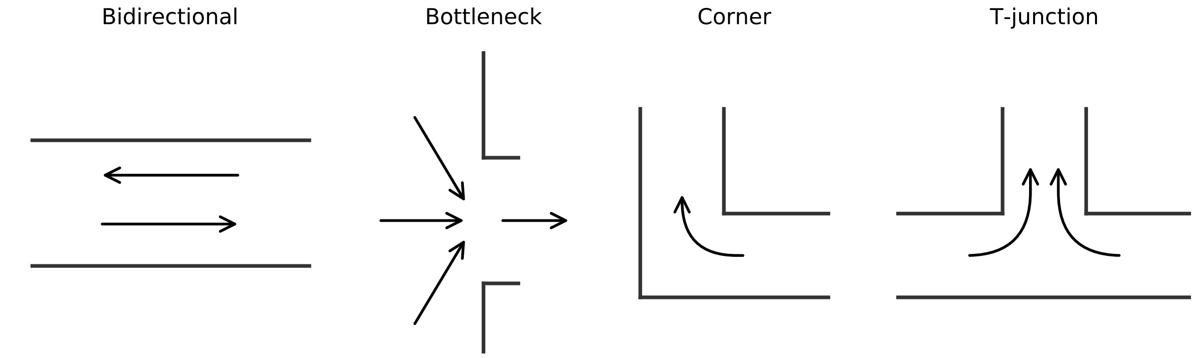


Figure 1: Overview of the movement base cases used in this research

Pedestrian Dynamics

During this research Pedestrian Dynamics ® (PD) has been used as the pedestrian simulation model. PD is a microscopic simulation model that contains the ability to model all three behavioural levels of pedestrian behaviour (strategic, tactical and operational). This research focussed on the calibration of the operational level and hence on the parameters that influence the route following and collision avoidance behaviour. A total of 11 different parameters influence these two behaviours and a sensitivity analysis was performed to obtain the model's sensitivity to those parameters. The sensitivity analysis investigated only the first-order effects but did include all seven scenarios in Table 1. The sensitivity of all parameters was investigated for a range of $\pm 25\%$ of the default value. Before a quantitative analysis was performed, a qualitative analysis was performed to ascertain if the model produced realistic behaviour for the maximum deviations of $\pm 25\%$ of the default value. It was found that the model did not produce unrealistic behaviour at those boundaries for any of the parameters. The quantitative analysis, which used the distribution of the instantaneous speeds as the sole metric, showed clear differences in the sensitivity of the model to the different parameters and also clear difference between the seven scenarios. It was concluded that the model was most sensitive to the relaxation time and the viewing angle and that these two parameters are the most important to calibrate.

How to deal with the stochastic nature

Like most pedestrian models PD is stochastic in nature. Hence this research also investigated how one should deal with this stochastic nature. The primary questions were how many replications one needs and how one should determine this number. The method to determine the number of replications was based on the test if the distribution of the instantaneous speeds had converged whereby the speed distribution contained the data of all replications up to that point. An Anderson-Darling test (A-D test) was used to test if for the last n consecutive replications the speed distributions were samples considered to be drawn from the same distribution. If this was the case the distribution was considered to have been converged. If this was not the case additional replications were run.

The method was tested for various values of n and for various significance levels required for the A-D test. However, it was found that regardless of the values used, the order of the seeds has influence on the results. Even when 500 replication were used, (a far larger number than was found by the method) the seed order still had a detectable and significant effect. Hence, during the calibrations efforts a fixed seed order and number of replications was used to ensure that any differences found were not the result of the stochasticities.

Multiple-objective calibration of pedestrian models

To test how the choice of scenarios and metrics influences the calibration results, the model was calibrated using 16 different combinations of objectives. These are: The seven individual scenarios using all four metrics, the four metrics using all scenarios, the combination of all scenarios of the same density

level using all metrics, the combination of all metrics of the same aggregation level using all scenarios and lastly a combination of all scenarios and metrics. The objectives were combined into a single objective using the weighted sum method whereby the objectives were weighed equally. Because the metrics have different units and different orders of magnitude a normalization method was used to ensure that the results of different metrics could be added up in a meaningful way. The normalization method is based on normalization values which are in turn based on the ratios between the different metrics in the reference data. The search space was made up out of the relaxation time, the viewing angle and the radius and a grid-search was used to determine the optimal parameter set.

It was found that the optimal parameter sets obtained using these 16 different combinations of objectives showed large differences. These differences in parameter sets already showed that the choice of objectives has a strong influence on the calibration results. To obtain a more detailed insight, a number of comparisons were made between the different combinations of objectives to answer three questions. Namely, 1) How does the choice of movement base case influence the calibration results? 2) How does the choice of density level influence the calibration results? And, 3) How does the choice of metrics influence the calibration results? The results of the comparisons between different combinations of objectives showed the following:

- For all movement base cases it was found that the model's performance, with respect to a particular movement base case, decreases when an optimal parameter set is used that has been obtained using either another movement base case or a set of movement base cases. However, on average, this decrease is smallest when the optimal parameter set is used that has been obtained using the combination of movement base cases. Furthermore, this was found to be far more apparent in the high density cases than in the low density cases.
- A calibration using solely low density scenarios cannot capture the behaviour found in the high density scenarios. This is not the case vice versa. It is not only the case when combinations of movement base cases were used but it was also found to hold in the cases when two scenarios of the same movement base case, but different levels of density, were compared.
- Different metrics lead to different optimal parameter sets. Furthermore, the model cannot obtain a good fit on all metrics simultaneously.

These findings are in line with the findings of previous research and the findings of the sensitivity analysis.

Conclusions

Based on the findings the following can be concluded:

- It is necessary to use multiple movement base cases, when calibrating a model, to capture all relevant behaviour. However, this does decrease the Goodness-of-Fit (GoF) of the individual movement base cases compared to the case where they are calibrated solely based on that particular base case.
- The level of density does influence the calibration results whereby it is particularly important to include scenarios with the highest density levels, one wants the model to be able to reproduce, in the set of scenarios one uses for the calibration.
- The choice of metric or combinations of metrics does influence the results. Whereby, depending on the combination of metrics, also the choice of objective function and normalization method influences the results. Also, the model cannot obtain a good fit on all metrics simultaneously.

A critical review of the used calibration methodology showed that there are multiple ways in which changing the methodology could affect the results. Examples are, using a more precise grid, another objective function or another seed order. However, how it would change the results exactly and if it would significantly change the conclusions presented above was not determined given that this would require more quantitative analyses for which there was no time in this research. However, the fact that the main findings are in line with previous research and the results of the sensitivity analysis and the fact that large differences are found between optimal parameter sets of different (combinations of) objectives, are considered strong indications that it is unlikely that the main findings would change significantly.

Implications

The findings and resulting conclusions have two main implications. Firstly, one should calibrate a pedestrian simulation model based on the intended application of the model. So, the calibration should include those scenarios that are likely to occur to ensure that all relevant behaviour is captured. However, including scenarios that are unlikely to occur, given the intended application, will likely influence the calibration results negatively and hence also the model's accuracy with which it can predict the traffic state. Furthermore, the metrics and accompanying objective functions and method for combining the multiple objectives into a single objective, should be chosen such that they represent which metrics are most important to the intended application and what level of accuracy one wants the model to have regarding the particular metrics

Secondly, the fact that the results of this study are in line with the previous studies, whilst using a different model than previous studies, raises an important question. Namely, is the fact that the models cannot obtain a good fit on different scenarios using only a single parameter set caused by the fact that:

a.) the models are simplifications of the behaviour of pedestrians and the models are too simple to capture the behaviour of pedestrian in different flow situations well using only a single parameter set. Or,

b.) the behaviour of pedestrians in different flow situations is so different that it might not be a valid approach to try to capture this using a single model (i.e. the assumption that the behaviour of the pedestrians is independent of the flow situation is not valid).

The results of this research cannot answer this question. However, it is important to answer this question given that, in the case it would be the second cause listed above, it would fundamentally change the way in which we need to model pedestrian behaviour.

Recommendations

Based on this research a number of recommendations were given. To practice, the main recommendations are to take into account the intended use of the model when choosing which objectives to include during the calibration. And, to validate it using a wider range of objectives to obtain insight into how reliable to model's predictions are for those objective that weren't included during the calibration.

To science the main recommendations are suggestions for further research. Besides further investigations into the effect of the choice of objectives on the calibration results, investigations into how to deal with the stochastic nature are considered worthwhile. Possibly the most important suggestion is to investigate the question if it is a valid modelling approach to use a single model for all flow situations. Given that the answer to this question could potentially warrant a fundamental change in how we model pedestrian behaviour.

Contents

Preface i

Summary ii

List of Figures viii

List of Tables x

1 Introduction 1
1.1 Research objective . . . 2
1.2 Research scope . . . 2
1.3 Research questions . . . 4
1.4 Contributions of this research . . . 4
1.5 Thesis outline . . . 5

2 State-of-the-art of calibrating pedestrian simulation models 6
2.1 State-of-the-art of calibration methods for pedestrian models . . . 6
2.2 Framework elements . . . 8
2.3 Conclusions . . . 17

3 Research methodology 19
3.1 Reference data . . . 19
3.2 Scenarios . . . 22
3.3 Pedestrian Dynamics . . . 29
3.4 How to deal with stochasticity . . . 33
3.5 Conclusions . . . 37

4 Sensitivity analysis 38
4.1 Methodology . . . 38
4.2 Qualitative analysis . . . 40
4.3 Quantitative analysis . . . 43
4.4 Conclusions . . . 50

5 Multiple-objective calibration 51
5.1 Calibration methodology . . . 51
5.2 Analysis of the individual objectives . . . 59
5.3 Analysis of the combined objectives . . . 61
5.4 Discussion on the calibration results and the methodology . . . 70
5.5 Practical implications of the results . . . 73
5.6 Conclusions . . . 74

6 Conclusions, discussion and recommendations 76
6.1 Conclusions . . . 76
6.2 Discussion . . . 77
6.3 Recommendations . . . 79

Acronyms 81

Bibliography 82

Appendix A Overview metrics 87

Appendix B Scenarios sensitivity analysis 91

Appendix C Results quantitative sensitivity analysis 93
C.1 Bidirectional straight - high . . . 93
C.2 Bidirectional straight - low . . . 94
C.3 Bottleneck . . . 95
C.4 Corner - high . . . 96
C.5 Corner - low . . . 97
C.6 T-junction - high . . . 98
C.7 T-junction- low . . . 99

Appendix D Scenarios calibration 100

Appendix E Measurement periods calibration 102

Appendix F Objective functions and normalization method 104

Appendix G Reference data - Metrics 106

Appendix H Detailed calibration results 108
H.1 Correlation between mean travel time error and pedestrian count error . . . 108
H.2 How is the decrease in GoF related to the changes in the errors? . . . 108
H.3 Tables split out per individual objective . . . 113

Appendix I Calibration - different objective functions 121

Appendix J Decrease in GoF versus the distance from the optimal parameter set 122

List of Figures

1 Overview of the movement base cases used in this research iii

1.1 Overview of the model development framework 1
1.2 Outline of the thesis 5

2.1 Overview of the multiple-objective framework 7
2.2 The eight movement base cases in pedestrian flow 9

3.1 Overview research steps 19
3.2 Overview of the available experiments of the TU Delft data set 20
3.3 Overview experiments Hermes 22
3.4 Overview of the bidirectional flow scenario 24
3.5 Difference in trajectories between the free and forced destination choice in the reference data for a bidirectional flow 25
3.6 Cumulative curves bidirectional scenario 25
3.7 Fundamental diagrams (ku) bidirectional scenario 25
3.8 Overview of the bottleneck scenario 26
3.9 Cumulative curve in the bottleneck 26
3.10 Fundamental diagrams (ku) bottleneck scenario AO-360-400 27
3.11 Overview of the corner scenario 27
3.12 Cumulative curves corner scenario 28
3.13 Fundamental diagrams (ku) corner scenario 28
3.14 Overview of the T-junction scenario 28
3.15 Fundamental diagrams (ku) T-junction scenario 29
3.16 Cumulative curves T-junction scenario 29
3.17 Example of the parameters determining the desired direction. 30
3.18 Required number of replications per combination for different values of $p_{\text{threshold}}$ and b . . 35
3.19 Influence of the order of the seed versus the number of replications 36

4.1 Overview of the methodology for the sensitivity analysis of a single parameter and scenario 39
4.2 Sensitivities - Bidirectional high flow 45
4.3 Sensitivities check - Bidirectional high flow 45
4.4 Sensitivities - Bidirectional low flow 46
4.5 Sensitivities check - Bidirectional low flow 46
4.6 Sensitivities - Bottleneck 46
4.7 Sensitivities check - Bottleneck 46
4.8 Sensitivities - Corner high flow 47
4.9 Sensitivities check - Corner high flow 47
4.10 Sensitivities - Corner low flow 47
4.11 Sensitivities check - Corner low flow 48
4.12 Sensitivities - T-junction high flow 48
4.13 Sensitivities check - T-junction high flow 48
4.14 Sensitivities - T-junction low flow 49
4.15 Sensitivities check - T-junction low flow 49

5.1 Measured and fitted free speed distributions 53
5.2 Box plots of the non-squared, non-normalized errors per scenario and metric. 59
5.3 The CDFs showing how the objective values are distributed for the four high density scenarios 66
5.4 The CDFs showing how the objective values are distributed for the all seven scenarios . . 68
5.5 The CDFs showing how the objective values are distributed for the four metrics 69
5.6 Example of the influence of the precision of the search space on the optimal parameter set 72

B.1 Overview bidirectional scenario 91
B.2 Overview bottleneck scenario 91
B.3 Overview corner scenario 92
B.4 Overview T-junction scenario 92

C.1 Influence of seeds versus the number of replications - Bidirectional straight high 93
C.2 Influence of seeds versus the number of replications - Bidirectional straight low 94
C.3 Influence of seeds versus the number of replications - Bottleneck 95
C.4 Influence of seeds versus the number of replications - Corner high 96
C.5 Influence of seeds versus the number of replications - Corner low 97
C.6 Influence of seeds versus the number of replications - T-junction high 98
C.7 Influence of seeds versus the number of replications - T-junction low 99

D.1 Overview bidirectional scenario 100
D.2 Overview bottleneck scenario 100
D.3 Overview corner scenario 101
D.4 Overview T-junction scenario 101

G.1 Spatial distributions 107

H.1 Correlation between mean travel time error and pedestrian count error - Bidirectional high density scenario (p is the Pearson correlation coefficient) 108

J.1 The decrease in the GoF versus the distance from the optimal parameter set for the individual scenarios - all metrics (combinations 1-7). 124
J.2 The decrease in the GoF versus the distance from the optimal parameter set for the individual metrics - all scenarios (combinations 8-11) 125
J.3 The decrease in the GoF versus the distance from the optimal parameter set for the different combinations of scenarios and metrics (combinations 12-15) 125
J.4 The decrease in the GoF versus the distance from the optimal parameter set for the whole range of decreases in GoF for the individual scenarios - all metrics (combinations 1-7). . . . 126
J.5 The decrease in the GoF versus the distance from the optimal parameter set for the whole range of decreases in GoF for the individual metrics - all scenarios (combinations 8-11). . 127
J.6 The decrease in the GoF versus the distance from the optimal parameter set for the whole range of decreases in GoF for the combinations of scenarios and metrics (combinations 12-15). 127

List of Tables

1 Overview of the scenarios used in this research . . . ii

2.1 The nine identified elements of the multiple-objective framework . . . 8
2.2 Overview of optimization methods found in the literature . . . 16

3.1 Overview of the available data TU Delft . . . 21
3.2 Overview of the available data Hermes project . . . 21
3.3 Coverage of the movement base cases by the reference data . . . 23
3.4 Overview of the scenarios used in this research . . . 23
3.5 Operational parameter of PD . . . 32
3.6 The values of $p_{\text{threshold}}$ and b used in the tests . . . 35

4.1 Combinations of scenarios and parameters investigated during the sensitivity analysis . . . 40

5.1 Overview of the OD-pairs and the demand per OD-pair . . . 52
5.2 Normalization values . . . 56
5.3 Search space definition . . . 58
5.4 Nr. of replications per scenarios . . . 58
5.5 Tested combinations of objectives. . . . 62
5.6 The optimal parameter set per combination of objectives whereby $O(\theta^*)$ represents the minimal value of the objective function. The cells in red indicate that the value is at the upper or lower boundary of the search space . . . 63
5.7 Difference in the GoF for the comparisons between different movement base cases. . . . 64
5.8 Difference in the GoF for the comparisons between different density levels for every movement base case. . . . 67
5.9 Difference in the GoF for the comparisons between different density levels for combinations of movement base cases. . . . 67
5.10 Difference in the GoF for the comparisons between different metrics. . . . 69

A.1 Overview of metrics, found in the literature, used to evaluate the performance of a pedestrian simulation model. . . . 88

B.1 Inflows per scenario in ped/s/m/entry . . . 92

C.1 Deviation from the default value [%] - Bidirectional straight high . . . 93
C.2 Deviation from the default value [%] - Bidirectional straight low . . . 94
C.3 Deviation from the default value [%] - Bottleneck . . . 95
C.4 Deviation from the default value [%] - Corner high . . . 96
C.5 Deviation from the default value [%] - Corner low . . . 97
C.6 Deviation from the default value [%] - T-junction high . . . 98
C.7 Deviation from the default value [%] - T-junction low . . . 99

E.1 Measurement periods . . . 102
E.2 Measurement period duration in seconds . . . 102

F.1 Ratios between metric values per scenario . . . 105

G.1 Measured flows [ped/s/m] per data set . . . 106
G.2 Measured effort [m/s] per data set . . . 106
G.3 Measured travel time per data set . . . 106

H.1 An overview of the changes in errors of the combination of low density cases when, instead of its own optimal parameter set, the optimal parameter set of the bidirectional low density scenario is used . . . 108

H.2 An overview of the decreases in GoF and the accompanying descriptive statistics of the errors whereby the base case is the combination of the high density scenarios 109
H.3 An overview of the decreases in GoF and the accompanying descriptive statistics of the errors whereby the base case is the combination of the low density scenarios 109
H.4 An overview of the decreases in GoF and the accompanying descriptive statistics of the errors whereby the base case is the high density bidirectional scenario 110
H.5 An overview of the decreases in GoF and the accompanying descriptive statistics of the errors whereby the base case is the bottleneck scenario 110
H.6 An overview of the decreases in GoF and the accompanying descriptive statistics of the errors whereby the base case is the high density corner scenario 110
H.7 An overview of the decreases in GoF and the accompanying descriptive statistics of the errors whereby the base case is the high density t-junction scenario 110
H.8 An overview of the decreases in GoF and the accompanying descriptive statistics of the errors whereby the base case is the combination of the low density bidirectional scenario 111
H.9 An overview of the decreases in GoF and the accompanying descriptive statistics of the errors whereby the base case is the combination of the low density corner scenario 111
H.10 An overview of the decreases in GoF and the accompanying descriptive statistics of the errors whereby the base case is the combination of the low density t-junction scenario . . 111
H.11 An overview of the decreases in GoF and the accompanying descriptive statistics of the errors whereby the base case is the flow metric 112
H.12 An overview of the decreases in GoF and the accompanying descriptive statistics of the errors whereby the base case is the spatial distribution metric 112
H.13 An overview of the decreases in GoF and the accompanying descriptive statistics of the errors whereby the base case is the effort metric 112
H.14 An overview of the decreases in GoF and the accompanying descriptive statistics of the errors whereby the base case is the travel time metric 112
H.15 Difference in the GoF for the comparisons between different movement base cases using the **flow** as the only metric. 114
H.16 Difference in the GoF for the comparisons between different movement base cases using the **spatial distribution** as the only metric. 114
H.17 Difference in the GoF for the comparisons between different movement base cases using the **effort** as the only metric. 114
H.18 Difference in the GoF for the comparisons between different movement base cases using the **travel time** as the only metric. 115
H.19 Difference in the GoF for the comparisons between different density levels for every movement base case using the **flow** as the only metric. 116
H.20 Difference in the GoF for the comparisons between different density levels for combinations of movement base cases using the **flow** as the only metric. 116
H.21 Difference in the GoF for the comparisons between different density levels for every movement base case using the **spatial distribution** as the only metric. 116
H.22 Difference in the GoF for the comparisons between different density levels for combinations of movement base cases using the **spatial distribution** as the only metric. 116
H.23 Difference in the GoF for the comparisons between different density levels for every movement base case using the **effort** as the only metric. 117
H.24 Difference in the GoF for the comparisons between different density levels for combinations of movement base cases using the **effort** as the only metric. 117
H.25 Difference in the GoF for the comparisons between different density levels for every movement base case using the **travel time** as the only metric. 117
H.26 Difference in the GoF for the comparisons between different density levels for combinations of movement base cases using the **travel time** as the only metric. 117
H.27 Difference in the GoF for the comparisons between different metrics using the **bidirectional high density** scenario as the sole scenarios. 118
H.28 Difference in the GoF for the comparisons between different metrics using the **bidirectional low density** scenario as the sole scenarios. 119
H.29 Difference in the GoF for the comparisons between different metrics using the **bottleneck** scenario as the sole scenarios. 119
H.30 Difference in the GoF for the comparisons between different metrics using the **corner high density** scenario as the sole scenarios. 119

H.31 Difference in the GoF for the comparisons between different metrics using the **corner low density** scenario as the sole scenarios. . . . 120
H.32 Difference in the GoF for the comparisons between different metrics using the **t-junction high density** scenario as the sole scenarios. . . . 120
H.33 Difference in the GoF for the comparisons between different metrics using the **t-junction low density** scenario as the sole scenarios. . . . 120

I.1 Comparison of the optimal parameter sets resulting from either using a squared objective function or an absolute error objective function. . . . 121

1 | Introduction

In the past decades many different pedestrian simulation models have been developed in different fields of science. Reviews, such as those by Papadimitriou, Yannis and Golias (2009), Schadschneider, Klüpfel, Kretz, Rogsch and Seyfried (2009) and Duives, Daamen and Hoogendoorn (2013), clearly show these models vary widely. The models use different aggregation levels to describe the behaviour, from macroscopic to microscopic and everything in between, and use different modelling approaches, using cellular automata, discrete choice and social forces among others.

In practice, these models have a multitude of applications. This is clearly shown by the variety of cases presented on the websites of some of the commercial packages (among others, Pedestrian Dynamics by INCONTROL Simulation Solutions (2016), PTV Viswalk by PTV Group (2016) and SimWalk by Savannah Simulations AG (2016)). These application range from evacuation studies and risk analysis to assisting in the design process of pedestrian facilities. Some examples are:

- evaluating if and where potential bottlenecks exist at an event terrain, a train station or an airport and how and if these bottleneck can lead to risky situations such as overcrowding.
- assisting in the design process of a train station or airport whereby the goal could be to ensure that all passengers can comfortably walk from their origin to their destination.
- an evacuation study of a stadium or a cruise ship to assess the time necessary for a successful evacuation

Ideally, a multitude of steps has to be taken before one has a (commercial) implementation of a pedestrian model that can be used in practice. In Figure 1.1 these steps, described in both (Klügl, 2008) and (Duives, Daamen & Hoogendoorn, 2016), are depicted. As the figure illustrates, the process starts with the proposition of the model whereby a conceptual model is created and implemented which results in a runnable model. After this the model is verified whereby it is determined if the implementation of the model is indeed consistent with the conceptual model. If this is indeed the case the model can be calibrated and validated. In the calibration step the main goal is to increase the accuracy of the predictions by finding values for the model parameters that allow for better replication of reality based on data collected from laboratory experiments or real-life situations. Unsatisfactory results, for example no parameter set can be found which results in a reasonable fit to the data, can lead to a revision of the proposed model as is indicated by the feedback loop in the figure. The validation step checks how valid and accurate the predictions of the model are also using data, however, different data then was used for the calibration process. Again, as the feedback loop indicates, unsatisfactory results in the validation step can lead to a revision of either the calibration procedure or the model itself.

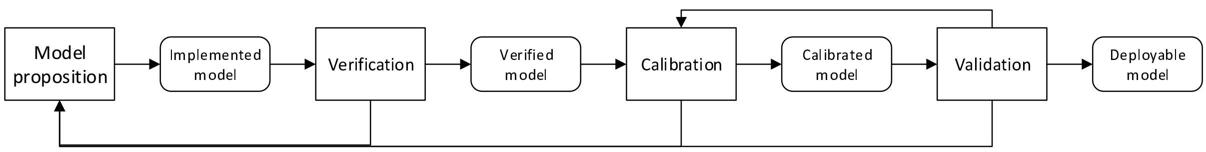


Figure 1.1: Overview of the model development framework

The calibration step is an important step because one wants to get a model whose predictions are as accurate as possible given all limitations such as data availability and computing power. And although the calibration step is clearly important, it has received relatively little attention within the field of pedestrian modelling (Seer, Brändle & Ratti, 2014; Rudloff, Bauer, Matyus & Seer, 2011). Most studies that have performed a calibration of a model have a limited focus (Berrou, Beecham, Quaglia, Kagarlis & Gerodimos, 2007; Davidich & Köster, 2013). The focus of these studies is limited because they, for example, only take into account one flow situation or only use one metric. Meanwhile, research by Campanella, Hoogendoorn and Daamen (2011) and Duives (2016) shows that using different flow situations and different metrics leads to different optimal parameter values and hence it is questionable how generally applicable a model is that has been calibrated using a limited focus. To overcome this problem

Campanella et al. (2011) have proposed a framework for calibrating pedestrian models using multiple objectives. Compared to a single objective approach (i.e. using one flow situation and one metric), the multiple-objective approach uses multiple flow situations and multiple metrics such that the resulting parameter set should result in a better overall performance instead of only performing well in one specific flow situation. Subsequent work by Campanella, Hoogendoorn and Daamen (2014) showed that it is indeed the case that using a parameter set obtained by calibrating using multiple objectives results in a better validation score whereby the validation also included different flow situations and metrics.

As far as the author knows, the only occurrences of the use of the multiple-objective framework for calibrating a pedestrian model in the literature are (Campanella et al., 2011) and (Duives, 2016). As will be shown in chapter 2 these studies clearly show the need for multiple objectives when calibrating pedestrian models. However, chapter 2 will also show that what these objectives should be is still largely unanswered.

1.1 Research objective

So, as is shown in the introduction and in chapter 2:

1. The applications of pedestrian models require accurate predictions of the traffic state
2. However, an essential step in the creation of a deployable and accurate model, calibration, is rarely performed and in most cases if it is performed it is done with a limited focus.
3. This limited focus leads to a questionable general applicability of a model
4. The proposed solution is the use of multiple objectives
5. However, what these objectives should be is a largely unanswered question

Given the statements above the goal of this research is formulated as follows:

> ***The objective of this research is to gain improved insights into how the choice of objectives influences the calibration results thereby improving the applicability of the multiple-objective framework.***

In the case of pedestrian models, an objective is a defined by a unique combination of a scenario and a metric. A scenario is the model implementation of a flow situation defined by a number of elements as will be shown in subsection 2.2.2. Out of these five properties, only the effect of the movement case has been investigated. So, the aim of improving insight into the effect implies two things: Firstly, by using another model and other data sets (see chapter 3) it can be investigated if the previous findings (regarding the movement base cases and the metrics) also hold when another model and data set is used (i.e. are the previous finding generalizable?). And secondly, by looking into other scenario properties, other than the movement base case, additional insights should be gained regarding the effect of other scenario properties.

1.2 Research scope

In this section the scope of the research is defined whereby it is discussed, which behavioural level is included, what model will be used, what the main goal of the calibration is and which elements of the calibration are part of this research.

Behavioural level

As has been put forward in (Hoogendoorn, 2001), pedestrian behaviour can be divided into three behavioural levels namely:

- Strategic: At the strategic level a pedestrian decides which activities he/she wants to perform. This is the process of activity choice and the result is a collection of activities called the activity set.

- Tactical: At the tactical level two processes take place. Namely, the pedestrian schedules at which location the activities in the activity set are performed and in which order. This is the process of activity scheduling and the result is an activity schedule. The second process that takes place at the tactical level if the planning of the route to the activities. The process of route choice results in a route the pedestrian follows from his/her current location to the next planned activity.
- Operational: The operational level entails the walking behaviour of a pedestrian. It controls both how a pedestrian follows his/her planned route (route following) and the avoidance of collisions with objects or other pedestrian (collision avoidance).

When using pedestrian models in practice, one needs to simulate the behaviour of the pedestrians at all three levels described above. All three levels also require calibration, however, this research will, in line with previous research, limit itself to the operational behaviour.

Model

In this research the choice is made to use Pedestrian Dynamics ® (PD) by InControl as the pedestrian simulation model. The reason for this choice is two-fold, namely I have experience with the program and it is a commercial implementation that is available due to the cooperation between InControl and the TU Delft.

It also has to be noted that the model is solely a tool and that it is not the aim of this research to provide a calibrated model (i.e. it is not the goal to find the optimal parameter set which should be used in practice).

Goal of the model

As is stated in (Campanella et al., 2014), the main applications of pedestrian models are related to either safety or comfort and thus they are mostly used to predict aggregate data. This means that the primary goal of the calibration is to increase the accuracy of the prediction of the aggregate flow and not necessarily the accuracy of the individual behaviour. Preferably one is able to capture both the individual behaviour and the resulting aggregate flow accurately. However, as will be shown in subsection 2.2.3, as of yet, models do not seem to be capable of producing accurate results for both the individual behaviour and the aggregate flow for a single parameter set.

Furthermore, as is clear from the examples of the applications in the introduction, most applications are focussed on non-panicky situations (an orderly evacuation is also considered to be a non-panicky situation). Hence it is assumed that the goal of the model is to predict the traffic state in a normal situation where pedestrian show calm, non-panicking behaviour.

Elements

As will be shown in section 2.2, calibrating a pedestrian model using multiple objectives includes a number of elements. For all of these elements a number of choices, such as which optimization method to use, have to be made when implementing the framework. Out of the nine elements listed in Table 2.1 two fall within the scope of this research. These are the scenarios and the metrics given that these two are directly related to the question of which objectives to use. The other six elements are relevant to this research in the sense that they are necessary to perform the calibration. So, for the remaining seven elements the most practical solution will be chosen, taking into account things such as the limited time available for this research and limitations to computing time and power.

1.3 Research questions

In order to structure the research a number of research questions are posed, whereby the main question is defined as follows:

> ***How can a microscopic pedestrian model be calibrated, using a multiple-objective approach, given its stochastic nature and differences in behaviour in different flow situations?***

To assist the answering of this main question a number of sub-questions are posed, namely:

1. *What is the state-of-the art of calibration frameworks in the field of pedestrian flow modelling?*

   A review of the literature should give an overview of the state-of-art of calibration frameworks in the field of pedestrian flow modelling. Furthermore, the review of the literature should also give an overview of which flow situations and which metrics have been used previously in calibrating pedestrian models. Finally, the review should explore the steps that make up the multiple-objective framework.

2. *How can we deal with the stochastic nature of the model and the different behaviours in different flow situations when determining the search space?*

   The answer to this question should give insight into what the most important parameters are that have to be taken into account when determining the search space, how different flow situations influence this and how one should deal with the stochastic nature of the model when doing quantitative analyses.

3. *How can we calibrate a microscopic pedestrian simulation model using the multiple-objective approach given the stochastic nature of the model and differences between the behaviour in different flow situations?*

4. *How do different choices regarding metrics and flow situations in the calibration procedure affect the state estimation of microscopic pedestrian models?*

1.4 Contributions of this research

This research contributes to both science and practice in a number of ways. The primary contributions to both are the increased insights into how different choices regarding metrics and flow situations in the calibration procedure affect the state estimation of microscopic pedestrian models. This provides improved information on how to calibrate a pedestrian flow given its intended usage and how this is reflected in the choice of flow situations and metrics.

It also contributes by giving insight into why it is important to deal with the stochastic nature of pedestrian models whilst calibrating the model and especially what the problems are one faces when trying to do this.

Lastly, it contributes by showing the influence of using different flow situations during the sensitivity analysis and hence the importance of taking the difference in behaviour into account during the sensitivity analysis.

1.5 Thesis outline

Based on the structure provided by the research questions this research is built up as follows. As Figure 1.2 illustrates, the research contains two major steps, the preparation step and the calibration step.

The preparation step contains the first three chapters of this thesis. Chapter 2 provides an overview of the state-of-the art regarding the calibration of pedestrian models. Furthermore, it analyses the elements involved in calibrating a pedestrian model and the possible options per element. Based on the outcomes of the literature review the research methodology is discussed in chapter 3. This chapter introduces, among other things, the reference data and the simulation model used during this research as well as the investigation into the question of how to deal with the stochastic nature of the model. Based on both the outcomes of chapter 2 and chapter 3, a sensitivity analysis is performed in chapter 4 to investigate to which parameters the model is sensitive.

The second step contains the actual calibration of the model and the investigation into how different choices of objectives influence the outcome. Section 5.1 introduces the calibration methodology based on which the simulations are run. Based on the results of the simulations and the reference data the model is calibrated using different combinations of scenarios and metrics. The results of the calibration are discussed in section 5.2 and section 5.3. Section 5.4 discusses how the results of the previous two sections could potentially be affected by the choices made in section 5.1. Based on this discussion and the results the implication for practice are described in section 5.5

Lastly, chapter 6 discusses the main conclusions of this research as well as the limitations of this research. It also provides a number of recommendations for both science and practice.

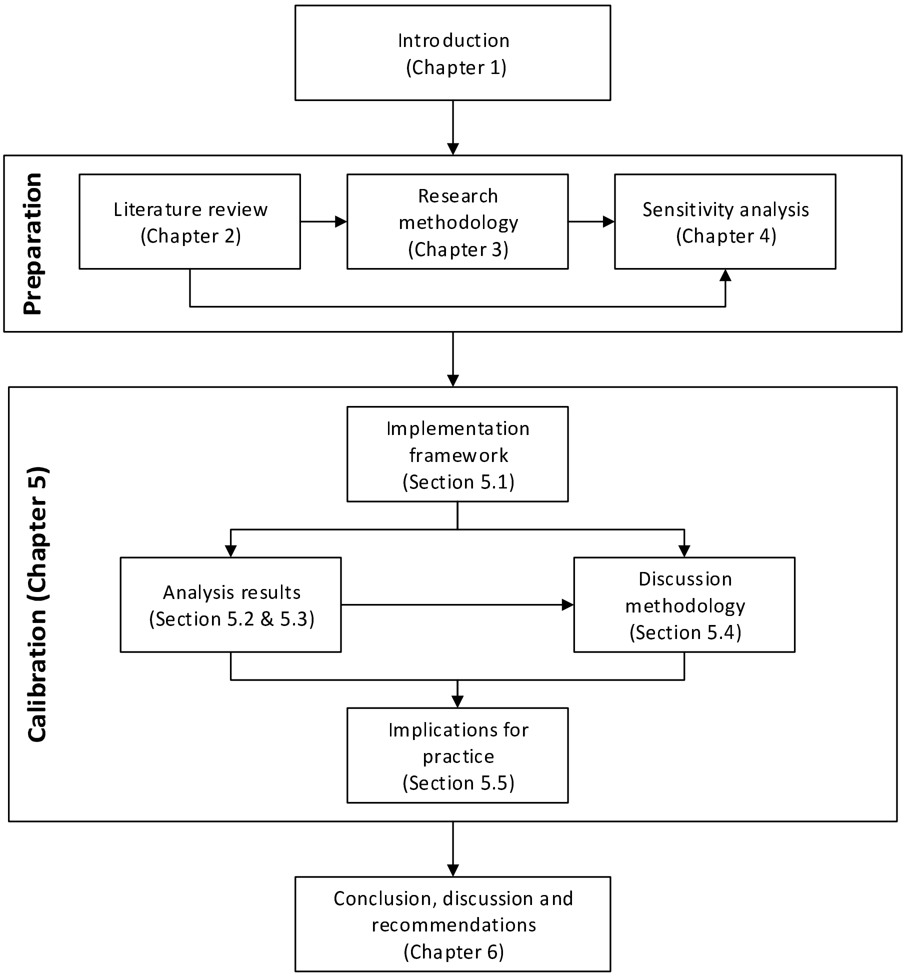


Figure 1.2: Outline of the thesis

2 | State-of-the-art of calibrating pedestrian simulation models

In this chapter a review will be given of the literature regarding the calibration of pedestrian simulation models. The goal of the review is to answer two questions, namely:

1. What is the state-of-the-art regarding the methods for calibrating a pedestrian model? (**Section 2.1**)
2. What steps are involved in calibrating pedestrian models and what methods and solutions do exist for these steps? (**Section 2.2**)

Together, the answers to the questions above will provide a basis for the remainder of this research. For the purpose of clarity, the definition of calibration is given below.

> ***Definition:***
>
> Calibration is the process whereby the parameters of the model are systematically adapted such that the model replicates reality more accurately. One has a model $f(\cdot)$ with a certain input I and a certain parameter set P. For the given parameter set and input the model results in a certain output $O_{sim} = f(I, P)$. The goal of calibration is to find a parameter set P_{opt} such that $O_{sim} \sim O_{real}$ whereby O_{real} is data representing the reality one wants the model to replicate accurately. Or in words, the goal is to find a parameter set that results to the simulated data being as similar as possible to the actual data given a certain level of desired accuracy.

2.1 State-of-the-art of calibration methods for pedestrian models

In this section an overview is given of the state-of-the-art of calibration methods for pedestrian models. The main goal is to answer two questions, namely: 1) What methods have been proposed and used and 2) are there generally accepted methods?

Compared to model development little attention has been given to calibration (Rudloff, Matyus, Seer & Bauer, 2011). This is mainly attributed to the lack of data (Abdelghany, Abdelghany & Mahmassani, 2016; Berrou et al., 2007; Davidich & Köster, 2013; Rudloff, Matyus, Seer & Bauer, 2011) especially at high densities.

Despite this there are a good number of studies where authors do calibrate a pedestrian model (e.g. (Davidich & Köster, 2012; Bauer, 2011; Robin, Antonini, Bierlaire & Cruz, 2009; Tang & Jia, 2011; Klein, Köster & Meister, 2010; Weichen et al., 2014)) usually by using the fundamental diagram (Campanella, Hoogendoorn & Daamen, 2009b). However, as multiple authors mention, the calibration attempts in these studies are limited and mostly focus on only one or a few aspects (Berrou et al., 2007; Campanella et al., 2009b; Davidich & Köster, 2013; Hoogendoorn & Daamen, 2007; Rudloff, Matyus, Seer & Bauer, 2011). Most studies focus on one specific movement base case (e.g. a bidirectional flow in a straight corridor), only use a single metric or do not look at various compositions of the population. Meanwhile, there are countless combinations of movement base cases, metrics and population compositions (Campanella et al., 2009b) and from studies such as (Campanella, Hoogendoorn & Daamen, 2009a) it is clear that these factors do affect the outcome and performance of the model. Hence there is the question of how transferable the models are if they are calibrated using a very limited focus.

Although the large number of possible combinations of movement base cases, metrics and population compositions together with the large number of parameters of the models (Rudloff, Matyus, Seer & Bauer, 2011; Seer, Brändle & Ratti, 2014) is seen as a difficulty (Berrou et al., 2007; Campanella et al., 2009b) there are three frameworks described in the literature which try to take a more inclusive approach (i.e. try to include more than one combination). These three frameworks, (Campanella et al., 2011), (Wolinski et al., 2014) and (Duives, 2016), will be described in more detail below.

All three frameworks can be described as using a multiple-objective approach but there are some differences. Wolinski et al. (2014) focus on using multiple metrics to compare the model results to the reference data whilst Campanella et al. (2011) present a framework that uses multiple movement base cases with multiple metrics. However, during the calibration procedure both used only one metric. The work by Duives (2016) uses multiple movement base cases with multiple metrics and furthermore includes different combinations of weights in the objective function and thus is the most extensive of the three.

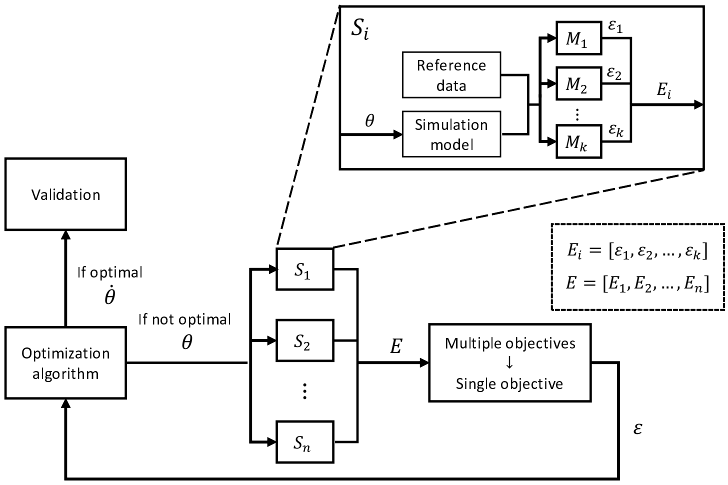


Figure 2.1: Overview of the multiple-objective framework

Generally speaking the frameworks mentioned above can be depicted as follows (see Figure 2.1): Following the multiple-objective approach the calibration framework uses multiple scenarios (S_1 to S_n) whereby every scenario has reference data to which the simulation results are compared using one or more metrics (M_1 to M_k) and their corresponding objective function. This results in k error values (ε 1 to k). The errors of all metrics of all scenarios are combined into a single error value (ε) which in turn is fed into the optimization algorithm which determines, based on stopping criteria, whether or not the current parameter set is deemed optimal and if the next step, validation, can be initiated. In the case that the set is not yet optimal the algorithm creates a new set which in turn is used to run the next set of simulations.

So, it is clear from the literature that a multiple-objective approach should be used when calibrating a pedestrian model, and hence this approach will also be used in this research. However, other than the need for multiple objectives the question of what these objectives should be is largely unanswered. Thus far only two dimensions, movement base cases and metrics, have been investigated. Both Campanella et al. (2011) and Duives (2016) have investigated the effect of using multiple movement base cases during the calibration versus only using one base case. These studies show that pedestrians seem to show different behaviour in different movement base cases from which two conclusion can be drawn. Namely, 1) Calibrating using a single movement base case, compared to using multiple base cases, will result in better performance on that particular base case. However, 2) for general usage (i.e. more than one base case) one does need to calibrate using multiple movement bases to capture all relevant behaviour. The effect of using different metrics during the calibration has only been investigated by Duives (2016). The study shows that different combinations of metrics clearly lead to different calibration results. However, the research also states that more research is necessary into which metrics to use.

So, clearly more research is necessary into what objectives should ideally be used. Besides further investigating the influence of the movement base cases and the metric, other scenario properties (see subsection 2.2.2) such as population composition and density level are of relevance for research.

2.2 Framework elements

In this section the elements which together form the multiple-objective framework for calibrating a pedestrian model will be discussed. Firstly, the elements will be identified after which each elements is discussed in more detail whereby it is discussed what choices have to be made, what the possible options are and which factors are to be taken into account when making the choices.

2.2.1 Element identification

As one can see in Figure 2.1 the framework contains multiple parts and implementing such a framework involves a number of questions that have to be answered. Based upon these questions all elements are identified.

The first question that has to be answered when implementing a multiple-objective framework is which scenarios to use. Answering this question involves three sub-questions, namely:

1. Which scenarios one should use based upon a number of scenario properties such as the movement base case and the population composition.
2. Which metrics should one use for the given scenarios?
3. What objective functions should one use and how should one compare the simulated data and the reference data for the given metrics?

The first three elements involve answering these three questions.

As one can see in the figure, the framework not only contains a set of scenarios but also an optimization algorithm whereby three other elements are involved namely:

1. Choosing an optimization algorithm,
2. Determining the stopping criteria, and
3. Determining the search space.

Lastly there are three elements related to the practical implementation of the framework, namely, how to deal with the stochastic nature of the model, what type of reference data to use and how to define the input of the model such that is closely matches the reference data.

So, nine elements are identified and in Table 2.1 an overview can be found. All elements listed in Table 2.1 will be discussed in more detail below whereby it is explained why the element is relevant for the multiple-objective framework, what the possible options are and what factors will influence the choice.

Table 2.1: The nine identified elements of the multiple-objective framework

1) Scenarios (2.2.2)
2) Metrics (2.2.3)
3) Objective functions and comparison methods (2.2.4)
4) Stochasticities (2.2.5)
5) Input definition (2.2.6)
6) Reference data (2.2.7)
7) Optimization methods (2.2.8)
8) Stopping criteria (2.2.9)
9) Search space definition (2.2.10)

2.2.2 Scenarios

In order to run a simulation of a pedestrian model one has to define a number of things, namely:

- Infrastructure: What is the exact geometry of the walkable space, the location of obstacles and the location of the entry and exit points?

- Demand patterns: Where and when do how many pedestrians with a certain destination enter the simulation?
- Population composition: What properties do the pedestrians have and how are they distributed?

So, these are the first three properties of a scenario and based on these two other properties can be identified, namely:

- Movement base case: The movement base case is a property that follows from the combination of the infrastructure and the demand pattern. For example, the bidirectional straight movement base case is defined by a straight corridor (infrastructure) and a bidirectional flow which is the result of the location of the entries and exits and the demand patterns.
- Density level: This property is closely related to the demand patterns however it describes the resulting/desired density level whilst the demand pattern determines how to reach this density level.

All five properties identified above are discussed in more detail below.

Movement base case

As is explained in (Duives, 2016) a movement base case is a movement case whereby there is only one predominant action performed by the pedestrians. In (Duives, 2016) eight of these base cases have been identified which can be found in Figure 2.2.

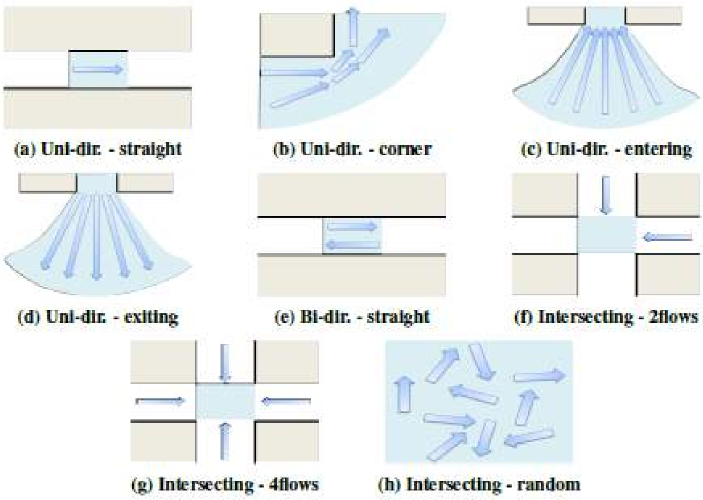


Figure 2.2: The eight movement base cases in pedestrian flow, identified in (Duives, 2016) (from Duives, 2016)

The eight movement base cases shown in Figure 2.2 are considered to "cover the range of pedestrian crowd movement dynamics occurring often during large-scale events" (Duives, 2016, p. 21) hence there might be other base cases when looking outside of the scope of large-scale events. Based on the literature review two cases have been identified which seem to involve a significantly different type of movement compared to the eight base cases such that is seems reasonable to add them to the list of base cases. These cases are: vertical movements such as flows up or down stairs or escalators (e.g. Daamen, 2004) and merging flows (e.g. Shiwakoti, Gong, Shi & Ye, 2015; Zhang, Klingsch, Schadschneider & Seyfried, 2011).

Research by both Campanella et al. (2011) and Duives (2016) has shown that using different base cases during calibration results in different optimal parameter sets and hence both studies show the importance of using multiple bases cases when calibrating a pedestrian model for general usage. This is further supported by the findings in (Campanella et al., 2014) which show that models perform better in a general setting when they are calibrated using multiple base cases.

Geometry of the infrastructure

The geometry of the infrastructure defines the exact dimensions and shape of the walkable space. For example, it defines the width and length of straight corridor, the width and shape of a bottleneck or the angle and width of a corner. Although the geometry is strongly linked to the movement base cases a movement base case is not solely defined by the geometry as, for example, a straight corridor can have either a uni-directional flow or a bidirectional flow.

Research has shown that the geometry can influence the flow but this is not necessarily always the case. For example, Zhang, Klingsch, Schadschneider and Seyfried (2012) found no clear influence of the width of the corridor on bidirectional flows. On the other hand Gorrini, Bandini, Sarvi, Dias and Shiwakoti (2013) found a clear dependence of the flow rate to the angle of a corner.

Density level

The density level describes what densities occur within the scenario and thus which part of the fundamental diagram is covered. The density level tells something about the number of interactions between pedestrians and how these interactions influence the movement of the pedestrians. Research has found that, in general, the speed decreases as the density increases (and hence the number of interactions increases), however, the exact shape of the relationship is still under discussion (Duives, 2016). Furthermore, different levels of density give rise to self-organizing phenomena such as lane formation in bidirectional flows, stripe formation in crossing flows and stop-and-go waves in unidirectional flows. A more detailed overview of self-organizing phenomena in pedestrian flows can be found in (Duives et al., 2013).

So, the level of density clearly affects the flow, however, the question is also does it affect the underlying behaviour and is it thus relevant to calibrate a model based on more than one density level? A review of the literature did not turn up any studies giving an clear answer to these questions except for (Campanella, 2016). In this study it is shown that poorness of data can be a problem when calibrating a pedestrian model. Poorness of data is described as: "the insufficient amount of information about important behaviours such as collision avoidance or following behaviour" (Campanella, 2016, p. 94). So, a model calibrated based solely on density levels where no or very few interactions are present might have a problem with the significance of parameters responsible for the interaction between pedestrians. However, this does not answer the question if this is also the case when one compares a case with few interactions (low density) with a case with many interactions (high density).

Given that is it unclear how the level of density affects the calibration results, other than the case of no interactions versus interactions, and given that it is clear that the density level affects the flow, the level of density is considered a relevant and interesting property of a scenario to investigate in more detail.

Population composition

The composition of the population can influence the flow in different ways. For example, a meta- analysis of Bohannon and Williams Andrews (2011) shows clear differences in walking speeds for different age groups. Research by Chattaraj, Seyfried and Chakroborty (2009) shows that cultural differences also impact the flow and multiple studies (Moussaïd, Perozo, Garnier, Helbing & Theraulaz, 2010; Duives, Daamen & Hoogendoorn, 2014; Gorrini, Bandini & Sarvi, 2014) show this also holds for groups whereby not only their size but also their composition (gender, ages etc.) influences the flow. On top of this studies by Campanella et al. (2009a) and Yang, Daamen, Hoogendoorn, Chen and Dong (2014) show that the heterogeneity of the population clearly influence simulation results.

Given that the composition of the population thus clearly influences the flow it is relevant to use different population compositions when calibrating a pedestrian model for general usage.

Demand pattern

The demand pattern determines where and when which pedestrians appear in the simulated area. Hence the demand pattern is closely related to the density level however it determines also elements like the flow ratios, in the case of multiple flows, and how the inflow is distributed over time.

The distribution of the inflow over time can influence the flow because pedestrian flows have the property that after the capacity is reached the flow will start to decrease when the density increases. So, for example, when the average inflow is close to the capacity and the inflow is uniformly distributed over time the flow will remain near capacity but when it is not uniformly distributed the inflow could temporarily exceed the capacity causing the flow to decrease and hence a bottleneck to form.

Research by Kretz, Grünebohm, Kaufman, Mazur and Schreckenberg (2006) shows that in a bidirectional flow, the flow ratios do seem to influence the flow. However, as they note themselves, the research has its limitations because the participants where mainly in their twenties and only a single corridor width was used. Hence the question how the flow ratio influences a flow with different population compositions, different geometries and different base cases remains.

Conclusions

So, based on the five properties described above it is clear that there are many different possible scenarios which can be used when calibrating and validating a model. The large possible number of scenarios raises the question if all these properties are equally relevant or that some are more relevant than others.

The movement base case and population composition seem most relevant because varying them will answer the question if the model is capable of modelling the different types of interactions well and if the model is able to capture the influence of differences in the population composition well. The density level is also important however primarily to ensure that all relevant self-organizing phenomena can occur and that the data is rich enough to estimate all relevant parameters. The geometry and demand patterns are also relevant however given that they are at a more detailed level they are considered to be less relevant than the other three.

So, it is clear that, when making a choice of which scenarios to use, the relevancy as described above should be an important factor. However, the availability of data will, in the end, decide whether or not it is possible to include different combinations of a certain property.

2.2.3 Metrics

Every scenario has one or more metrics which are used to compare the simulation and the reference data. In Table A.1 a comprehensive overview of metrics, used in evaluating the performance of pedestrian models, can be found. As the table illustrates, these metrics have been categorized based on three properties, namely whether they are: 1) Macroscopic, mesoscopic or microscopic, 2) quantitative or qualitative, and 3) generally usable or linked to a specific set of scenario properties. These categories will be described in more detail below.

Macroscopic, Mesoscopic or Microscopic

This property describes the aggregation level at which we look at the pedestrian flow and when describing pedestrian flows one can identify three of these aggregation levels.

At the **microscopic** level one looks at the behaviour of a single entity which can be either a single pedestrian or a single group of pedestrians. For example, one compares the actual trajectory of a single pedestrian with the simulated trajectory to determine whether or not the model is able to capture the individual behaviour accurately.

At the **mesoscopic** level one looks at the distribution of a single metric whereby the metric is obtained for every single pedestrian. For example, one compares the actual travel time distribution with the simulated travel time distribution. This level can give insight into whether or not the pedestrians behave well on average (e.g. the actual mean travel time and the simulated travel time are similar) but more importantly, it can give insight into how well the heterogeneity is captured by the model (i.e. do the actual travel time distribution and the simulated travel time distribution have similar shapes and variances?).

The **macroscopic** level describes the aggregate and collective behaviour of the pedestrians and also properties of the system resulting from this collective behaviour. Examples of this are the formation of self-organizing structures such as lanes which form as a result of collective behaviour or the capacity of a bottleneck which is a property of the system resulting from this collective behaviour.

As Table A.1 illustrates, macroscopic metrics are most frequently used and especially the three macroscopic flow variables (flow, speed and density) or the relationship between them (the fundamental diagrams). Trajectories, a microscopic metric, are also quite commonly used. Mesoscopic metrics, on the other hand, are less commonly used and if they are used it is primarily in the form of travel time.

Ideally, one would like the model to produce behaviour that is realistic at all three levels. However, the question is if the model captures the behaviour well enough for this to be possible or that the model is simply not capable of producing realistic results at all three levels using a single parameter set. Work by Campanella (2016) has shown that parameter sets solely based on calibration of trajectories (microscopic level) did not result in accurate predictions on the macroscopic level. Furthermore, work by Duives (2016) has shown that using solely macroscopic metrics, compared to using solely mesoscopic metrics, when calibrating a model results in different optimal parameter sets. So, previous research thus shows that using metrics at different levels results in different optimal parameter sets and that a model calibrated using only metrics at one level not necessarily produces accurate results at the other levels. Hence, it is clearly useful to use metrics of different aggregation levels when calibrating a pedestrian model.

However, given that the findings above implicate that, currently, pedestrian simulation models do not seem capable of obtaining accurate results at all three aggregation levels, there is the question of priority. Namely, does one level take priority over the other levels when calibrating a pedestrian model? The answer to this question is will depend on the intended use of the model. For example, when the main goal of the model is to accurately predict the flows the macroscopic level takes priority over the microscopic level whilst if the goal is to produce animations of pedestrian movement the microscopic level might be more relevant. As is stated in the scope (section 1.2), the main goal of the calibration is to improve the model's capability to predict the aggregate data and hence the macroscopic level takes priority over the mesoscopic and the microscopic level. So, when choosing which metrics to use this prioritisation will be leading.

Quantitative or Qualitative

A second property which can be used to categorize the metrics is whether they are quantitative or qualitative. Note that a metric is defined as quantitative if it is expressed numerically and that this is independent of the method used for comparing simulation results with the reference data.

Table A.1 shows that mostly quantitative metrics are used however there are also some occurrences of the use of qualitative metrics. These qualitative metrics are either used at the macroscopic level or at the microscopic level. At the macroscopic level they are clearly used to check whether or not a model is capable of producing certain self-organizing patterns. This is done qualitatively instead of quantitatively because it has been proven difficult to quantitatively describe self-organizing patterns and hence qualitative metrics provide an easier, but less strictly defined, method for determining a models capability to produce these patterns. At the microscopic level they are primarily used for face-validating the model.

Quantitative metrics are preferable over qualitative metrics because they are strictly defined by their mathematical formulas whilst qualitative metrics depend heavily on the judgement made by an individual researcher and hence different researcher might come to different values. Furthermore, using qualitative metrics during the calibration process is not feasible given the large number of iterations. However, qualitative metrics do have their use is this research as section 4.1 will shows.

General or specific

The last property that will be used to categorize the metrics is whether or not they are related to a specific set of scenario properties or whether they can be used regardless of the properties of the scenario. Metrics that can be used regardless of the scenario properties are used to compare how a model performs on different scenarios. Examples of these kinds of metrics are travel time and the fundamental diagram. Metrics that are related to a specific set of scenario properties, for example the bottleneck capacity, are useful to check whether or not the model is able to reproduce the specific behaviour related to that scenario.

From Table A.1 it can be derived that metrics that are generally applicable are used far more often than those that can only be used in specific scenarios. It can also be noted that all qualitative macroscopic metrics are related to a specific scenario which is logical given that these are used to check for self-organizing patterns which are related to specific movement base cases.

Conclusions

The findings above have a number of implications for this research. Firstly, it is clear that preferably metrics from all three levels should be included in every scenario to ensure that the simulated behaviour is accurate at all these three levels. However, given that it is questionable if the model is capable of producing accurate results at all three levels and given the goal of the model, macroscopic metrics take priority over meso -and microscopic metrics and mesoscopic metrics take priority over microscopic metrics. Secondly, during calibrating the model solely quantitative metrics are used as using qualitative metrics is not feasible given the large number of iterations. Thirdly, it is preferable to use general metrics in all scenarios because it makes it possible to easily compare the performance of the model on these different scenarios. However, some specific metrics, such as the bottleneck capacity, might also be very relevant in certain scenarios given that they capture behaviour specific to that scenario.

2.2.4 Objective functions and comparison methods

In the calibration process the function that describes the difference between the simulation and reference data given a certain parameter set is called the objective function whereby the goal is to minimize this function. Mathematically this can be written down as follows:

$$f : \Theta \rightarrow \mathbb{R} \tag{2.1}$$

Whereby Θ is the set of feasible solutions (i.e. the search space) and $\mathbb{R}$ describes the objective function space (Zak & Chong, 2013). Given that the goal is to minimize the difference the problem can be written down as follows:

$$\begin{aligned} \text{minimize} \quad & f(\theta) \\ \text{subject to} \quad & \theta \in \Theta \end{aligned} \tag{2.2}$$

However, the definitions above assume a single objective whilst in this research the focus is on a multiple-objective approach. In the case of multiple objectives Equation 2.1 and Equation 2.2 are transformed to:

$$\mathbf{f} : \Theta \rightarrow \mathbb{R}^n \tag{2.3}$$

and

$$\begin{aligned} \text{minimize} \quad & \mathbf{f}(\theta) \\ \text{subject to} \quad & \theta \in \Theta \end{aligned} \tag{2.4}$$

Whereby $\mathbf{f} = [f(\theta)_1, ..., f(\theta)_n]^{\mathrm{T}}$ and n defines the number of objectives. As one can see Equation 2.3 and Equation 2.4 reduce to respectively Equation 2.1 and Equation 2.2 when $n = 1$.

In order to solve Equation 2.4 one cannot simply use standard optimization methods. Two commonly used approaches to tackle a multiple-objective optimization problem are: 1) The use of optimization algorithms that aim at producing the Pareto optimal solutions. And, 2) Transforming the multiple objective function into a single objective function so standard optimization methods can be used. The literature review found only two cases where a pedestrian model was calibrated using multiple objectives (Campanella et al., 2011; Duives, 2016) and in both cases the multiple-objective function was transformed to a single objective function.

As described in (Toledo & Koutsopoulos, 2004) there are multiple ways to compare simulation and reference data. Besides technical points such as, are the underlying assumptions of a certain test valid and can it thus be used, the main point is whether one wants to test if the model captures the behaviour well *on average* or if it capture the *dynamics* of the behaviour well. For example, comparing the mean travel times of simulated and reference data gives insight into the average behaviour however tells little about the dynamics. On the other hand, comparing the flow over time along a certain line does give insight into whether or not the dynamics are captured well. Toledo and Koutsopoulos (2004) describe three different approaches which can be used, namely 1) Goodness-of-Fit measures, 2) Hypothesis testing and Confidence intervals, and 3) Test of Underlying Structure. The Goodness-of-Fit measures are useful to get insight into the overall performance of the model. Examples of these measures are the Mean-Squared Error (MSE) and the Root-Mean-Squared Error (RMSE). Hypothesis testing is used to test whether or not two distributions are equal given a certain confidence interval and given certain assumptions. With these kinds of tests it is important to check whether these assumptions, such as the assumption of Independent Identical Distributions (IID) when using a two-sample t-test, hold and thus if the method can be used. The last approach is testing the underlying structure by using metamodels and checking whether or not the metamodels obtained from the simulated data are equal to those obtained from the reference data. This can be used, for example, to test certain relationships such as those described by the fundamental diagram.

In the only two examples of using a multiple-objective approach to calibrate a pedestrian model (Duives, 2016; Campanella, 2016) Goodness-of-Fit measures were used whereby the MSE was minimized. The research by Duives (2016) was the only study which commented on the effect of the choice of objective function whereby it was found that the choice of the objective function and the method for combining the multiple-objective functions into a single objective function severely influence the calibration results.

So, in a multiple-objective framework it has to be decided, for every metric, which of the comparison approaches described above has to be used and the exact mathematical form of the objective function. Furthermore, it has to be decided how to deal with the results of two or more objective functions. This has to be done in conjunction with the choice of the optimization algorithm whereby, broadly speaking, the two options are using multiple objectives with an optimization function capable of finding the Pareto optimal solutions or combining the objectives into a single objective and using conventional optimization algorithms.

2.2.5 Stochasticities

Pedestrian models are often stochastic by nature (Duives, 2016). For example, they use distributions in the input, as is common for the preferred speed, or when determining the acceleration of a pedestrian they use a fluctuation term (e.g. see the social force model (Helbing & Molnar, 1995)). This stochastic nature means that two simulations using the same input and parameter set but different seeds can give different results. In order to deal with this multiple simulations have to be run using different seeds, the result of which is a distribution representing the influence of these stochasticities on the simulation results. However, the question is how many runs are necessary to ensure that the resulting distribution approximates the actual distribution (the one one would obtain when running an infinite amount of simulation with different seeds) well enough.

In the literature three approaches were found which can be used to tackle this problem. The first one is a two-step approach presented in (Toledo & Koutsopoulos, 2004). In this case the minimum number of replications is determined by:

$$R_i = \left(\frac{S_{R_0}(Y_i)t_{\alpha/2}}{d_i}\right)^2 \tag{2.5}$$

where

$S_{R_0}(Y_i)$	=	Sample standard deviation of simulation outputs based on R_0 replications
$t_{\alpha/2}$	=	Critical value of the t-distribution at significance level α
d_i	=	Allowable error

This method makes two assumptions, namely that the outputs (Y_i) from different replications are normally distributed and secondly that the standard deviation does not change significantly when one were to do more replication than R_0.

The other two methods are both sequential methods whereby the first, also presented in (Toledo & Koutsopoulos, 2004), is effectively the sequential approach of the first method. The stopping criterion is given by:

$$R \geq R_i = \left(\frac{S_R(Y_i)t_{\alpha/2}}{d_i}\right)^2 \tag{2.6}$$

where R is the current number of replications. In (Ronchi, Kuligowski, Reneke, Peacock & Nilsson, 2013) another sequential method is proposed that is based on multiple convergence measures. They see the results of the runs for a given metric as a series converging to an expected value (i.e. the expected value of the specified metric). If one takes $M_{i,j}$ as the result for the j^{th} run of metric i and $\overrightarrow{M}_{i,j}$ as the series of j results of metric i (e.g. $\overrightarrow{M}_{i,3} = (M_{i,1}, M_{i,2}, M_{i,3})$) then the expected value of $\overrightarrow{M}_{i,j}$ is given by $E(\overrightarrow{M}_{i,j})$. The stopping criterion in (Ronchi et al., 2013) is that, for all metrics, the difference between two consecutive expected values is below a certain threshold for n consecutive runs (i.e. $\left|\frac{E(\overrightarrow{M}_{i,j})-E(\overrightarrow{M}_{i,j-1})}{E(\overrightarrow{M}_{i,j})}\right| \leq T_i$ for n consecutive runs).

The difference between the three methods is two-fold. Firstly, there is a clear difference between a two-step and a sequential approach, whereby the two-step approach has the advantage that the amount of replications is determined before the calibration procedure and hence one can make a much better estimation of the required computational time for both procedures. However, one also has to assure that the number of replications is determined based on a representative selection of scenarios, metrics and parameter sets and thus one has to decide beforehand what this representative set looks like and the larger this set the larger the computational time it requires to determine the number of replications. Secondly, there is a difference between the two sequential approaches primarily in the choices that have to be made. In case of the method in (Toledo & Koutsopoulos, 2004) one has to determine the allowable error (d_i) for every metric whilst in the case of the method in (Ronchi et al., 2013) one has to determine the threshold value for every metric. Furthermore, in the case of the method in (Toledo & Koutsopoulos, 2004) one has to ask the question of how well the calculated standard deviation approximates the actual standard deviation when the number of replications is small and in the method of (Ronchi et al., 2013) one has to determine the value of n.

So, based on the difference describe above, a choice has to made between using a two-step approach and a sequential approach and in the case of a sequential approach a choice has to be made between the two possible sequential approaches. Furthermore, it has to be decided which metric or metrics are to be used and what the values should be for the method's parameters (e.g. the allowable error or the number of consecutive replications).

2.2.6 Input definition

This step involves defining the input to the simulation such that it matches the reference data well. As is stated in (Toledo & Koutsopoulos, 2004) ideally one would want to feed the simulation with exactly the same inputs as is the case in the reference data. It is important to do this well given that this is a potential source of discrepancy between the simulation and the reference data and hence could cause a wrongful estimation of the parameters. Below a number of examples is given of potential inputs used when calibrating a pedestrian model with regards to the operational behaviour.

- Geometry of the infrastructure: Besides using the right shape and dimensions defining the infrastructure also involves defining the borders of the walkable area and how pedestrians interact with these borders. For example, how to define a border when in the data the border is a line on the ground people are instructed not to cross and in the model a border is defined as a wall? More examples of this can also be found in (Duives, 2016)
- Speed distribution: Depending on the composition of the population in the reference data the distribution of the free speeds might differ significantly from those average values from the literature (see for example table 3.4 in (Daamen, 2004)). So, this begs the question how closely should this distribution match the reference data and how to estimate this in the case there is no free-flow?
- Demand pattern: The demand pattern defines where and when pedestrians enter the simulation. The question is, should these locations and times be replicated exactly or can they also be approximated using distributions?
- Route choice: In the case of simple scenarios this defines what the destination is, in the case of multiple possible destinations, and the exact location of this destination. This is strongly related to the demand patterns because when using distributions in the demand patterns it is logical to do the same for the destination choice and location whilst when defining the demand patterns more strictly it is also logical to define the destination more strictly. In both cases it boils down to the question whether or not one want the model the replicate the reference data exactly or replicate it on average.

From the list presented above, it is clear that, besides determining some metrics related to performance of the model some other data has to be extracted from the reference data to ensure the input of the simulation model matches the references data as closely as possible.

2.2.7 Reference data

Every scenario has reference data to which the simulation results are compared. Needless to say the reference data should have the same properties as the scenario, however, there is also a property specific to the reference data namely the source. The data can be either obtained from literature or the data can be collected through a controlled experiment or in real-life.

The review of the literature showed that data obtained from controlled experiments and real-life are used far more often than those taken from the literature. The examples whereby literature values were used all used a smoothed fundamental diagram. This has the disadvantage that due to the smoothing of the fundamental diagram it can only be confirmed whether or not the model reproduces the diagram on average and nothing can be said about the variance.

In the case of data from controlled experiments or real-life, the data is most often obtained via cameras from which trajectories can be extracted. Controlled experiments have the advantage that one has control over many factors influencing the flow and that information about things such as the population composition can be easily obtained. However, the data from real-life has the advantage that one knows that one is capturing the actual behaviour of pedestrians whilst during experiments there is always the question of how much the behaviour of the participants is influenced by the experiment itself.

So, the main difference between the three types of reference data described above is the level of detail. The values from literature seem to be the least detailed whilst trajectories extracted from data provide far more detail. For calibration using a multiple-objective approach a higher level of detail seems preferable given that based on the trajectories one can extract metrics at all three aggregation levels.

2.2.8 Optimization methods

The optimization algorithm is responsible for two things. Firstly it determines whether or not the current parameter set is optimal and secondly if this is not the case it defines a new parameter set which is tested. Two properties of the optimization algorithm are of relevance, namely: 1) The ability and likelihood to find the global optimum, and 2) the computational burden. The first one is important because the search space might contain multiple local minima and if an algorithm can get stuck in such a local minimum the result of such an algorithm might be a sub-optimal solution. The second property is relevant because an inefficient algorithm might take too long to find the solution to be used in practice.

In the literature a multitude of methods were found, all of them using a single objective function. An overview of these methods can be found in Table 2.2. Wolinski et al. (2014) also combined a number of these methods, namely: Genetic Algorithm (GA) + Greedy, GA + Simulated Annealing (SA), Covariance Matrix Adaptation (CMA) + Greedy and CMA + SA.

Table 2.2: Overview of optimization methods found in the literature

Method	Studies
fmincon (Matlab)	Hoogendoorn and Daamen (2007)
Nelder-Mead	Rudloff, Matyus and Seer (2014)
Genetic Algorithm (GA)	Rudloff, Matyus and Seer (2014), Wolinski et al. (2014)
Greedy	Wolinski et al. (2014)
Grid search	Duives (2016)
Simulated Annealing (SA)	Wolinski et al. (2014)
Covariance Matrix Adaptation (CMA)	Wolinski et al. (2014)
GA + Simplex	Campanella (2016)
Box's Complex Algorithm (from vehicular research)	Toledo, Ben-Akiva, Darda, Jha and Koutsopoulos (2004)

In principle all these methods can find the global optimum, however, the greedy method is more likely to get stuck in a local minimum than the other methods and the grid search is the only exact method and thus the only one that will certainly find the global minimum. On the other hand the grid search method is likely to be the slowest and the other methods will outperform it with regards to the speed.

Two of the studies found did compare the performance of the various methods described above. Rudloff, Matyus and Seer (2014) compared the Nelder-Mead and GA methods. They concluded that, when using trajectories, the Nelder-Mead approach was not recommended given that it resulted in bad estimates unless the errors in the data were very small which is not to be expected when using real data sets. They also found that, although the Nelder-Mead method was much faster than the GA method, it was more likely to get stuck in local minima and thus should be used with caution. Wolinski et al. (2014) compared four methods plus four combinations of methods with regards to both the optimality of the results and the speed. The general pattern of their comparison is that the better the method performs regarding the optimality of the solution the slower it is.

Because the goal of this research does not include finding the best/optimal optimization method the choice of method will be based on what is the most practical method. What is practical is determined by how easy the method is to implement and how likely it is to find the optimal solution within the limited time available within this research.

2.2.9 Stopping criteria

The stopping criteria define whether or not a found solution is optimal. None of the studies found explicitly define their stopping criteria, however some can be derived based on the described method. The grid search method in (Duives, 2016) implies that the stopping criterion is simply that all possible parameter sets have been explored and hence the whole search space has been explored according to a predefined resolution. Most methods described in (Rudloff et al., 2014) and (Wolinski et al., 2014) seem to use a convergence criterion whereby a solution is deemed optimal if within n subsequent iterations no new optimal solution is found. Other possibilities are, stopping when the solution satisfies some minimum criterion (i.e. the error is smaller than a certain value or the goodness-of-fit is larger than a certain value.), stopping after a fixed number of iterations or after a certain maximum computation time and stopping after manual inspection.

The choice of the stopping criteria will depend on the chosen optimization method.

2.2.10 Search space definition

The search space defines the set of possible parameter sets that can be used to calibrate the model. The size of the search space is determined by the number of parameters and the range of these parameters. The larger the number of parameters and their ranges the larger the search- space will be. The size of the search space is one of the factors that will determine the computational burden of the calibration procedure together with the complexity of the search- space and the chosen optimization algorithm. In the literature the search space is seldomly defined with the exception of (Duives, 2016). In order to determine which parameters to include and what range to choose for each parameter a sensitivity analysis can be performed. Firstly, this will give insight into what the most sensitive parameters are whereby the logic is, the more sensitive the simulation results are to a small change in a parameter the more important it is to ensure a good estimate of the value of the parameter. Secondly, it can give insight within what range of a parameter the model produces feasible results. Of course it is of great importance that the sensitivity analysis covers multiple scenario and metrics to ensure that the results are representative.

2.3 Conclusions

In this chapter a review of the literature has been performed with the goal of getting an overview of the state-of-the-art of calibrating pedestrian models and to identify the steps involved in calibrating pedestrian models.

The review of the state-of-the-art showed that calibration has received relatively little attention in the field of pedestrian modelling. Most studies that did perform a calibration did so using a very limited focus (e.g. a single movement base case, a single metric etc.). Meanwhile, other research has shown that different choices regarding, among other things, movement base cases and metrics, lead to different optimal parameter sets and hence that the transferability of models, which have been calibrated using a limited focus, is questionable. The solution proposed in the literature is to use multiple objectives when calibrating pedestrian models. However, only a few studies investigated the use of a multiple-objective framework and hence there are still questions about what the effects of different choices of scenarios, metrics and objective functions are on the results of the calibration.

Based on the description of the multiple-objective framework and Figure 2.1 nine steps, involved in calibrating a model, were identified and these nine steps can be found in Table 2.1. In all of these steps choices have to be made whereby in case of the steps regarding the stochasticities, the input definition, the reference data, the optimization method, the stopping criteria and the search space the most practical solution will be chosen given that these steps are not within the scope of this research.

3 | Research methodology

The literature review in the previous chapter answers the first of the sub-questions of this research. This chapter will discuss how the remainder if this research is set up and how this chapter and the following two chapters will answer the remaining sub-questions. Taken together, the four sub-question will be used to answer the main research question in the last chapter.

To answer the three remaining sub-questions three steps are performed. Firstly, a sensitivity analysis will be performed in order to gain insight into the model's sensitivity to the different parameters. As subsection 2.2.10 explains this information will be used, among other things, to determine the search space. Secondly, the calibration framework will be implemented such that, in the third step, the model can be calibrated using different combinations of objectives to explore how the use of different objectives affects the calibration results.

As the sub-questions imply, these steps have a number of things in common. Firstly, the question of how to deal with the stochastic nature of the model. Secondly, how to deal with the differences in behaviour in the different flow situations. And thirdly, the model which is calibrated and hence whose sensitivities to its parameters will be investigated during the sensitivity analysis.

Figure 3.1 shows that those elements that both the sensitivity analysis and the calibration have in common are discussed in this chapter. Firstly, in section 3.1 the reference data, available to this research will be discussed. The reference data will, together with the conclusions of subsection 2.2.2, determine which scenarios (the model implementations of the flow situations) can and will be used (section 3.2) during both the sensitivity analysis and the calibrations. The reference data is, of course, also used during the calibration to determine how well the simulation results reproduce reality. Section 3.3 will introduce the simulation model and the parameters of the model which are of relevance for this research. Lastly, section 3.4 will discuss the method which will be used to deal with the stochastic nature of the model.

In both chapter 4 and chapter 5, the methodologies for respectively the sensitivity analysis and the calibration will be discussed in more detail.

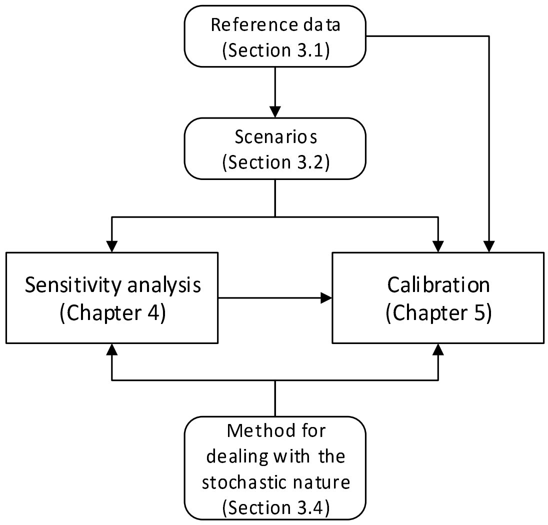


Figure 3.1: Overview research steps

3.1 Reference data

In order to calibrate a pedestrian model one needs data. In this research two sources of data were available. Namely, data from the Transport & Planning (T&P) department of the Delft University of Technology and data from the Hermes project. In this section both data sets will be discussed in more detail based on which a conclusion will be drawn as to which data is used in the remainder of this research.

3.1.1 Data from the T&P department of the Delft University of Technology

The data, available from the T&P department of the Delft University of Technology, was collected during experiments performed in 2002 at the university. During multiple experiment pedestrians were tracked using cameras and from the resulting videos trajectories were extracted. For more details about the experimental setup the reader is referred to (Daamen, 2004, Section 4.4).

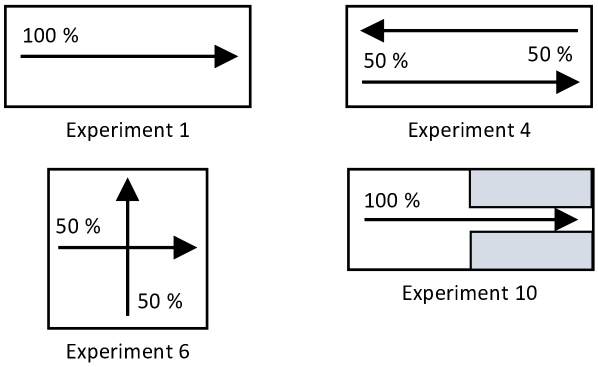


Figure 3.2: Overview of the available experiments of the TU Delft data set(adapted from (Daamen, 2004))

In total, data from four of the ten experiments (see Table 4.3 in (Daamen, 2004)) was available. In Table 3.1 and Figure 3.2 one can find an overview of these four experiments. The table includes the experiment number (according to Table 4.3 in (Daamen, 2004)), the movement base case and a description of the geometry of the walkable area and the inflows. In all four experiments the pedestrians were instructed to walk at normal speeds and in all four cases the population consisted of a mix of children, students, adults and seniors. Furthermore, in the first three experiments (1,4 and 6) there were no physical borders but the participants were instructed to stay within the walkable area indicated by markings on the floor.

3.1.2 Data from the HERMES project

The data from the HERMES project[1] was collected during experiments performed in Germany in 2009. The set contains data from numerous experiments and is publicly available here[2]. Comparable to the data collected at the TU Delft, the movement of the pedestrians was captured using cameras and from the collected images the trajectories were extracted. For a more detailed overview of the experimental setup the reader is referred to (Keip & Ries, 2009).

The participants of the experiments were mostly students with an average age of 25 ± 5.7 years (Zhang, Klingsch, Schadschneider & Seyfried, 2012). In total, the experiments covered six of the movement base case listed in subsection 2.2.2. An overview of these experiments can be found in Table 3.2 and Figure 3.3. Comparable to the TU-delft experiments, participants were instructed to walk normally.

3.1.3 Conclusion on the choice of reference data

Based on the descriptions of the data sets in the two previous subsections and the findings in subsection 2.2.2, it will be determined if both data sets can be used within this research or that only one of the two will be used.

In subsection 2.2.2, the movement base case, the population composition and the density level were deemed the most relevant scenario properties for this research. Together, the two data sets have two different population compositions which would make it possible to vary this during the calibration. However, only two experiments (straight unidirectional and bidirectional flows) are somewhat comparable, and would therefore be candidates for a comparison, and even these two differ in the sense that the geometry of the infrastructure differed and the protocols of the experiments differed. Therefore it would be difficult to assess, in case differences are found, how much of this difference can be contributed to the difference in population composition and how much to difference in infrastructure and experimental protocol. So, for this reason the population composition won't be taken into account in this research and hence only one of the two data sets will be used. Given that the data set from the Hermes project contains more movement base cases, more density levels and different geometries of the infrastructure, it is the data set that will be used in this research.

[1] http://www.fz-juelich.de/ias/jsc/EN/Research/ModellingSimulation/CivilSecurityTraffic/Projects/Hermes/_node.html
[2] http://www.fz-juelich.de/ias/jsc/EN/Research/ModellingSimulation/CivilSecurityTraffic/PedestrianDynamics/Activities/database/databaseNode.html

Table 3.1: Overview of the available data TU Delft

Experiment nr.	Movement base case	Description
1	Unidirectional straight	Straight corridor of 10x4 meters with a unidirectional flow and an inflow of 0.5 ped/s/m
4	Bidirectional straight	Straight corridor of 10x4 meters with a bidirectional flow with equal inflows on both sides and inflows of 0.3 ped/s/m/direction
6	Two-way crossing	Borderless area of 8x8 meters with two flows crossing at an angle of 90 degrees and equal inflows on both sides and inflows of 0.1875 ped/s/m/direction
10	Unidirectional entering	A long narrow bottleneck whereby the walkable area changes from a width of 4 meters to a width of 1 meter and the flow is unidirectional and an inflow of 0.375 ped/s/m

Table 3.2: Overview of the available data Hermes project

Experiment name	Movement base case	Description
Unidirectional - closed	Unidirectional straight	Two straight walkways with a length of 6 meters connected by two semi-circular walkways with walls on both sides of the walkway. In total 24 experiment were performed with different combinations of walkway widths and amount of pedestrians within the infrastructure.
Unidirectional - open	Unidirectional straight	An 8 meter long straight corridor with walls on both sides. 27 experiments were performed with different widths of the walkway, different inflows and different widths of the exit of the corridor.
Short bottleneck	Unidirectional entering and exiting	A 1 meter long bottleneck with about 400 participants standing in front of it whereby the density is about 3 ped/m2. Five different experiments were performed with five different widths of the bottleneck.
Bidirectional	Bidirectional straight	An 8 meter long straight corridor with walls on both sides. In total 22 experiments were performed with varying corridor widths, varying inflows, asymmetric inflows and forced destinations whereby the participants were instructed to leave the corridor on a particular side.
T-junction	Unidirectional merging	A t-junction with two unidirectional flows merging at the junction and continuing as a single unidirectional flow. Seven different experiments were performed using two different walkway widths and different inflows.
Corner	Unidirectional corner	A unidirectional flow around a corner whereby a right-handed turn of 90 degrees needs to be made. 10 experiments were performed with two different walkway widths and five different inflows.

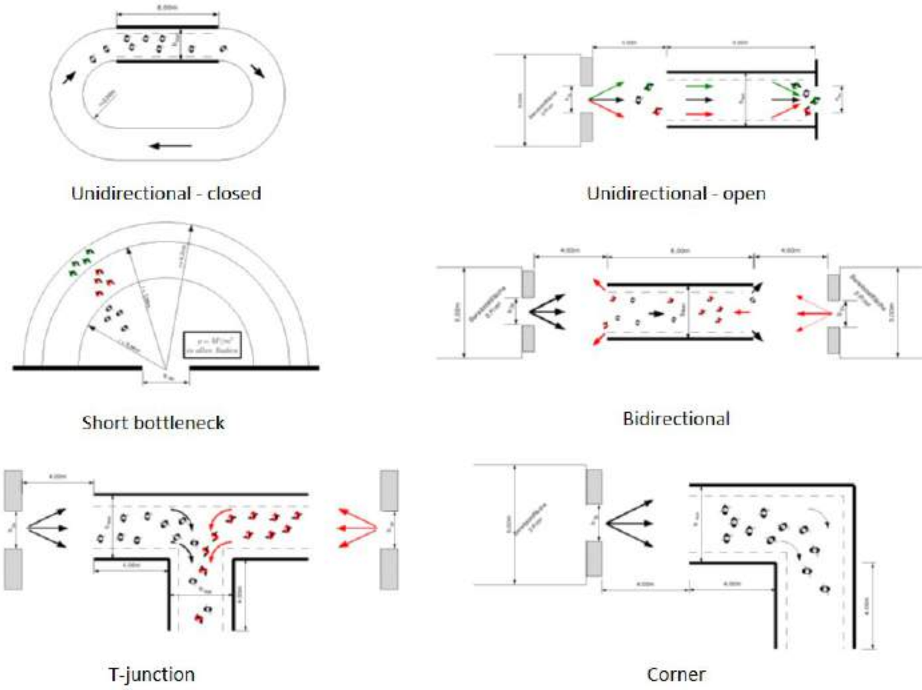


Figure 3.3: Overview experiments Hermes (adapted from (Keip & Ries, 2009))

3.2 Scenarios

In this section it will be explained which scenarios will be used during the sensitivity analysis and the calibration and why these are the scenarios that will be used. The first part of this section will discuss which scenarios will be used based on the five properties described in subsection 2.2.2 and the availability of data. After a choice is made, the second part of the section will describe the chosen scenarios in more detail.

3.2.1 Choice of scenarios

In subsection 2.2.2 five properties were identified based on which a scenario is defined. These five are the infrastructure, the demand patterns, the population composition, the movement base case and the density level. Ideally, one would like to research how varying these five properties influences the calibration results in order to gain more insight into what objectives are necessary. However, constraints in time and data availability make it impossible to do this within this research.

Based on the review of the literature, subsection 2.2.2 concluded that the population composition, the movement base case and the density level are the most relevant properties. Subsection 2.2.2 also concluded that the infrastructure and the demand patterns are of a lesser relevance and hence these are not considered in this research. The population composition will also not be considered in this research given that section 3.1 concluded that there is not enough data to properly investigate the effect of different population compositions on the calibration results. So, this leaves the movement base cases and the density level as properties to be researched. The remainder of this section will explain which movement base cases and density levels will be included and which combinations are used to form the scenarios.

Table 3.3 shows that the data set from the Hermes project covers six of the ten movement base cases. Out of these six, the unidirectional straight movement base case will not be used during the calibration for the following reason. The main interaction in this movement base case is the front-to-back interaction and this interaction is present in all other movement base cases. Furthermore, as research by Campanella (2016) shows, the walking behaviour found in non-congested bidirectional flows seems transferable to unidirectional flows. Out of the five remaining movement base cases, two cases, the entering and exiting cases, are captured in one experiment. Hence, the choice is made to capture these two base cases in the same experiment. So, data from four different experiments will be used and subsequently the scenarios will contain four different infrastructures.

Table 3.3: Coverage of the movement base cases by the reference data

Covered	Not covered
Unidirectional straight	Intersecting flow: 2 flows
Unidirectional corner	Intersecting flow: 4 flows
Unidirectional entering	Intersecting flow: random
Unidirectional exiting	Vertical movements
Bidirectional straight	
Merging flows	

As subsection 2.2.2 concludes, the density level has a strong relation to the number of interactions between pedestrians and that the density level clearly affects the flow. However, given that only the case of no interactions versus interactions has been investigated (Campanella, 2016), the question if using high levels of density (many interactions) versus low levels of density (few interaction) affects the calibration results is yet unanswered. So, in this research the comparison will be made between a low level of density and a high level of density. In the case of the short bottleneck experiment only a high density level will be used given that at low densities the flow would become a unidirectional straight flow which would not capture the behaviour of interest (entering and exiting).

So, as Table 3.4 shows, seven scenarios will be used in this research. The implementation of all these seven scenarios will be discussed in more detail in the next subsection.

Table 3.4: Overview of the scenarios used in this research

Movement base case(s)	**Density level**		**Reference data**
	Low	**High**	
Bidirectional straight	x	x	Bidirectional straight
Unidirectional entering and exiting		x	Short bottleneck
Unidirectional corner	x	x	Corner
Merging unidirectional flows	x	x	T-junction

3.2.2 Scenario definitions

All the seven chosen scenarios will be discussed in more detail below whereby the following aspects will be described:

- The geometry and lay-out of the infrastructure
- The route(s) of the pedestrians
- The measurement area
- The reference data set that is used to compare the results to during the calibration

The measurement area determines the area that is of interest during collection of metrics for both the sensitivity analysis and the calibration. The measurement area is used for two reasons. Firstly, to ensure that only the behaviour representative for the given scenario is captured. For example, in the bidirectional scenario one is interested in the behaviour within a bidirectional flow and one is not interested in how pedestrian enter or exit the experiment or the simulation. Secondly, in the case of the reference data, the cameras did not always capture the whole walkable area of the experiment which in turn limits the area in which metrics can be collected.

Per scenario it will be defined which of the data sets described in (Keip & Ries, 2009) is used to compare the simulation results against. To get insight into the traffic states we want to reproduce during the calibration, the traffic states in reference data will be described using cumulative curves and fundamental diagrams. The cumulative curves are obtained at a certain flow line, the location of which will be described for every scenario. The fundamental diagrams are obtained using the XT-method (Duives, Daamen & Hoogendoorn, 2015) to calculate the density and the instantaneous speeds. The fundamental diagrams only contain data obtained in the measurement areas. Furthermore, for all scenarios, except the bottleneck scenario, measurements obtained the first few seconds and the last few seconds are omitted. This is done because, as the cumulative curves will show, all experiments, except the bottleneck scenario, include a warming up and cooling down period during which the traffic states are not representative for the given scenario. The traffic states are not considered representative because the pedestrians do not experience the same density level as during the rest of the experiment. Hence, is can not be assured that their behaviour is representative for the scenario.

Note: The implementations described below are the implementations as they are used during the calibration. The implementations for the sensitivity analysis are very similar but can differ slightly. This is because after the sensitivity analysis was performed the implementations were critically reviewed to ensure that they matched the reference data as closely as possible. This did lead to some minor changes. In the experimental setup section of the sensitivity analysis these changes will be discussed.

Bidirectional straight - low & high density

The bidirectional straight low and high density scenarios are both based on a bidirectional flow through a straight corridor. Figure 3.4 depicts the lay-out of the scenarios. As the figure illustrates, the area of interest is a corridor section of 8 meters long and 3.6 meters wide bounded by walls on both sides. Pedestrians flow in from both sides, walk through the corridor and at the end of the corridor leave the corridor either at the top or the bottom. During the experiments, the participant were instructed beforehand on which side of the corridor (top or bottom) they should exit.

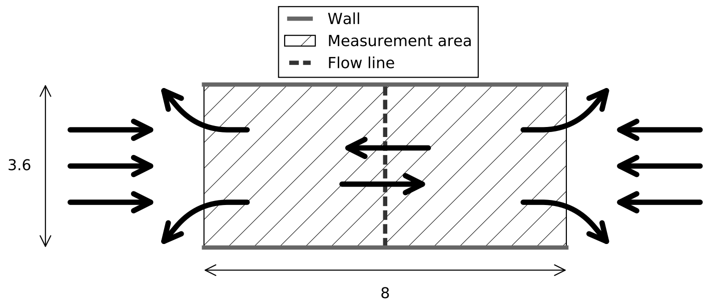
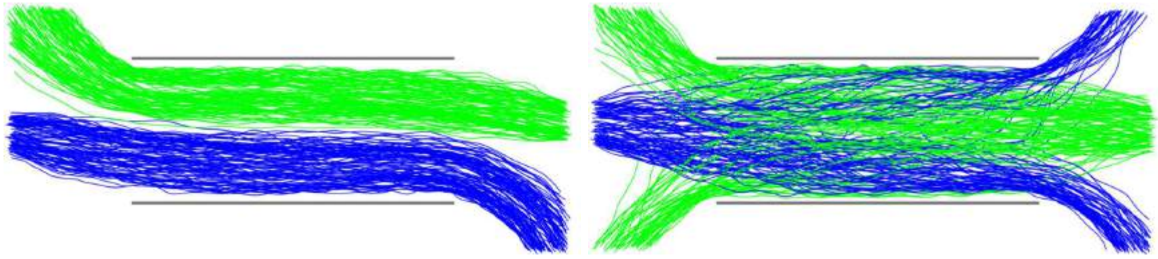


Figure 3.4: Overview of the bidirectional flow scenario

The data also contains experiments where participants were free to choose the side of the corridor they would exit. However, these experiments are not used because, as Figure 3.5 illustrates, this results in two large lanes. Because of these two large lanes most of the interactions between pedestrian is front-to-back whilst the main goal of using a bidirectional scenario is to calibrate for the front-to-front interactions. As the figure also illustrates, the front-to-front interaction occurs far more often in the experiments where pedestrians are forced to exit the corridor at a certain side.

During the calibration, data from the BOT-360-075-075 experiment will be used for the low density scenario. For the high density scenario data from the BOT-360-200-200 experiment will be used. Figure 3.6 illustrates the cumulative curves obtained at the flow line in the centre of the corridor (see Figure 3.4). The curves clearly show that at the start of the experiment the infrastructure is still empty and that it takes a few seconds before it is populated. The curves also show that in the case of the low

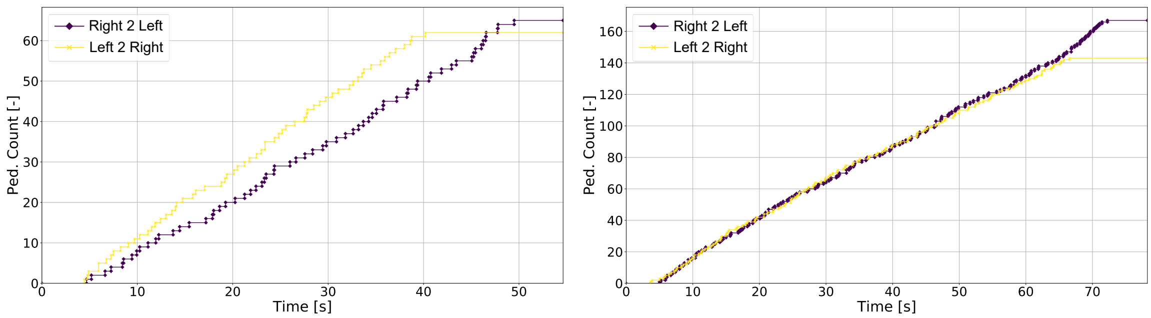


(a) Trajectories resulting from a free destination choice (b) Trajectories resulting from a forced destination choice

Figure 3.5: Difference in trajectories between the free and forced destination choice in the reference data for a bidirectional flow

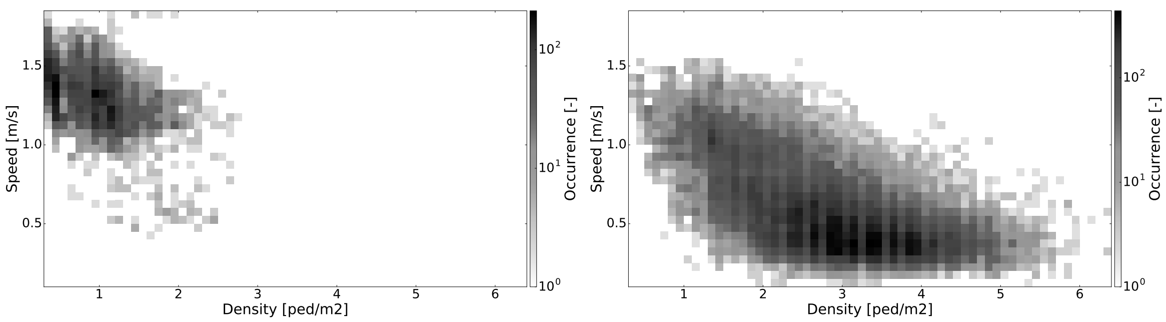
density experiment the flow from left to right is clearly somewhat higher than the flow from the right to the left. This can be explained by the fact that the inflows were not exactly equal (0.49 ped/s/m versus 0.40 ped/s/m). The flows in the high density are very similar although here the inflows are also not exactly equal (0.95 ped/s/m for the flow from left to right versus 0.91 ped/s/m for the flow from right to left).

Figs. 3.7a and 3.7b show the speed-density relationships for both scenarios. The two scenarios clearly cover different parts of the fundamental diagram. The low density scenario covers an area with high speeds and low densities whilst in the case of the high density scenario the speeds are primarily low and the densities high. Furthermore, the states obtained in the high density scenario clearly vary a lot more indicating that during the experiment the state varied significantly over time and space.



(a) Low density (BOT-360-075-075) (b) High density (BOT-360-200-200)

Figure 3.6: Cumulative curves bidirectional scenario



(a) Low density (BOT-360-075-075) (b) High density (BOT-360-200-200)

Figure 3.7: Fundamental diagrams (ku) bidirectional scenario

Bottleneck, short

The short bottleneck scenario covers two different movement base cases, namely, the unidirectional entering case and the unidirectional exiting case. As Figure 3.8 illustrates, the scenario contains a start area in front of the bottleneck which, at the start of the simulation, is filled with a density of 3 ped/m^2. As soon as the simulations starts the pedestrians will start moving through the bottleneck towards the exit. The bottleneck itself has a width of 3.6 meters and a length of 1 meter. After the bottleneck the walkable area widens and pedestrians can spread out whilst walking towards the exit. The measurement area contains part of the area before the bottleneck, the bottleneck itself and a part of the area behind the bottleneck. The measurement area only captures parts of the areas before and after the bottleneck because the cameras did only capture these parts of the areas during the experiments.

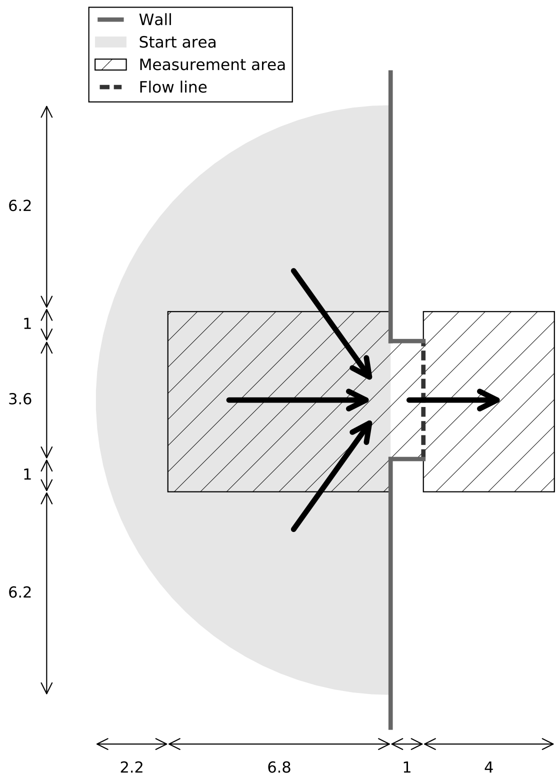


Figure 3.8: Overview of the bottleneck scenario

During the calibration, data from the AO-360-400 experiment will be used for comparison with the simulation results. Figure 3.10 clearly illustrates the differences between the traffic states before, in and after the bottleneck. Before the bottleneck the speeds are low, primarily below 0.5 m/s, and the densities high, mainly between 4 and 6 ped/m^2. In the bottleneck itself the speeds increase and the density decreases. After the bottleneck the speeds increase even more and the densities, again, decrease. As Figure 3.9 shows, once the first pedestrians pass the bottleneck there is a steady flow of about 2.5 ped/s/m in the bottleneck. As the last pedestrians pass through the bottleneck the flow decreases. In total, 349 pedestrian pass the bottleneck.

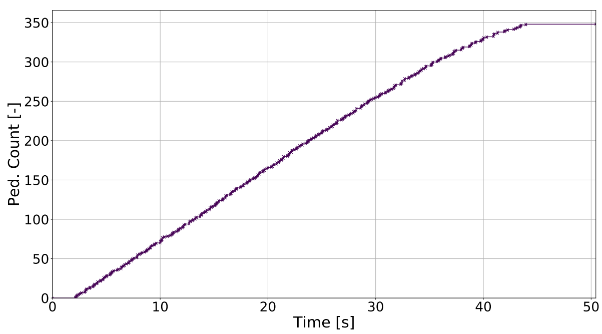
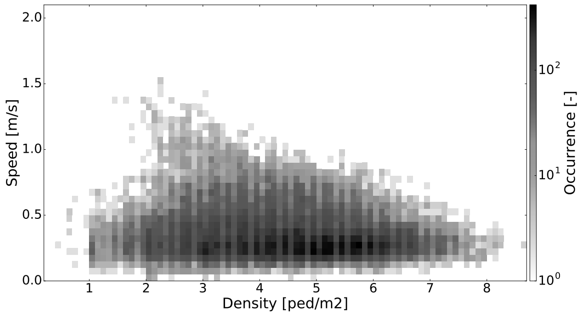
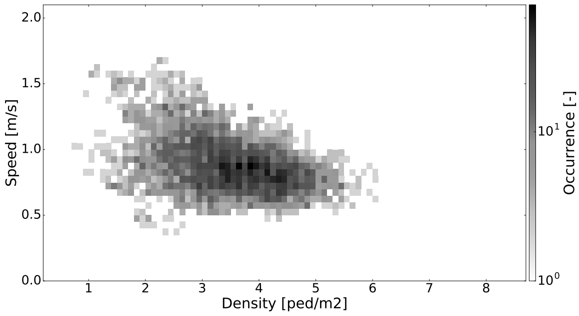


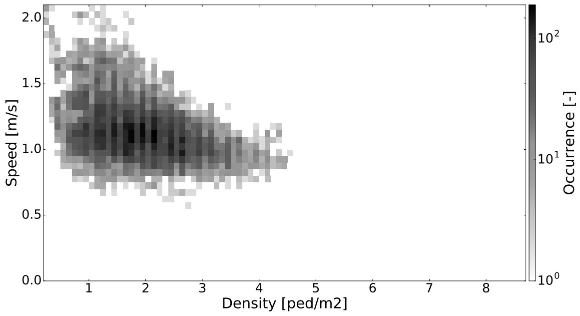
Figure 3.9: Cumulative curve in the bottleneck



(a) Before bottleneck



(b) In bottleneck



(c) After bottleneck

Figure 3.10: Fundamental diagrams (ku) bottleneck scenario AO-360-400

Corner - low & high density

In the corner scenario a unidirectional flow around a corner is simulated whereby, in accordance with the data, the turn is a right handed turn. As Figure 3.11 illustrates, the walkway is 2.4 meters wide and the pedestrians start at the bottom right side and from there move towards the top left part of the infrastructure. The measurement area covers the corner itself and a part of the area before and after the corner.

For the calibration the data sets from the experiments EO-240-060-240 and EO-240-150-240 are, respectively, used for the low and the high density scenarios. The cumulative curves in Figure 3.12 show that in both cases there is a warming up period as the first pedestrian pass through the infrastructure. After this, the flow remains at a relatively steady rate of about 0.6 ped/s/m in case of the low density scenario and 1.2 ped/s/m in the case of the high density scenario. The fundamental diagrams in Figure 3.13 show a similar patterns as the diagrams of the two bidirectional scenarios. In the case of the high density scenario the speeds are generally lower and the densities higher and there is a larger variance in the states. However, the difference between the scenarios are smaller compared to the difference between the two bidirectional scenarios.

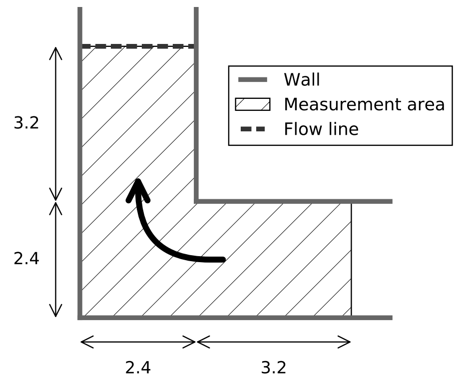
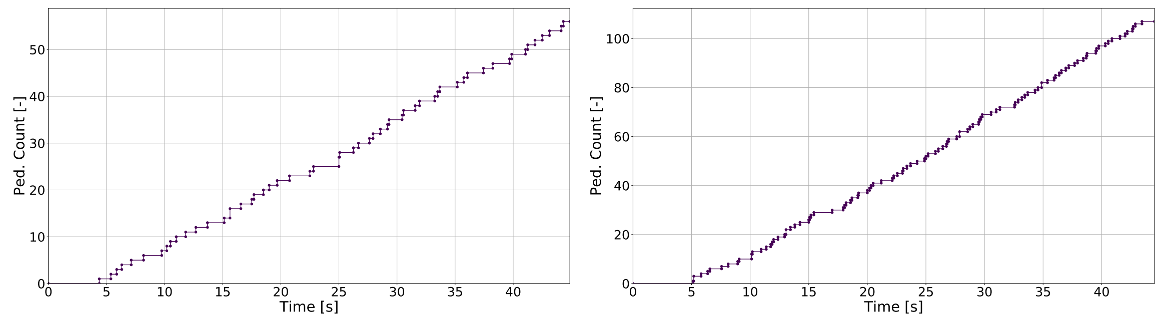


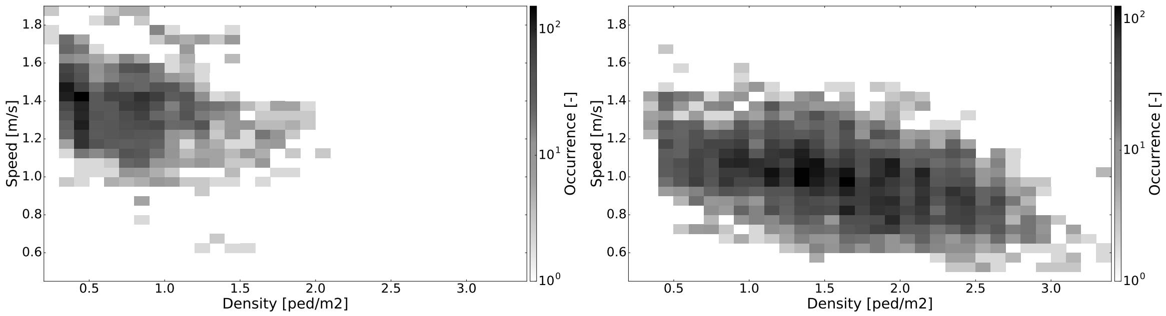
Figure 3.11: Overview of the corner scenario



(a) Low density (EO-240-060-240)

(b) High density (EO-240-150-240)

Figure 3.12: Cumulative curves corner scenario



(a) Low density (EO-240-060-240)

(b) High density (EO-240-150-240)

Figure 3.13: Fundamental diagrams (ku) corner scenario

Merging unidirectional flows in a T-junction - low & high density

In the T-junction scenarios the main movement base case is merging. As Figure 3.14 illustrates, every part of the walkable area is 2.4 meters wide and pedestrians walk from both sides towards the junction where the flows merge and the pedestrians move towards the exit. The measurement area consists of the lower part where the two unidirectional flows join and the upper part where the flows have joined.

During the calibration, data from the KO-240-060-240 and the KO-240-150-240 experiments will be used to compare the simulation results against. The cumulative curves in Figure 3.16, again, show a warming up period as the first pedestrians move through the infrastructure. After this warming up period the flows remain relatively stable at a rate of, respectively, 1.1 and 1.6 ped/s/m for the low and high density scenarios. The fundamental diagrams, depicted in Figure 3.15, also show the same pattern as the corner and bidirectional scenarios. The low density scenario generally has higher speeds, lower densities and a lower variance of the states.

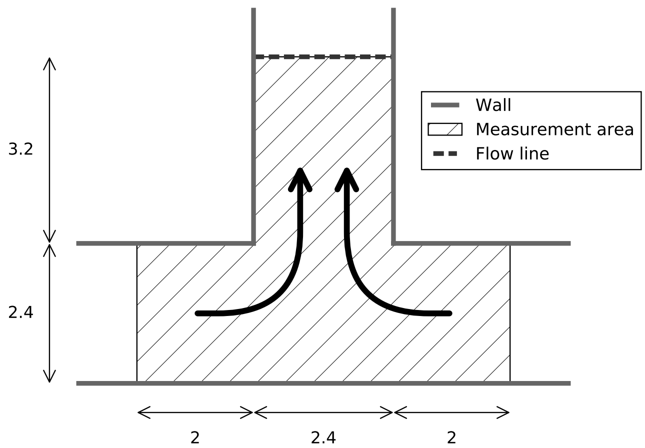
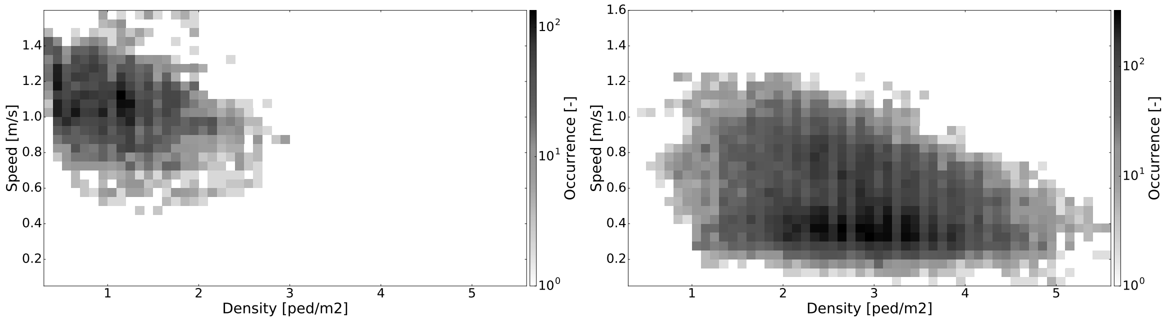
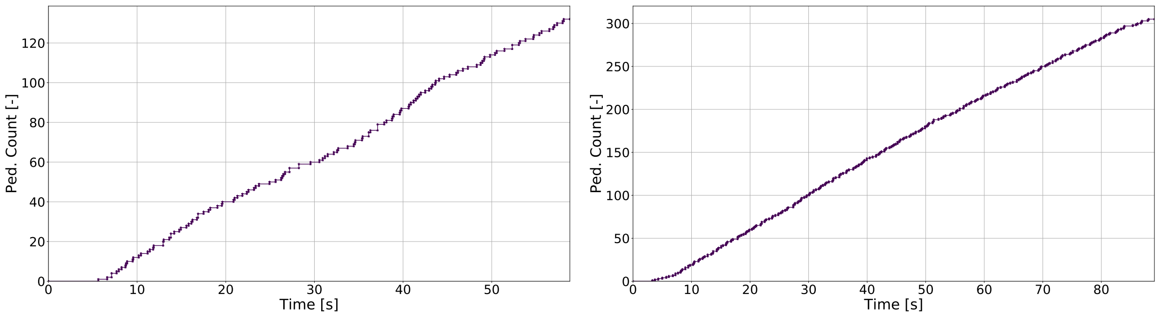


Figure 3.14: Overview of the T-junction scenario



(a) Low density (KO-240-060-240) (b) High density (KO-240-150-240)

Figure 3.15: Fundamental diagrams (ku) T-junction scenario



(a) Low density (KO-240-060-240) (b) High density (KO-240-150-240)

Figure 3.16: Cumulative curves T-junction scenario

3.3 Pedestrian Dynamics

This section introduces Pedestrian Dynamics ® (PD), a microscopic pedestrian simulation model by INCONTROL Simulation Solutions (2016). Pedestrian Dynamics ® (PD) offers a user the ability to model the behaviour of pedestrians at all three behavioural levels (strategic, tactical and operational). Models for both the operational behaviour (route following and collision avoidance) and the route choice (tactical level) are already implemented and these processes are thus influenced by changing the parameters. The processes of activity choice (strategic level) and activity scheduling (tactical level) do not have fixed implementations and are thus left to the user to implement. In the case of this research the pedestrians only have one activity, walking from their origin to their destination, and hence there is no need to model the activity choice or the scheduling. The route choice and walking behaviour are both modelled and will be discussed in somewhat more detail whereby the focus is especially on the walking behaviour given that this will be the focus of the calibration. Furthermore, an overview of the parameters, that influence the operational behaviour, will be given and it will be discussed if they are relevant in the steps to come given the scope of this research.

3.3.1 Route choice

The route choice algorithm, implemented in PD, uses the concept of an Explicit Corridor Map (ECM) (Geraerts, 2010) in combination with the A-star algorithm to determine the global route for a pedestrian. The ECM is able to compute the shortest path given a certain minimal clearance which ensures that a pedestrian will always be able to traverse the given path. PD offers two approaches for computing the global route. The first one is simply using the shortest path (i.e. the path with the shortest *distance* between A and B). The second option is using a least-effort approach whereby the cost is the estimated travel time. This estimated travel time can be computed including the estimated delay caused by an increased density on an edge and including some discomfort factors.

In this research the global route choice is not a relevant part given that all scenarios only have one possible global route for every Origin-Destination pair (OD-pair). Hence, in all scenarios the route choice option is set to the shortest path option given that this is the least computationally heavy option.

3.3.2 Operational behaviour

The operational behaviour consists of two parts, route following and collision avoidance, which together determine the acceleration of a pedestrian at every time step. In PD the acceleration of an agent is determined as follows:

$$\frac{d\vec{v}_i}{dt} = \frac{\vec{v}_{\text{des};i} - \vec{v}_i}{\tau} + \sum_j \vec{f}_{ij} + \sum_W \vec{f}_{iW} \qquad [\text{m/s}^2] \tag{3.1}$$

In the equation above $\vec{v}_{\text{des};i}$ and $\vec{v}_i$ are, respectively, the desired and current velocity of agent i at the current time step. $\vec{f}_{ij}$ and $\vec{f}_{iW}$ are physical forces that occur when an agent comes into contact with another agent or an obstacle. τ is the relaxation time which is a parameter that determines how strongly an agent reacts to deviations from its desired velocity whereby a lower value indicates that the agent's velocity will converge faster to its desired velocity.

The desired velocity ($\vec{v}_{\text{des}}$) is determined according to the method proposed by Moussaïd, Helbing and Theraulaz (2011). The method uses a vision based approach to avoid collisions and combines the collision avoidance with the preferred speed and the desired destination to determine the desired velocity. The desired velocity is a combination of the desired speed and the desired direction. The desired speed is given by $v_{\text{des}} = \min(v_i^0, d_h/\tau)$ where v_i^0 is the preferred speed of the agent (which is determined in the input) and d_h is the distance between the agent and the first expected collision in the desired direction. d_h is determined as follows:

$$d_h = \begin{cases} d_{i;\text{exp:col}} - d_{i;\text{pers}} & \text{If the expected collision is with another} \\ & \text{agent travelling in the same direction} \\ d_{i;\text{exp:col}} & \text{Otherwise} \end{cases} \qquad [\text{m}] \tag{3.2}$$

Whereby $d_{i;\text{exp:col}}$ is the distance to the first expected collision in the desired direction α of agent i and $d_{i;\text{pers}}$ the personal distance of agent i. The personal distance is the distance an agent wants to keep between itself and another agent. The desired direction is determined by minimizing the following function:

$$d(\alpha) = d_{\text{max}}^2 + f(\alpha)^2 - 2d_{\text{max}}f(\alpha)\cos(\alpha_0 - \alpha), \text{ for } \alpha \in [-\phi, \phi] \tag{3.3}$$

As Figure 3.17 illustrates, ϕ is the viewing angle of the agent and together with d_{max}, the viewing distance of the agent, it determines which obstacles are taken into account when calculating the desired direction. $f(\alpha)$ is the distance to the closest expected obstacle in the direction of α whereby it is equal to d_{max} if there are no obstacles within the viewing range. Finally, α_0 is the angle towards the desired destination.

The desired destination is determined by the location of the attraction point which in turn is determined by the Indicative Route Method (IRM) (Karamouzas, Geraerts & Overmars, 2009). The IRM uses the ECM to produce an indicative route that runs from the agent's origin to its destination whereby it is ensured that the agent can traverse the whole route (i.e. the clearance at every point along the route is larger than the radius of the agent).

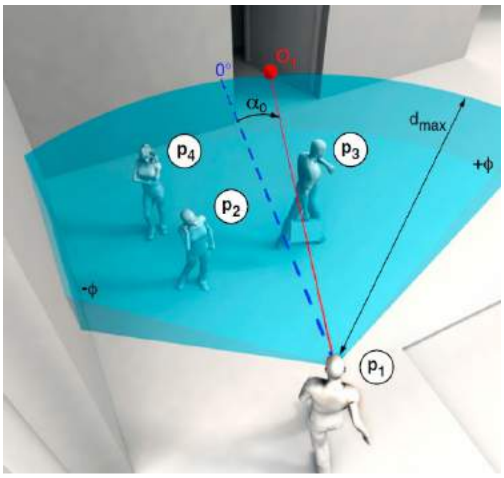


Figure 3.17: Example of the parameters determining the desired direction. (Fig. 1 (A) from (Moussaïd et al., 2011))

3.3.3 Stochastic nature

PD is stochastic by nature which means that two simulations using the same parameters and inputs but different seeds can lead to different results. Within in this research there are three main causes for this stochasticity. Namely, the preferred speed, the initial destination and the exact location of the origin. All three are shortly discussed below.

By default, the preferred speed is a triangular distribution from which a preferred speed for every agent is drawn. So, depending on the seed, an agent spawning at the same time and location in a particular simulation can have a different preferred speed.

For every agent, its initial destination is a randomly chosen point in a destination area. The exact location of the destination within this area is updated based on the current location of the agent whilst it is walking through the infrastructure. However, the exact location of this initial point does influence the indicative route and therefore the location of the attraction point.

The indicative route, and thereby the attraction point, is also influenced by the exact location at which an agent enters the simulation. Comparable to the exact location of the destination, the origin is also a randomly chosen point in a starting area.

The fact that the model is stochastic in nature has to be taken into account. How to deal with this stochastic nature is the topic of the next section.

3.3.4 Parameters operational behaviour

PD contains 11 parameters that influence the operational behaviour. In Table 3.5 an overview of these parameters can be found whereby the parameters have been divided into those that influence the location of the attraction point and those that influence the walking behaviour itself. The parameters are all deterministic and hence the population is homogeneous regarding these parameters. Although there is the possibility to make almost all of these parameters stochastic, this is considered outside of the scope of this research.

Not all of these 11 parameters might be relevant for this research, given for example the limitations set by the choice of scenarios (section 3.2). Parameters that are not influencing the results of any of the chosen scenarios cannot be calibrated and hence do not have to be included. In order to determine which are relevant, for every parameter the following two questions can be posed:

- Does the parameter influence the traffic state in any of the chosen scenarios?
- Does the parameter influence behaviour in a significant way?

Based on these two question four of the eleven parameters can be eliminated. If we take the first question posed above, it is clear that the fixed speed multiplier is not relevant for this research given that fixed speed surfaces do not occur in any of the chosen scenarios and hence this parameter is discarded. Based on the second question three other parameters can be discarded as well. These are the maximum shortcut distance, the side clearance factor and the avoidance preference. The avoidance preference can be discarded because it is considered irrelevant for the results whether or not all pedestrians prefer to pass another pedestrian on the right or left side as long as the preference is equal for all pedestrians. And given that by default it is the same for all pedestrians and given that, as mentioned earlier, the effect of making a homogeneous parameter heterogeneous is outside of the scope of this research this is the case. The maximum shortcut distance and the side clearance factor are both parameters which can be used to fix unrealistic local path finding behaviour in case this would occur in a given part of the infrastructure. These two parameters should thus only be changed when this unrealistic behaviour occurs and for every scenario this should be checked. However, once the values are found which produce realistic behaviour, if they need to divert from the default at all, these values will remain constant for the remainder of the research.

So, 7 of the 11 parameters, presented in Table 3.5, will be taken into account during the sensitivity analysis.

Table 3.5: Operational parameter of PD

Name	Description	Unit	Default value
	Attraction point location		
Preferred clearance	The preferred distance an agent wants to keep between itself and obstacles when planning its indicative route through a corridor.	m	0.3
Max. shortcut distance	The maximum distance the attraction point can be from an agent whereby $<= 0$ means there is no restriction.	m	0
Side clearance factor	This factor determines how the agent plans its local route in relation to the corridor's wall whereby 0 means close to the wall and 1 in the centre.	-	0
Side pref. update factor	Determines how fast the indicative route converges towards the current position of the pedestrian.	-	1
	Walking behaviour		
Min. desired speed	The minimum desired walking speed of agents. Once the speed of an agent drops below this threshold it will stop walking and only start walking again when it can start moving at a speed higher than or equal to this threshold.	m/s	0.06
Fixed speed multiplier	The fraction of the agent's walking speed that is used on fixed surfaces such as escalators.	-	0
Relaxation time	A parameter that determines how strongly an agent reacts to deviations from its desired velocity (τ in Equation 3.1).	1/s	0.5
Viewing angle	The angle of the agent's field of view (ϕ in Equation 3.3).	degree	75
FoV avoidance range	The distance of the agent's field of view (d_{max} in Equation 3.3).	m	8
Avoidance preference	The bias that determines the preferred side of passing an obstacle.	-	Right
Personal distance	The desired personal distance an agent wants to keep between itself and another agent (d_{pers} in Equation 3.2).	m	0.5

3.3.5 Conclusions on the microscopic model

This section introduced the microscopic model used within this research. The model is capable of modelling all three levels of behaviour (strategic, tactical and operational). However, the processes of activity choice (strategic level) and activity scheduling (tactical level) won't be modelled within the research given that all agents only have one activity. The route choice algorithm offers two approaches, a shortest path and a least effort approach. Given that all OD-pairs in all scenarios only have one possible global path the decision is made to use the shortest path approach given it is the least computationally heavy option of the two.

In total 11 parameters were identified which could influence the operational behaviour. Four of those, the fixed speed multiplier, the maximum shortcut distance, the side clearance factor and the avoidance preference are not considered relevant for this research and hence won't be taken into account during the sensitivity analysis and the creation of the search space.

3.4 How to deal with stochasticity

Pedestrian models often contain stochastic elements. This is also the case for PD and hence it is necessary, for both the sensitivity analysis and the calibration, to run multiple replications of the same simulation in order to assure than any differences found between two different parameter sets are not caused by the stochasticities but by the difference in parameter values. So, this section will explain how it is determined how many replications need to be performed.

As is clear from the literature review, there are multiple methods for determining the required number of replications and a choice has to be made which method is most suitable for this particular research. Besides the method itself, two other choices have to be made, namely, which metric or metrics are used and what values to use for elements such as the confidence level, the allowable error or the number of consecutive replications.

Regardless of the choice of method, metric or confidence level, the underlying principle is the following: One has enough replications if the results of the n replications together form a good approximation of the actual probability distribution of the results (i.e. the distribution of results one would get when the number of replications would go to infinity). In other words, if one were to have two sets of n replications of the same simulation the probability distribution of the results of these two sets would be considered to be samples drawn from the same distribution given a desired level of confidence.

As stated in subsection 2.2.5, the first choice to be made is whether to use the two-step method or one of the two sequential methods. Given that it is unknown how the different combinations of parameter sets and scenarios influence the necessary number of replications and testing, beforehand, which number of replications to use for every combination is too time consuming for this research, the decision is made to use a sequential method.

In the remainder of this section the following elements will be discussed. First, it will be explained which metric/metrics will be used. Second, it will be explained which method, out of the two remaining ones, will be chosen. Lastly, the chosen method will be tested to get insight into its performance and to determine which values are going to be used for elements such as the confidence level, the allowable error or the number of consecutive replications, depending on the chosen method.

3.4.1 Choice of metric

Before a choice regarding the metric/metrics is made, the question - What are the requirements within this research regarding the metric/metrics which should be used for determining the required number of replications? - should be answered first.

The requirements for the metric/metrics follow from the goal of this research, which is, among other things, to assess how different combinations of metrics influence the calibration result. As is explained in subsection 2.2.3, different metrics describe the flow at different aggregation levels. The multiple-objective approach, ideally, includes using metrics at these different aggregation levels. Hence, it is important that the metric/metrics used in determining the required number of replications take into account these different levels. So, the metric or metrics used should give insight into how the stochasticities effect the overall flow. This can be expressed by metrics such as the flow, the fundamental diagram or the mean speed or travel time. Moreover, the metric should provide insight into the underlying behaviour of which the flow is the result (i.e. metrics at the meso or microscopic level).

Besides the requirements described above, one also has to take into account the fact that a sequential method has to be run after every replications (or n replications). Hence, from a practical point of view, it is preferable that the calculation of the metric/metrics is not computationally heavy. For example, calculating the density using the XT-method (Duives et al., 2015) is more time consuming than calculating the travel time or the distribution of the speeds.

So, based on both the requirements and the preference from a practical point a view, the choice is made to use the distribution of the speeds as the only metric. The distribution is comprised of the instantaneous speeds of all pedestrians whereby the speeds are only included for those time step the pedestrian is in the measurement area. This choice is made for the following reasons: Firstly, it is a computationally light metric, given that the distribution is made up of the instantaneous speeds which are easily determined based on the trajectories. Secondly, the distribution of the speeds is able to give insight into both the efficiency of the flow (a higher mean speed indicates a more efficient flow) and into the underlying behaviour (e.g. a high variance can indicate that interactions are not solved efficiently causing pedestrians to change their speed a lot).

3.4.2 Choice of method

As concluded in the introduction of this section, a sequential method will be used to determine the number of required replications. As subsection 2.2.5 shows, there are two possible sequential methods out of which one has to be chosen. The chosen method must determine whether the speed distribution has converged such that data from any additional replications won't change the distribution significantly. To test whether two speed distributions are samples drawn from the same distribution one can use the Anderson-Darling test (A-D test)[1] (NIST/SEMATECH, 2012a). The use of the A-D test has direct implications for the choice of the method given that it cannot be used in combination with the method proposed in (Toledo & Koutsopoulos, 2004). So, the method proposed in (Ronchi et al., 2013), which is based on one or more convergence criteria, will be used.

The following procedure is used to determine whether the distribution of the speed has converged:

if $p_i > p_{threshold}\ \forall\ i \in I$ **then**
  converged = $True$;
else
  converged = $False$;
  Run additional simulation to obtain more replications;
end

where

p_i	=	AD$(S_{i-1},\ S_i)$, The significance level at which the null hypothesis, that the two samples come from the same distribution, can be rejected.
I	=	$\{m-b+1, m-b+2, ..., m\}$
S_i	=	Speed distribution consisting of the speeds of all i replications
m	=	Current number of replications
b	=	Number of consecutive replications for which the test should hold

The pseudo-code above illustrates that the speed distribution has converged if for b consecutive runs the null hypothesis cannot be rejected (i.e. the p-value is larger than the chosen threshold value ($p_{\text{threshold}}$)). Both the number of consecutive runs and the threshold value have to be chosen beforehand. The next subsection will explore which values should be used within this research.

3.4.3 Testing the method

In this subsection the convergence method using the A-D test will be tested whereby two aspects will be looked at. Firstly, how do different values for the number of consecutive runs (b) and the threshold value ($p_{\text{threshold}}$) influence the required number of replications and the performance? Secondly, how does the exact order of the seeds influence the performance and the required number of replications? The method is considered to perform well if the required number of replications is such that adding any number of extra replications won't change the results (i.e. if the method has converged after m replications the following holds AD$(S_m, S_{m+i}) > p_{\text{threshold}}\ \forall\ i \in \mathbb{N} \geq m$).

All tests in this section are performed using data from two different scenarios. Both scenarios are simulations of a bidirectional flow and use the default settings of PD. In the first scenario a flow of 0.2 ped/s/m/dir is used and in the second scenario a flow of 0.4 ped/s/m/dir. The first flow is chosen such that the no breakdowns [2] occur whilst in the case the second flow this is likely to occur. For both scenarios 900 replications, all with a unique seed value, are simulated using the default parameter sets. The results of these 900 simulations are then placed in 100 different orders simulating 100 different orders of seeds.

To explore how different values for b and $p_{\text{threshold}}$ influence the results, nine different combinations of values are tested. In Table 3.6 one can find the values used.

[1]The A-D test was chosen over the Kolmogorov-Smirnov (KS) test as it is generally considered to be more powerful (NIST/SEMATECH, 2012b)

[2]Breakdown: The speed of the pedestrians has dropped significantly being zero or almost zero and hence pedestrians are only capable of moving towards their destination with large difficulty if it is at all possible

Table 3.6: The values of $p_{\text{threshold}}$ and b used in the tests

$$p_{\text{threshold}} = 0.25, 0.10, 0.05$$
$$b = 10, 15, 20$$

In Figure 3.18 one can find the results for all nine combinations for both the low and high density scenarios. The figure clearly illustrate two things. Firstly, the value $p_{\text{threshold}}$ does influence the required number of replications whereby the larger the threshold values the more replications are required. The value of b does seem to have little impact on the required number of replications compared to the threshold value. Secondly, the order of the seeds does have a clear influence on the resulting number of required replications given the large spread of the results. In this case the value of b does seem to have an influence on the results whereby a larger value of b results in less variance. Furthermore, the figure also illustrates that the variance in the required number of replications is larger for the high density scenario compared to the low density scenario. Due to the possibilities of the flow breaking down in the high density scenario, one would expect the results of different replications to vary more and hence a larger number of required replications.

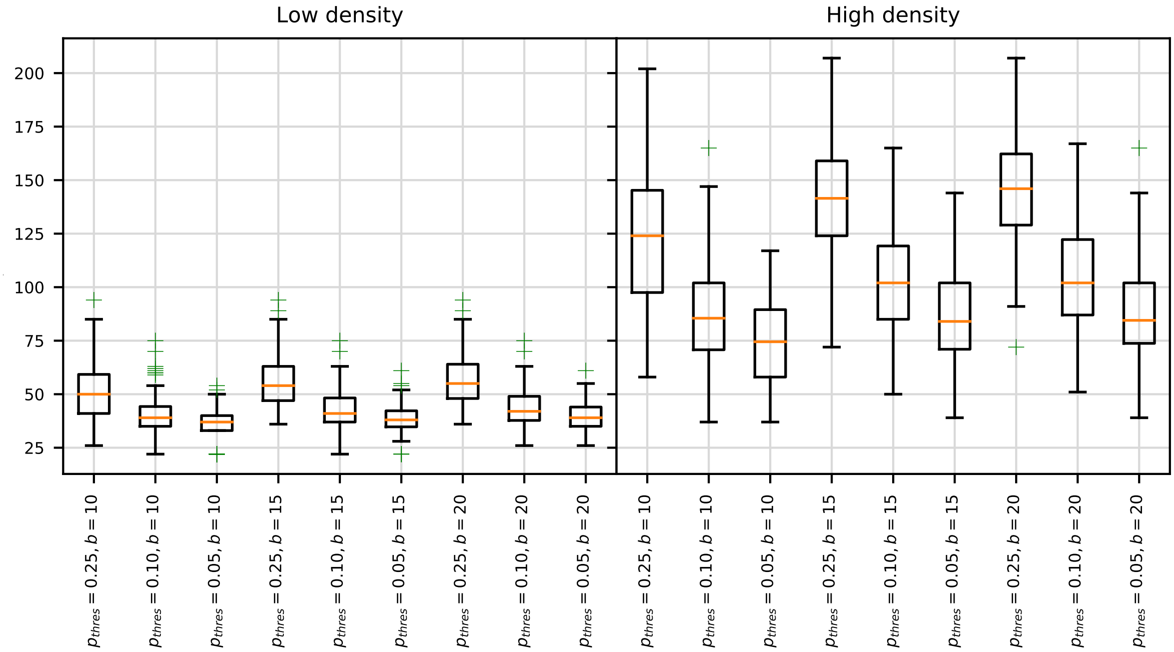
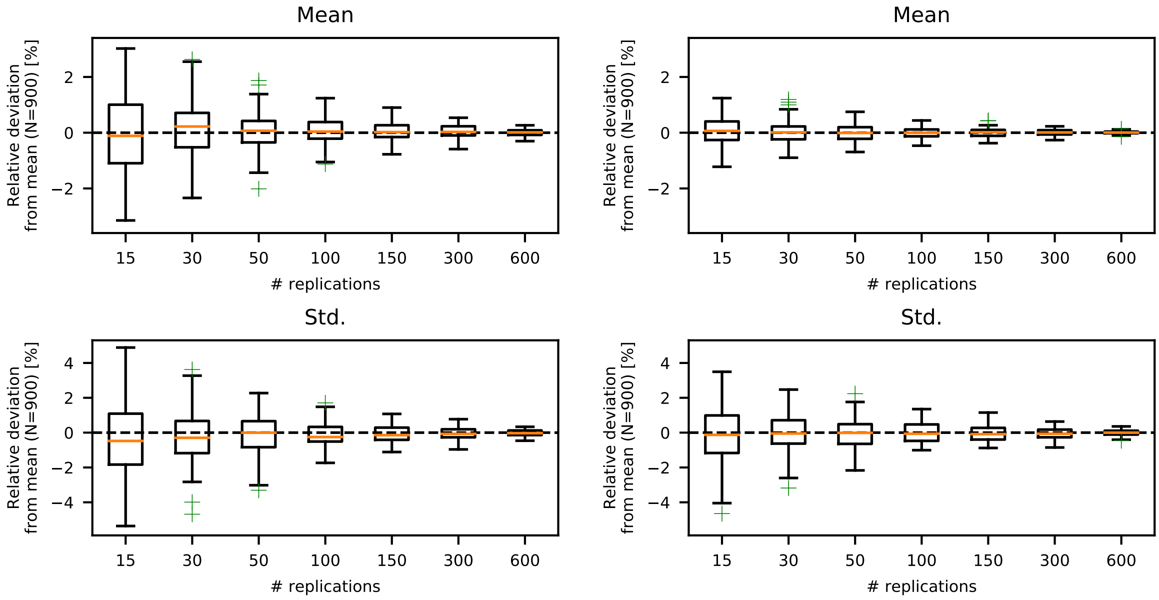


Figure 3.18: Required number of replications per combination for different values of $p_{\text{threshold}}$ and b

Based on the obtained required number of replications, a second test is performed to get insight into the performance. The test determines if the speed distribution, containing the data from the required amount of replications, is drawn from the same distribution as the distribution containing the data from all 900 replications. The test uses the A-D test whereby the distributions are not considered to be drawn from the same distribution if the null-hypothesis can be rejected at the 5% level. As mentioned before in this section, the method is considered to perform well if the distribution does not change significantly when any number of additional replications is added. In the case of the low density scenario, between 58 and 63 percent of the distributions is not considered to be drawn from the same distribution as the distribution containing the data from all 900 replications. In the case of the high density scenario 35 to 40 percent of the distributions are not considered to be drawn from the same distribution as the distribution containing all 900 replications. So, overall the method does not seem to perform very well given that for both scenarios the distribution changes significantly in many cases when additional replications are added. The stopping criteria, however, do not seem to have a large influence on the performance given that the difference between the different sets of stopping criteria are at most 5%. Furthermore, plotting the required number of replication versus whether or not the null-hypothesis was rejected showed no correlation. This indicates that it is the exact order of seeds that determines if the distribution changes significantly if more replications are added and not so much the required number of replications. The implication of the marginal performance of the method is that, regardless of the stopping criteria one uses, differences between sets of simulations can still be caused by the stochasticities.

To explore how big the impact of the stochasticities can be on the results and how this correlates with the required number of replications, a third test is performed. Again, the 900 replications are placed in 100 different orders and for six different numbers of replications the speed distributions are obtained. Figure 3.19 illustrates, per number of replications, how the mean and standard deviation of the 100 speed distribution vary. As the figures illustrate, the variance decreases as number of replications increases. However, even at 500 replications the mean and standard deviation can still vary depending on the exact order of the seeds. The figure also illustrate that, although the patterns are similar for both scenarios, the variances differ whereby the variance is larger in the case of the high density scenario.



(a) High density scenario (b) Low density scenario

Figure 3.19: Influence of the order of the seed versus the number of replications

Overall, the tests show that the performance of the method is not very good regardless of the choice of stopping criteria. As the figures in Figure 3.19 illustrate, one has to take into account the possible variance caused by different seed orders when comparing results of different simulations. For a given required number of replications, one can expect a certain variance in the results. Hence, for differences between simulations that fall within this range one cannot exclude the possibility that this is caused the stochastic nature of the model.

3.4.4 Conclusions on how to deal with the stochastic nature of the model

So, based on the subsections above the following can be concluded. The method for determining the required amount of replications will use the speed distribution as the metric and a convergence based method using the A-D test. Given that the stopping criteria do not seem to strongly influence the performance of the method the choice if made to use 0.25 and 10 as the values for, respectively, $p_{\text{threshold}}$ and b. The lowest number of consecutive replications was chosen because the higher the number of consecutive replications the more computationally heavy the method becomes. The largest value for the threshold value was chosen because, as Figure 3.18 illustrates, this values clearly leads to a higher number of replications and as Figure 3.19 illustrates, a higher number of replications leads to a smaller influence of the exact order of the seeds. However, regardless of the required number of replications one obtains using this method, the exact order of the seed will still have to be taken into account when comparing different simulations. Based on figures, such as those shown in Figure 3.19, one should get an estimate of the size of the differences that can be caused by differences in the order of seeds. This should be done for every scenario as Figure 3.19 indicates that differences between scenario exist.

3.5 Conclusions

In this chapter a number of elements relevant for the following chapter have been discussed. The chapter answers the questions: Which reference data to use?; Which scenarios to use?; Which parameters to use?; and; How to deal with the stochastic nature of the model?

Based on an evaluation of the available reference data the choice was made to use data collected during the Hermes project. The evaluation of the available data also showed there to be not enough data to investigate the effect of different population compositions on the calibration.

Based on the literature review and the availability of reference data, seven scenarios were chosen to be used in the upcoming sensitivity analysis and calibration. These seven scenarios include five of the ten movement base cases and two different density levels. All seven scenarios are also coupled to an reference data set which will be used during the calibration to compare the simulation results against.

An analysis of PD showed it to be stochastic in nature. Furthermore, 11 parameters were identified which influence the operational behaviour. Out of these 11 parameters, four are not considered relevant within the scope of this research and hence won't be taken into account during the sensitivity analysis or the formation of the search space for the calibration.

To determine the required number of replications a sequential method, based on convergence criteria, and using the A-D test, will be used. Tests of the method showed its performance to be problematic and hence an estimate of the possible difference caused by the stochasticities, for a certain number of replications, is necessary for all scenarios. This estimate will indicate if differences between simulations are likely caused by stochasticities or that they can be attributed to differences in parameter values.

4 | Sensitivity analysis

In this chapter the sensitivity analysis will be discussed. The goal of the sensitivity analysis is to obtain insight into the model's sensitivity to changes in the operational parameters. These insight will be used, among other things, to determine the search space for the calibration.

The insight into the sensitivities are used, when determining the search space, to limit the number of parameter included in the search space. The reason for limiting the number of parameters included in the search space is the following: The size of the search space and the number of parameters are exponentially related (k^n, whereby n is the number of parameters and k is the number of parameter values that will be taken into account during the calibration). An optimization algorithm will take at worst $k^n * t$ to determine the optimal parameter set whereby t is the amount of time it takes to obtain the fitness of a given parameter set. In the case of this research this t is fairly high given that for every point one has to run many simulations due to the need for multiple replications and the usage of multiple scenarios. For example, running simulations for all seven scenarios using 50 replications per scenario takes about 5 minutes on the computer available to this research.

When deciding whether or not to include a parameter in creating the search space the following rationale is used: The more sensitive the model is to changes in a particular parameter, the more important it is that one obtains a good estimate of its value given that a small deviation can lead to significantly different results.

The chapter is build up as follows: In the first section (section 4.1) the methodology for the sensitivity analysis is discussed. As this section will show the methodology is based on a qualitative analysis, discussed in section 4.2, and a following quantitative analysis discussed in section 4.3.

4.1 Methodology

This section introduces the methodology used to perform the sensitivity analysis. Ideally, the sensitivity analysis gives insight into both the model's sensitivity to changes in a single parameter (first-order effect) and simultaneous changes in multiple parameters (higher-order effects). A review of the literature did not turn up any methodologies specifically aimed at pedestrian simulation models. However, a brief review of the literature of the closely related field of microscopic (motorized) traffic simulation models did turn up a methodology. In (Punzo, Montanino & Ciuffo, 2015) a variance-based method is proposed using a sensitivity index, including both first and higher-order effects, to rank the parameters. Although the method has the advantage of both incorporating first and higher-order effects, compared to, for example, a one-at-a-time method which only takes into account the first-order effects, it is computationally expensive. Certainly if one takes into account that, in the case of pedestrian simulation models, determining the sensitivities based on one scenario is impossible as a consequence of the differences in behaviour between different movement base cases (Duives, 2016). Furthermore, an insight into just the first-order effects is deemed sufficient for the goals of this research. So, in the remainder of this section a three-step methodology, based on a one-at-a-time principle, will be discussed.

4.1.1 Three-step methodology

In this research the choice is made to perform the sensitivity analysis in three steps. This methodology is chosen primarily because it limits the amount of simulations necessary to obtain the first-order sensitivities of all parameters in all scenarios. The three step of the methodology are depicted in Figure 4.1. As the figure illustrates, the process is performed using a combination of one scenario and one parameter. Exactly which combinations of scenarios and parameters are tested is explained in the next subsection. Below, the three steps are introduced:

1. The first step in the process is to find the maximum deviation from the default value, for the given parameter and scenario, that still results in realistic behaviour. This analysis is based on a

qualitative analysis which is explained in more detail in section 4.2. This step results in an upper and lower boundary which are the input to the next step. Because PD has undergone some basic calibration, it is assumed that the optimal values will not deviate much from the default values. Hence, a maximum deviation of $\pm 25\%$ is used. The added advantage of this step is that it also gives insights into elements that are not easily quantified in a metric. For example, this step gives insight into whether or not the model is capable of producing lanes in a bidirectional flow.

2. The second step is a quantitative analysis which determines if the upper and lower boundary result in significantly different results compared to the default. If neither of the boundaries significantly differs from the default, it is assumed that the model is not sensitive to changes in this parameter for the given scenario and boundaries. If one or both of the boundaries result in significant differences we continue to the third step.
3. In the third step the development of the sensitivity over the whole range between the upper and lower boundaries is investigated. In case only one of the two boundaries produces significant differences the development of the sensitivity is only investigated for the range between the default value and the boundary that produced significant different results.

Both step two and three, involving a quantitative analysis, are explained in more detail in section 4.3.

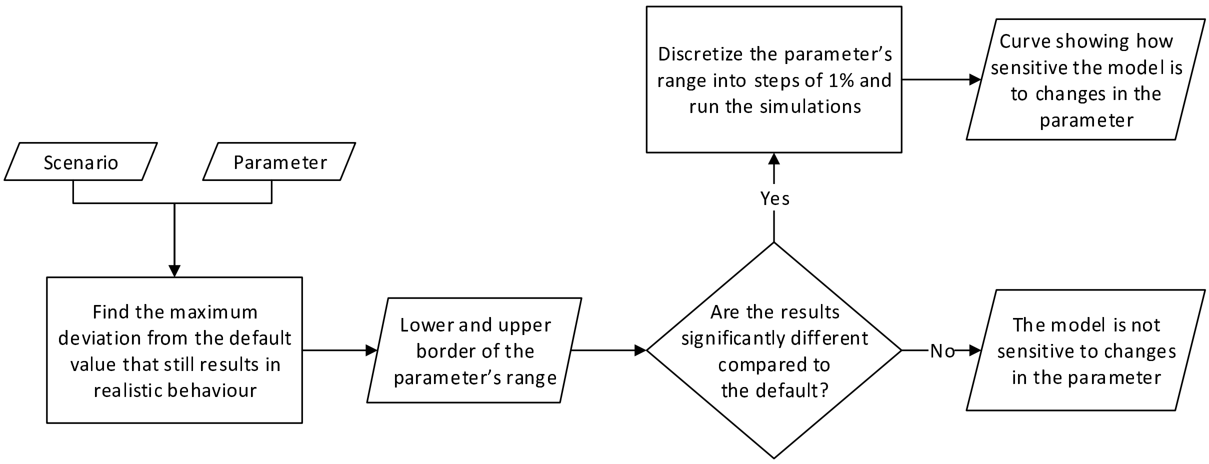


Figure 4.1: Overview of the methodology for the sensitivity analysis of a single parameter and scenario

4.1.2 Combinations of scenarios and parameters

The methodology chapter (chapter 3) concluded that, in total, seven scenarios will be used during this research and that seven parameter are of interest. This leads to 49 combinations of scenarios and parameters that need to be investigated during this sensitivity analysis.

For every one of these 49 combinations the question - Is a change in the parameters value unlikely to influence the simulation results? - is posed. Based on this question, three combinations are excluded from the analysis and in the case of four other combinations the analysis will only be performed for a decreased value. The three combinations that are excluded from the analysis are the combinations between the minimal desired speed and the low density scenarios. The low density scenarios are defined such that the speeds of pedestrians should remain high and never come close to the minimal desired speed, even when it is increased by the maximum deviation of 25%. In the case of the four high density scenarios the default value for the FoV avoidance range is used as the upper boundary. A value larger than the default value is not expected to make any difference given that PD only takes the four closest pedestrians into account and in the case of the high density scenario it is highly unlikely that these four closest pedestrians are not within the 8 meter range (the default value).

So, as Table 4.1 illustrates, for 42 of the 49 combinations the sensitivity will be investigated for both an increase and a decrease in the parameter's value, in four of the 49 cases it will only be investigated for a decrease in the value and three of 49 cases won't be investigated at all.

As noted in section 3.2, the scenarios used for the sensitivity analysis differ slightly from those used for the calibration. In appendix B one can find the exact lay-out for all scenarios and the inflows used.

Table 4.1: Combinations of scenarios and parameters investigated during the sensitivity analysis. The acronyms identify the scenarios (i.e. B-L = bidirectional low, B-H = bidirectional high, B = bottleneck, C-L = corner low, C-H = corner high, T-L = t-junction low, T-H = t-junction high).

	B-L	B-H	B	C-L	C-H	T-L	T-H
Preferred clearance	-+	-+	-+	-+	-+	-+	-+
Side pref. update factor	-+	-+	-+	-+	-+	-+	-+
Min. desired speed	x	-+	-+	x	-+	x	-+
Relaxation time	-+	-+	-+	-+	-+	-+	-+
Viewing angle	-+	-+	-+	-+	-+	-+	-+
FoV avoidance range	-+	-	-	-+	-	-+	-
Personal distance	-+	-+	-+	-+	-+	-+	-+

-+ Relevant to check both an increased and a decreased value (compared to the default)

+ Relevant to check an increased value (compared to the default)

- Relevant to check a decreased value (compared to the default)

x Not considered relevant

The inflows are chosen such that the high density scenarios have a flow twice that of the low density scenarios. In the case of the bidirectional scenario, the high density inflow was chosen such that during the simulation time of 1.5 minutes the flow would not break down.

4.2 Qualitative analysis

In this section the results of the qualitative analysis are presented whereby the goal of this analysis is to determine the boundaries for the quantitative analysis. Before the results are presented it will be explained how the analysis is performed. In subsection 4.2.1 the setup of the analysis is discussed. This is followed by the discussion of what is considered realistic behaviour in subsection 4.2.2. Subsection 4.2.3 will present and discuss the results.

4.2.1 Setup of the qualitative analysis

The basis of the analysis are 86 separate simulations (42 combinations which both a simulation with an increased value of 25% and a simulation with an decreased value of 25% plus four simulations with only a simulation with a 25% decreased value). All these 86 simulations are reviewed by hand to determine whether or not the behaviour is considered realistic according to the criteria explained in the next subsection.

In case the behaviour in a simulation is found to be realistic, the boundary is kept at the maximum deviation of 25%. If this is not the case the bisectional method is used to find the maximum boundary, for which the behaviour is considered realistic, within a precision of 5% point. This takes at most 3 or 5 steps including the first test with the maximum deviation.

4.2.2 Assessment of the behaviour

As the methodology states, it has to be determined what constitutes realistic behaviour. Because it is considered easier to describe unrealistic behaviour, the assessment is based on whether or not the simulations show unrealistic behaviour. First, a general description of characteristics that signify unrealistic behaviour is given. These descriptions are independent of the movement base case. Secondly, a description is given of additional characteristics of unrealistic behaviour that are related to a specific base case. Work by (Campanella et al., 2014) forms the basis for these descriptions.

Because it is not known how realistic the model's behaviour is when using the default values, for every scenario, a simulation is performed using the default values. These seven simulation runs are used as a baseline to which the other simulations are compared. This is done to assess whether any unrealistic behaviour found in the simulations is the result of a change in the parameter's value or whether it is a case of the model not being able to model this behaviour realistically by default. Only the first case is reason for continuing the search for the maximum boundary that does produce realistic behaviour.

General description

To assess whether or not the simulated behaviour is realistic a number of elements are used to assess the results.

Number of collisions: How often do pedestrians collide with nearby pedestrians? In this case a collision is defined as physical contact, the two circles representing the pedestrians overlap, that causes a clear and sudden change in either the speed and/or direction for one or both of the pedestrian involved. Examples of this are: Two pedestrians colliding head-on, the circles overlap, and as a result one of them, or both, bounce back. Or, two pedestrians colliding under an angle whereby one or both of them bounce to the side. Two pedestrian who brush each other, whereby the circles show a small overlap, but whose speed and direction does not show a sudden change, is an example which is not considered to be a collision. The more collisions the more unrealistic the behaviour is consider to be.

Straying outside the walkable area: Do pedestrian stray outside of the boundaries of the walkable area? The more they stray outside the walkable area the less realistic the simulation.

Erratic movement: Does the pedestrian move erratically, making many sudden changes to its speed and direction (not caused by a collision)? Given that, within this research, the scope is normal walking behaviour and not, for example, running behaviour or panicking behaviour, a pedestrian isn't expected to make many sudden changes in his/her movement and the more this happens the less realistic the behaviour.

Bidirectional stream

In a bidirectional stream two additional elements are considered, namely, **lane formation** and **pedestrian being pushed backwards**. In the first case one would expect lanes to form in a bidirectional stream and hence some leader-follower behaviour. Lack of this leader-follower behaviour is considered unrealistic. Secondly, given that the bidirectional flow causes head-to-head interactions there exists the possibility that a pedestrian is pushed backwards by one or more other pedestrian travelling in the opposite direction. This is considered to be unrealistic behaviour.

Bottleneck

In the bottleneck scenario there are three additional elements which are considered. These are:

Pedestrians near the wall are not able to enter the bottleneck: A simulation is considered less realistic if pedestrian near the wall of the bottleneck are consequently trapped and are not able to enter the bottleneck.

Movement towards the wall: One would expect pedestrians, who are not already close to the wall, generally not to move towards the wall given that it is more difficult to enter the bottleneck when standing next to the wall.

Pedestrians do not fan out after exiting the bottleneck: After the bottleneck the walkable area widens and hence it is expected that the pedestrians exiting the bottleneck will fan out and walk over a larger width than the bottleneck width.

Corner

In the case of a corner scenario, a smooth turning movement thought the corner is expected so lack of this indicated unrealistic behaviour. Furthermore, work by Duives (2016) showed that pedestrians temporarily reduce their speed upstream of the corner and increase their speed again downstream of the corner. Hence, no reduction of speed is considered unrealistic behaviour.

T-junction

Based on research by Zhang, Klingsch, Rupprecht, Schadschneider and Seyfried (2012) two additional elements are taken into account when assessing the simulations. Firstly, as the data in (Zhang, Klingsch, Rupprecht et al., 2012) shows, the two flows show a high level of segregation even after they have merged. So, lack of this high level of segregation is considered to identify unrealistic behaviour. Secondly, the research shows that one would expect a low density area with a triangular shape where the two flows merge. So, the more pedestrians pass through this area the more unrealistic the behaviour.

4.2.3 Results of the qualitative analysis

In this part of the section the results of the qualitative analysis are presented. For every scenario it is described what: a) the behaviour is when the default values are used, and b) for every parameter, what the boundaries are that still result in realistic behaviour.

Bidirectional straight - high

The assessment of the simulation using the default values showed that the pedestrians walk in a non-erratic manner, form lanes which are stable and don't push each other backwards. However, pedestrians do tend to cross the boundaries of the walkable area at multiple occasions, although the centre of the circles always remain within the walkable area. During the simulation, collisions occur quite often and if they occur they are usually between two pedestrians moving in opposite direction and under an angle such that the pedestrians bounce to the side. So when the default parameter values are used, the model seems to lack some level of realism when it comes to staying within the walkable area and when it comes to avoiding collisions.

The assessment of the simulations for the different parameters showed little differences in the behaviour when compared to the default case. Only in the case of the increase of 25% of the relaxation time did the number of collisions seem to increase somewhat compared to the default case. However, the difference is still relatively small. So, overall the changes in parameters with $\pm 25\%$ do not seem to cause the behaviour to become significantly less realistic than is the case when the default values are used. Hence all boundaries are set to the 25% deviation.

Bidirectional straight - low

The simulation of the scenario with all parameters at their default value showed behaviour comparable to the high density case of the same scenario. So, in this scenario it is also the case that the model cannot prevent the occasional crossing of the boundaries and collisions.

The inspection of all the other simulations did not show clear differences in behaviour compared to the default case. Hence, all parameter boundaries are kept at the maximum deviation of 25%.

Bottleneck

Assessing the results of the simulation performed with the default values showed some collisions, primarily side-to-side or front-to-back, and some cases where pedestrians cross the boundaries slightly. Overall, the movements are non-erratic and pedestrians positioned near the wall at the bottleneck entry are able to enter the bottleneck and hence do not get stuck. After the pedestrians exit the bottleneck they do fan out whereby especially faster pedestrian do this in order to overtake slower pedestrians.

All combinations of deviations and parameters were assessed and none of the cases showed large deviations in the behaviour. So, in all cases the boundaries for the deviations remain at the maximum value of 25%.

Corner - high

The results of the simulation using the default values for all parameters showed smooth turning movements, a few collisions, primarily pedestrians brushing each other, and some minor crossings of the boundaries of the walkable area. Furthermore, pedestrians do slow down upstream of the corner to speed-up again once they leave the corner.

The assessment of all other simulations did not show any clear unrealistic behaviour which leads to the conclusion that all boundaries are set to the maximum deviation of 25%.

Corner - low

The simulations of the corner scenario with low flows all showed similar behaviour. The turning movement is smooth, collisions do not occur, the boundaries of the walkable area are not crossed and the pedestrians do seem to slow down somewhat upstream of the corner and speed up again downstream of the corner especially at the inside of the corner. So, none of the simulations showed behaviour which can be considered unrealistic and hence all boundaries are kept at the 25% deviation.

T-junction - high

The assessment of the simulation using the default values showed that the pedestrian walk in a non-erratic manner. However, the boundaries are crossed quite frequently and a good number of collisions, primarily side-to-side, occur especially in the area where the two flows meet. Once the flows have merged, boundary crossings and collision don't occur. Once the flows have merged pedestrians do primarily walk on the same side of the corridor as the branch from which they originated. However, there are also quite a few pedestrians who take a path that causes them to end up on the other side of the corridor. This seems to correlate with the lack of the triangular shaped low density region which would be expected at the point where the two flows meet. The primary cause of this seems to be the route choice algorithm which is outside of the scope of this research.

The other simulations showed similar behaviour compared to the simulation with the default values and large differences were not found. Hence, the boundaries of all parameters are kept at the maximum of 25%.

T-junction- low

The results of all of the simulations showed little differences compared to the default case. In all cases the movements where smooth and boundaries were not crossed. Very few collision occurred and if they occurred they were primarily side-to-side and occurred in the area where the two flows meet. Similar to the high flow case the flows after merging were not as segregated as one would expect and in the low flow case this is far more apparent probably because in the high flow case the higher density level forces pedestrians more to one side than in a low density case. Given that the simulations showed little differences the boundaries all remain at the 25% deviation level.

4.2.4 Conclusions on the qualitative analysis

As is clear from the assessments of the different scenarios, the model does lack some level of realism primarily due to the number of collisions that occur and the number of times the boundary of the walkable area is crossed. This primarily seems to be the case at higher densities and when the interactions are not simply front-to-back.

The assessment of the effects of a change in a parameter's value did show that in all cases the change in the parameter's value with plus or minus 25% did not cause the behaviour to become unrealistic. So, the boundaries of all parameters for the quantitative analysis are set at the maximum deviation of 25%.

4.3 Quantitative analysis

This section presents the quantitative analysis. As described in section 4.1, it is performed to determine whether the results at the boundaries, obtained in section 4.2, differ significantly from the default. If the boundaries do significantly differ, a further analysis is performed to determine the development of the sensitivity over the whole range between the lower and upper boundaries.

In subsection 4.3.1 the setup of the analysis is discussed. Based on this setup the analysis is performed, the result of which are presented in subsection 4.3.2.

4.3.1 Setup of the quantitative analysis

In order to determine whether the results at the boundaries differ significantly one needs to answer two questions: 1) Which metric or metrics to use?, and 2) How to test whether a difference is significant?

In this research the distribution of the speeds is used as the metric, for the same reasons as it is used for determining the number of replications (see section 3.4). The speed distribution is made up of all instantaneous speed measurements for all pedestrians during the time they spend in the measurement area. For all scenarios except the bottleneck scenario, the first 15 seconds of a simulation are used as a warming up period to populate the scenario. Speeds collected during these first 15 seconds are thus not part of the speed distribution.

To test whether the speed distribution, obtained using the boundary value of a parameter, differs significantly from the default, the A-D test is used. If the null hypothesis of the A-D test, the two samples are drawn from the same distribution, can be rejected at the 0.05 significance level, the speed distributions are considered to differ significantly. However, as section 3.4 also showed, depending on the scenario and the number of replications one cannot reject the possibility that a certain difference exists due to the stochastic nature of the model. Hence, it should be determined if the difference is larger than the difference which could exist purely due to the stochastic nature. This is done by comparing the means and standard deviations of the two speed distributions and, together with the number of replications, compare those differences with the differences found when testing for the influence of the order of the seeds. For example, in the case of the bidirectional scenarios one would compare the differences to those presented in Figure 3.19.

To assure that the speed distribution has indeed captured the relevant behaviour a second metric will be used to assess any differences. This second metric is the number of pedestrians that exit the measurement area during the measurement period. This metric is used because it is an easy to determine metric for the efficiency of the flow. Using this metric, it will be assessed if the speed distribution indeed captures the efficiency of the flow well whereby one would expect a lower mean speed to correlate with a lower number of pedestrians that exited the measurement area. This metric is used for all scenarios except the bottleneck scenario because, contrary to the other scenarios, the bottleneck scenario does not have a fixed end time but a fixed number of pedestrians. So, in case of the bottleneck scenario the check is performed based on the time the last pedestrian exited the bottleneck whereby one would expect that a lower mean speed correlates with a longer time necessary for the last pedestrian to exit the bottleneck.

4.3.2 Results of the quantitative analysis

In this part the results for all seven scenarios are presented. First, the results of the tests with the maximum deviations are analysed to obtain which parameters deviate such that the model can be considered sensitive to changes in this parameter for that specific scenario. Afterwards, for the parameters that are considered sensitive, a further analysis is performed to assess how the sensitivity relates to the size of the deviation of the parameter's value.

As section 4.3 concludes, no unrealistic behaviour was found at the maximum deviations of 25%. So, for every scenario, the relevant boundaries (see Table 4.1) are all set to the 25% level and analysed accordingly.

Bidirectional straight - high

The A-D test showed that all speed distributions, for all parameters and boundaries, were significantly different from the default speed distribution. However, if one looks at the size of the deviations (see Table C.1), the data presented in Figure C.1 and the number of replications (N) it is clear that most deviations fall within the range one would expect to occur just because of the stochasticities (As Figure C.1 shows, at around 100 replications one would expect deviations of respectively 2.3% and 3.1% to occur for the mean and the standard deviation). Only two parameters resulted in deviations larger than would be expected (these values are highlighted in Table C.1) based on stochasticities. Hence, the sensitivity of the model to changes in the relaxation time and the viewing angle are investigated further below.

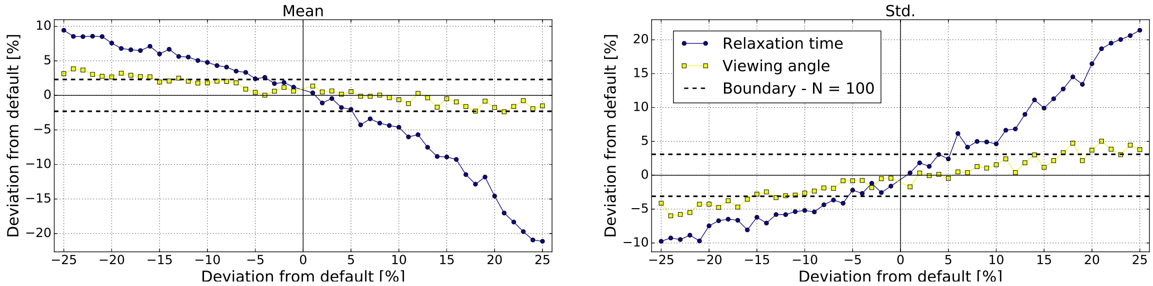


Figure 4.2: Sensitivities - Bidirectional high flow

Figure 4.2 visualises the model's response to changes in the relaxation time and the viewing angle. From the graphs it is clear that the model is far more sensitive to the relaxation time than to the viewing angle given the larger deviations. Furthermore, in the case of the relaxation time, increasing the value seems to have a stronger effect on the results than decreasing the value. An increased relaxation time results in less efficient flows, given the lower mean speed, whilst a decreased relaxation time results in a more efficient flow.

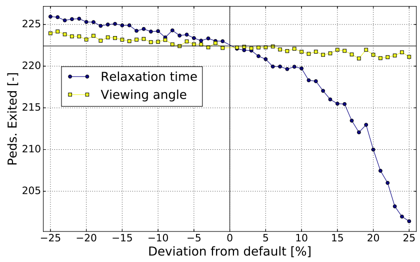


Figure 4.3: Sensitivities check - Bidirectional high flow

A change in the viewing angle primarily affects the results when the viewing angle is decreased. In this case the flow becomes a little bit more efficient. In case of an increased viewing angle there is no effect on the flow, however, a larger standard deviation does indicate some change in the underlying behaviour.

Lastly a check is performed in order to assure the mean speeds are indeed a good indicator of the flow. Figure 4.3 illustrates how many pedestrians exited the measurement area during the measurement period. From the graph it is clear that a decrease in the speed indeed correlates with a decrease in the number of pedestrian that exited the area and hence correlates with a lower flow. Furthermore, the graph also shows a larger effect when the relaxation time is increased than when it is decreased, which is consistent with the earlier findings.

So, in the case of the bidirectional high density scenario, the model is sensitive to changes in both the relaxation time and the viewing angle. However, in the case of the viewing angle, this in only the case when the deviation from the default value is near the maximum deviation.

Bidirectional straight - low

The results of the simulations showed, again, that all speed distributions were significantly different compared to the default. However, based on the data in shown Table C.2 and Figure C.2 one can conclude that only the increase in the relaxation time results in deviations larger than one would expect based upon the stochastic nature of the model. And, as Figure 4.4 shows, this is only the case for a large deviation from the default value.

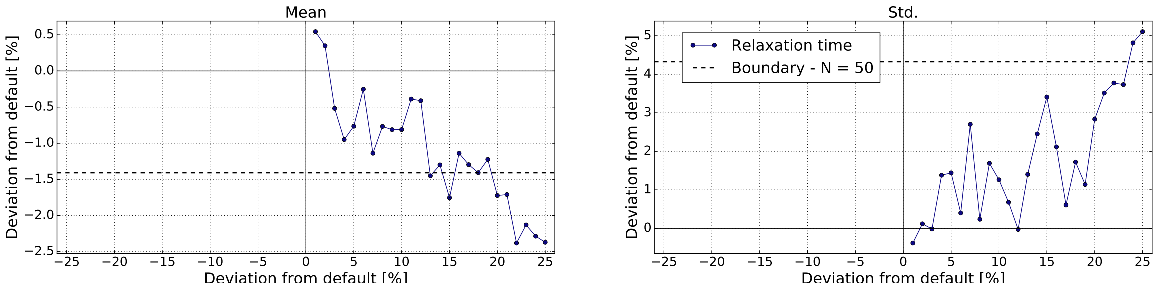


Figure 4.4: Sensitivities - Bidirectional low flow

The graph (Figure 4.5), displaying the number of pedestrians that left the measurement area, does show a similar pattern as the mean speed. An increase in the relaxation time correlates with a decrease in the speed and a decreased number of pedestrians leaving the measurement area.

So, the model does not seem to be sensitive to changes of 25% of most of the parameters except for the increase in the relaxation time which does lead to a decreased efficiency of the flow.

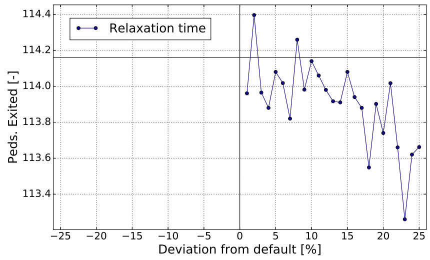


Figure 4.5: Sensitivities check - Bidirectional low flow

Bottleneck

The A-D tests, performed on the results of bottleneck scenario, did show that all speed distributions differed significantly when compared to default. Based on the deviations at 25% (see Table C.3) and the influence of the seeds (see Figure C.3) it is clear that the model can be considered insensitive to changes of 25% in three of the seven parameters. These three are the side preference update factor, the preferred clearance and the field of view avoidance range.

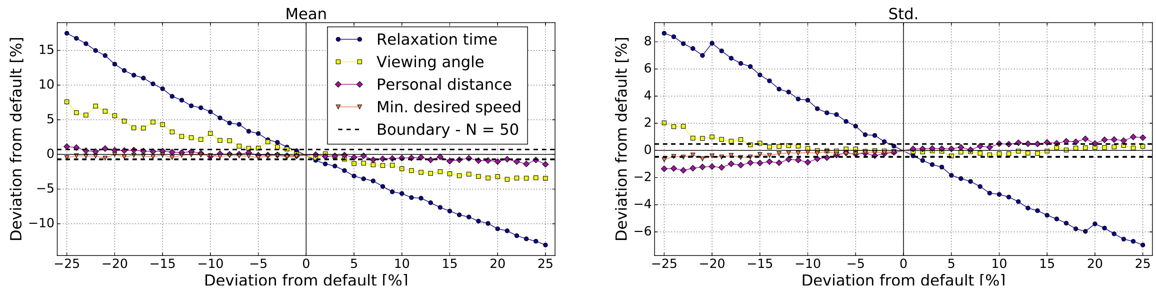


Figure 4.6: Sensitivities - Bottleneck

As Figure 4.6 illustrates, a change in the relaxation time has the largest impact on the results whereby it has to be noticed that contrary to the bidirectional scenarios decreasing the relaxation time has a larger impact than increasing it. The viewing angle also clearly impacts the results and, comparable to the relaxation time, this is also more apparent when the value is decreased compared to when it is increased. The personal distance and the minimum desired speed do also impact the results slightly, primarily when the increase or decrease of the value is large.

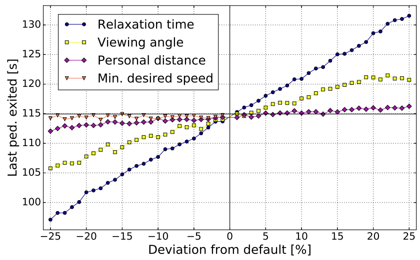


Figure 4.7: Sensitivities check - Bottleneck

Figure 4.7 shows that the speed correlates well with the time the last pedestrian exited the bottleneck given that a decrease in the mean speed indeed correlates with an increase in the time the last pedestrian exited the bottleneck.

So, in case of the bottleneck scenario the model is clearly most sensitive to changes in the relaxation time. Changes in the viewing angle also have a clear impact on the results. The personal distance and the minimum desired speed can have an impact on the results when the deviations from the default value are large. However, the model is not very sensitive to changes in either of these parameters.

Corner - high

In the case of the corner high density scenario the results of the simulations showed that all speed distributions differed significantly from the speed distribution obtained using the default values. However, when combining the data presented in Table C.4 and Figure C.4 it is clear that only two parameters differ more than one would expect based on the stochasticity of the model.

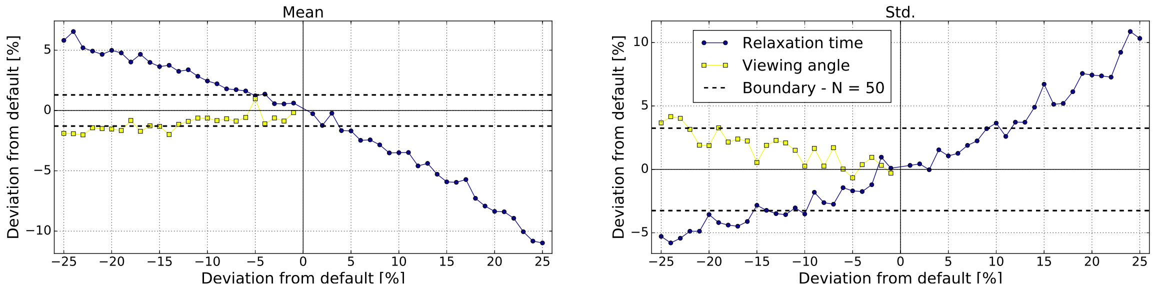


Figure 4.8: Sensitivities - Corner high flow

As Figure 4.8 shows, the model is primarily sensitive to changes in the relaxation time. This sensitivity is not symmetrical because an increase in the relaxation time clearly has a larger impact than a decrease. The viewing angle has a small impact when the deviation from the default value is large and the viewing angle is decreased.

Figure 4.9, again, shows that the changes in the mean speed correlate well with the number of pedestrians that exited the measurement area.

So, as is the case in all aforementioned scenarios, the model is most sensitive to changes in the relaxation time and comparable to the bidirectional scenarios the sensitivity is not symmetrical. The viewing angle has a small impact but the model clearly is not very sensitive to changes in this parameter.

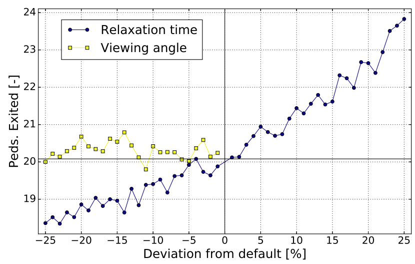


Figure 4.9: Sensitivities check - Corner high flow

Corner - low

All simulations of the low density corner scenario did result in a significant different speed distribution compared to the default simulation. Table C.5 and Figure C.5, however, show that only a change in the relaxation time had an effect larger than one would expect given the stochasticity of the model.

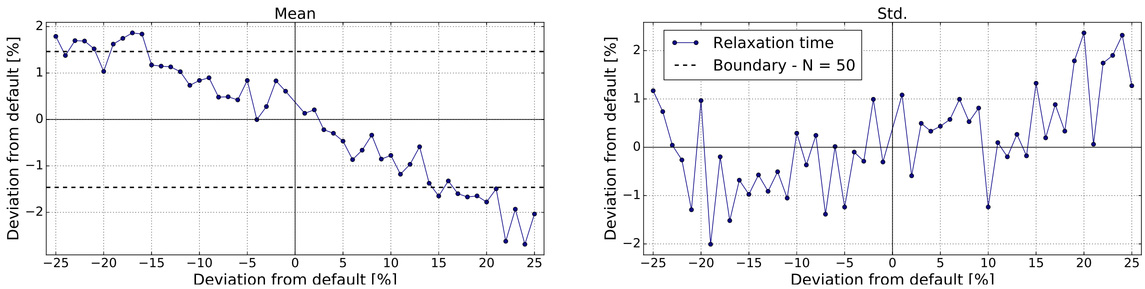


Figure 4.10: Sensitivities - Corner low flow

Figure 4.10 shows that, only when the deviations are near the maximum deviation, the change is larger than one would expect. Furthermore, this only holds for the mean and hence it is clear that in the case of this scenario the model is not very sensitive to any change in the parameters smaller or equal to 25%.

The graph in Figure 4.11 shows, again, a good correlation between the change in the mean speed and the number of pedestrians that exited the measurement area.

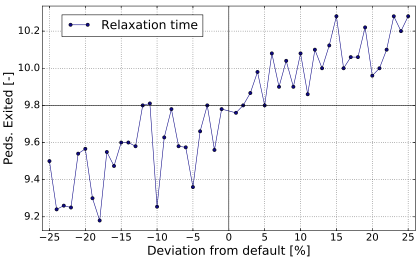


Figure 4.11: Sensitivities check - Corner low flow

T-junction - high

As is the case for all aforementioned scenarios, all simulations did result in significantly different speed distributions. In four of the seven parameters these differences were also larger than one would expect based on the stochasticities of the model. The preferred clearance, the minimum desired speed and the FoV avoidance range did not seem to have any impact on the results as can be derived from Table C.6 and Figure C.6.

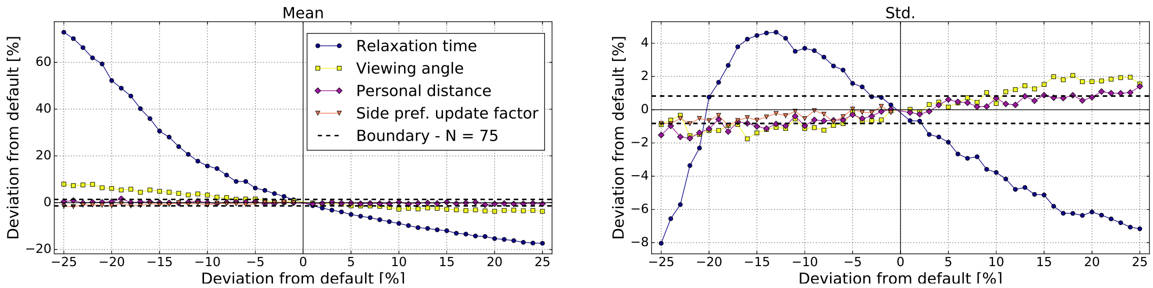


Figure 4.12: Sensitivities - T-junction high flow

As Figure 4.12 shows there is a clear difference regarding how much each parameter impacts the results. The relaxation time is clearly the parameter to which the model is most sensitive in this scenario. The sensitivity to the relaxation time is, again, not symmetrical and a decrease in the relaxation time clearly has a larger impact than an increase. This asymmetry is comparable to the one found in the bottleneck scenario. The viewing angle also has a clear impact on the model results, although this impact is clearly smaller than is the case with the relaxation time. Comparable to the relaxation time, and also to the bottleneck scenario, a decrease in the viewing angle has a larger impact than an increase in the viewing angle. The other three parameters have some impact on the result, however, this is primarily the case when the deviations are large.

The graph in Figure 4.13 does indeed show that as the mean speed increases the number of pedestrians that exited the measurement area also increases. However, the asymmetry found in the mean speeds does not exist in case of the number of pedestrians exited. A possible explanation for this discrepancy could be that as the relaxation time decreases the pedestrians keep walking at higher speeds, however, the behaviour changes such that on average their path is longer. The graph of the standard deviation does indicate a change in behaviour when the relaxation time in decreased below -12%.

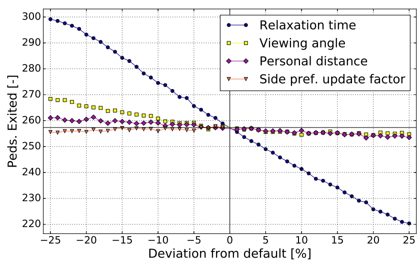


Figure 4.13: Sensitivities check - T-junction high flow

Overall, in the case of this scenario, the model is most sensitive to changes in the relaxation time. Changes in the viewing angle also clearly impact the results however the size of this impact is much smaller compared to the relaxation time. The personal distance, the side preference update factor and the minimum desired speed do impact the results when the deviation from their default value is near the maximum deviation. However, the model is not very sensitive to changes in these parameters.

T-junction - low

The simulations using the maximum deviation resulted in speed distributions that were all significantly different from the default speed distribution. Out of the six parameters tested, only changes in the relaxation time resulted in deviations that were larger than one would expect based on the stochasticity of the model. As Table C.7 and Figure C.7 show the other five parameters did not differ more than one would expect even when their values were increased or decreased by 25%. Figure 4.14 shows that an increase in the relaxation time has a smaller impact on the results than a decrease. This is opposite of the finding in the high density case of this scenario.

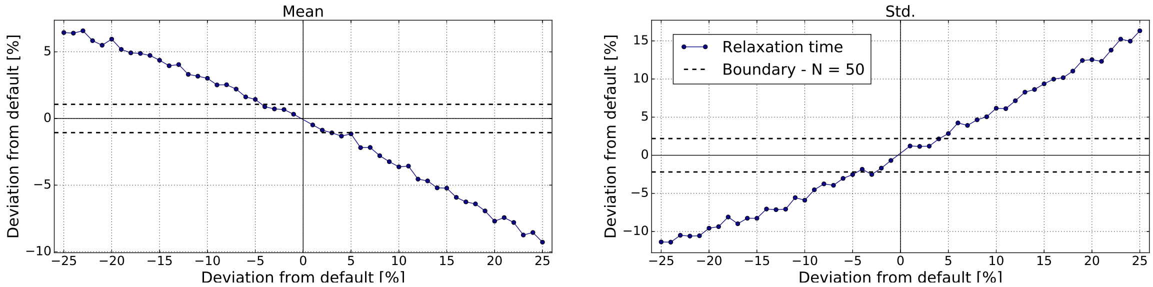


Figure 4.14: Sensitivities - T-junction low flow

If one compares the mean speeds with the number of pedestrians who exited the measurement area (see Figure 4.15) it is clear that they correlate as one would expect. An increased relaxation time results in a decreased mean speed and a decreased number of pedestrians who exited the area. Contrary the the high flow case the graphs show the same asymmetry.

So, in the case of the low density t-junction scenario the model is only sensitive to changes in the relaxation time.

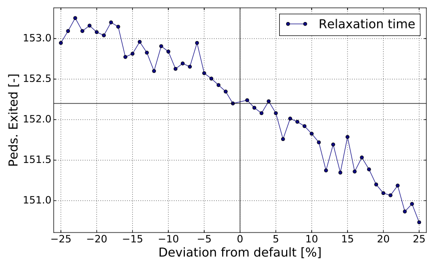


Figure 4.15: Sensitivities check - T-junction low flow

4.3.3 Conclusions on the quantitative analysis

Based on the results of all seven scenarios a number of conclusions can be drawn about the sensitivity of the model to changes in the values of the parameters.

- Firstly, the relaxation time is the only parameter to which the model is sensitive in all seven scenarios. In all cases the mean speed increases as the relaxation time decreases and vice versa. However, how sensitive the model is to a change in this parameter differs per scenario and also the asymmetry differs.
- Secondly, the model did not show any sensitivity to changes in either the preferred clearance and the FoV avoidance range. It has to be noted of course that this only holds for deviations of up to 25% percent given that larger deviations were not tested.
- Thirdly, out of the four remaining parameters that showed some significant impacts in one or more of the scenarios, the model is clearly more sensitive to the viewing angle than to the other three. However, the sensitivity of the model to the viewing angle is clearly smaller than its sensitivity to the relaxation time. The other three parameters, the personal distance, the side preference update factor and the minimum desired speed only showed some significant impact at large deviations and hence the model is not very sensitive to changes in these parameters.
- Overall, the speed distribution does seem to represent the flow well except in the case of the high flow t-junction scenario where a discrepancy can be found between the size of the impact a lower relaxation time has on the mean speed and the number of pedestrians that exited the measurement area.

4.4 Conclusions

In this chapter a sensitivity analysis was performed with the goal to get insight into how sensitive the model is to changes in the operational parameters. These result are necessary to define the search space of the calibration. In total seven different parameters were tested using seven different scenarios and a three step approach.

The qualitative analysis showed that deviations of 25% percent of the default value did not lead to unrealistic behaviour for any of the parameters in any of the scenarios. So, based on those results the boundaries for all parameters were set at the maximum deviation of 25%.

The quantitative analysis showed that the model was not sensitive to changes of 25% in the values of the preferred clearance and the FoV avoidance range. Out of the other five parameters, the relaxation time was clearly the parameter to which the model is most sensitive. The model is also sensitive to changes in the viewing angle, however, less sensitive than it is to changes in the relaxation time. Changes to the personal distance, the side preference update factor and the minimum desired speed only showed some significant impact in some scenarios at large deviations and hence the model is not very sensitive to changes in these parameters.

The quantitative analysis also showed differences between the scenarios regarding the sensitivities. These differences are on multiple levels, namely:

- To which parameters the model is sensitive.
- How sensitive the model is to the different parameters.
- How the sensitivity is related to the size of the deviation from the default value.

So, it is clear that the sensitivity of the model to parameter changes depends on the scenarios used. Hence, during a sensitivity analysis of a pedestrian simulation model one should use multiple scenarios to get a complete insight into the sensitivities.

Overall it can be concluded that the model is primarily sensitive to changes in the relaxation time and the viewing angle and hence these are the most important parameters to take into account during the calibration. Furthermore, it can also be concluded that one needs to use multiple scenarios to obtain complete insight into the sensitivities.

5 | Multiple-objective calibration

This chapter discusses the calibration of the model using different combinations of objectives. The main goal is to explore how different choices regarding the objectives influence the calibration.

The chapter discusses the following elements: In section 5.1 the methodology is discussed which explains how the calibration is performed. This is followed by a discussion of the results of the individual objectives in section 5.2. Section 5.3 explores how different choices regarding the objectives impact the calibration result by comparing different combinations of objectives to each other. In section 5.4 the methodology is systematically discussed in order to obtain insight into how changes in the methodology could potentially affect the results. Lastly, section 5.5 discusses the practical implication of the results.

5.1 Calibration methodology

In this section the calibration methodology is discussed. First, an overview will be given of the elements which are part of the multiple-objective calibration framework. After this, all these elements are discussed in more detail.

In section 2.2 nine elements have been identified which are involved in calibrating a pedestrian model using multiple objectives. These are:

1. Scenarios
2. Metrics
3. Objective functions and comparison methods
4. Optimization method
5. Stopping criteria
6. Search space definition
7. Stochasticities
8. Input definition
9. Reference data

Some of these elements are discussed simultaneously given their strong interdependency. The first subsection (subsection 5.1.1) discusses the scenarios in combination with the reference data and the input definition. Subsection 5.1.2 discusses the metrics followed by the discussion of the objective functions and comparison methods in subsection 5.1.3. In subsection 5.1.4 the choice for the optimization method is discussed including the choice of the stopping criteria and the definition of the search space. Lastly, in subsection 5.1.5 it is discussed how the stochastic nature of the model is dealt with.

5.1.1 Scenarios, reference data and input definition

Chapter 3 already concluded that the calibration will be performed using seven different scenarios and the data from the HERMES project. Hence, this subsection will only discuss the implementation of the scenarios in PD.

As the review of the literature in subsection 2.2.6 concluded, it is important that the input to the model matches the reference data as closely as possible. The four input types, identified in subsection 2.2.6, are discussed in more detail below.

Geometry of the infrastructure

The geometry is implemented such that it exactly matches that of the experiment (i.e. both the geometry of the walkable area and the location of walls). Figs. D.1 to D.4 give a more detailed view of the implementation.

Route choice

In the case of PD, the routes the pedestrians intend to follow are influenced by four elements: The route choice algorithm, the locations of the origins and destinations, the location of waypoints and the parameters of the route following algorithm. As concluded in subsection 3.3.1, all OD-pairs in all scenarios have only one possible global route. Hence, the shortest path option of the algorithm, implemented in PD, is used.

D.1 to D.4 show the locations of all origin and destination areas. Every pedestrian is assigned an origin and a destination area. A randomly chosen point within these areas determines the exact origin and destination of the pedestrian. The exact location of the destination can change slightly whilst the pedestrian traverses its path. However, it will always be located within the assigned destination area.

In both the bidirectional scenarios and the t-junction scenarios waypoints are used to ensure a better match between the trajectories in the simulations and the data. The decision to use waypoints and the choice of their location are based upon visual comparisons of the trajectories of the data and the simulations.

The last of the four elements, influencing the route of a pedestrian, are the parameters of the route following algorithm. As Table 3.5 showed, the route following algorithm has four parameters. As subsection 3.3.2 concluded, two of those four parameters, the maximum shortcut distance and the side clearance factor, should be used to fix unrealistic local path finding behaviour. The visual comparison of the trajectories did not show unrealistic local path finding and hence these parameters remain at their default value. The other two parameters, the preferred clearance and the side preference update factor, also use their default value because, as subsection 5.1.4 will explain, they are not used to define the search space.

Demand patterns

The demands for each OD-pair in each scenario are determined based on the cumulative curves of the accompanying data set. In Table 5.1, the demands obtained from the cumulative curves are presented. For all scenarios the demands are assumed to be constant. A test, comparing the cumulative curve with the curve based on the constant demand, resulted in R^2 values between 0.919 and 0.999. Hence, the use of a constant demand is considered to be a good approximation of the actual demand pattern. In the case of the bottleneck scenario, there is no constant demand. But, at the start of the scenario the start area is filled with 349 pedestrians which is exactly the same number as in the data.

Table 5.1: Overview of the OD-pairs and the demand per OD-pair

(a) Bidirectional scenarios

	O	D	Demand [ped/s]	
			High	**Low**
1.	1	a	1.7055	0.881
2.	1	b	1.7055	0.881
3.	2	c	1.6455	0.725
4.	2	d	1.6455	0.725

(b) Corner scenarios

	O	D	Demand [ped/s]	
			High	**Low**
1.	1	a	3.242	1.441

(c) T-junction scenarios

	O	D	Demand [ped/s]	
			High	**Low**
1.	1	a	2.213	1.348
2.	2	a	2.432	1.382

Speed distribution

Similar to the demand patterns, the distribution of the preferred speed speeds is also obtained from the reference data such that, in the simulation, it matches the data. During the experiments, two data sets were collected with the intention of measuring the preferred speed of the participants. These are the EOF-300-1 and the EOF-300-2 experiments (see (Keip & Ries, 2009)). The participants walked through the corner on their own whilst being instructed to walk at their preferred speed. In an area of 2.5 meters long, starting 1 meter after the pedestrian has rounded the corner, the instantaneous speeds of every pedestrian is measured. The preferred speed of a pedestrian is the mean of the measured instantaneous speeds. Figure 5.1 shows the measured speeds in the histogram.

In PD the default speed distribution is a single triangular distribution (i.e. no distinction is made between groups based on properties such as gender and age). During the simulation also a single triangular distribution is used to represent the preferred speed distribution of the population. The choice was made to use a single distribution because the data and accompanying information did not contain enough information to make any distinction between groups of participants.

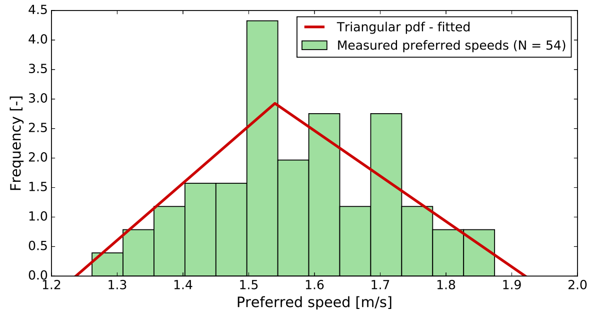


Figure 5.1: Measured and fitted free speed distributions

Figure 5.1 shows the fitted triangular distribution. The fit resulted in the following parameter values:

- Min: 1.24 m/s
- Mean: 1.54 m/s
- Max: 1.92 m/s

5.1.2 Metrics

A part of the multiple-objective framework is the usage of multiple metrics. Subsection 2.2.3 concluded that during the calibration quantitative metrics should be used, preferably covering all three aggregation levels (macro, meso and micro). In this research the choice is made to use two metrics at the macroscopic level, the flow and the spatial distribution, and two at the mesoscopic level, the effort distribution and the travel time distribution. Microscopic metrics, usually the trajectories, are not used for three reasons. Firstly, calibrating using trajectories requires a different approach, not compatible with approaches that are used when calibrating based on macro and mesoscopic metrics. Secondly, current approaches for calibrating a pedestrian model based on trajectories do not deal with the stochastic nature of the model and developing an approach that could deal with the stochastic nature of the model when calibrating the model based on trajectories is outside of the scope of this research. And thirdly, as subsection 2.2.3 concludes, given the goal of the calibration, defined in the scope (section 1.2), and the finding by Campanella (2016) that a calibration based on trajectories did result in inaccurate predictions on the macroscopic level, in this research macroscopic metrics take priority over the microscopic metrics.

The four chosen metrics are chosen such that at both aggregation levels they measure different aspects of the flow. The flow and travel time distribution are included as they are commonly used performance indicators and the spatial distribution and the effort are included as measures that give more insight into the behaviour and the resulting spatial patterns that do not necessarily bear a strong relation to the flow and travel time. By including these different types of metrics at both aggregation levels insight can be gained into how the calibration is affected by the type of metric/metrics used.

In the remainder of this subsection the four metrics are explained in more detail including why they are chosen out of the many other possibilities. However, first it is explained why a measurement area and a measurement period are used when collecting the data.

Measurement area and period

The measurement area and period together determine where and when measurements are collected. Limiting the area and time frame, within which the measurements are obtained, is necessary because pedestrians are not expected to exhibit the behaviour, relevant to the particular scenario, in all parts of the walkable area and during every moment of the simulation. The assumption is made that a pedestrian can be expected to show relevant behaviour in a given area and time frame, when a) the flow in the area is such that it represents the movement base case the scenario is intended to capture and b) the density level in the area is at the level it is intended to be. Capturing only the relevant behaviour is important because without doing so one is not able to capture any differences between the scenarios based on the properties that make them different (i.e. the movement base case and the density level). Parts outside the measurement area are a necessary part of the simulation though to make sure that the simulation does represent the experiments, whose data is used, as closely as possible (i.e. and thus also to make sure that the behaviour in the measurement area is as representative as possible).

As an example of why these limitations are used one can take a look at the bidirectional high density scenarios. In the case of the bidirectional high density scenario, the measurement area is limited to the eight meter long stretch of corridor since this is where the flow is bidirectional. Outside of this measurement area the flow is no longer really bidirectional but either unidirectional or crossing. The measurement period starts only after the measurement area has reached the required density level. Any data from pedestrians that traversed the area before this time is discarded given that they experienced a density level below the required density level and their behaviour is thus not representative for the scenario.

In figs. D.1 to D.4 one can find the locations, shapes and sizes of the measurement areas for all scenarios. The exact measurement period differs per metric as Table E.1 in appendix E indicates. In appendix E it is explained, in detail, how these measurement periods are determined.

Flow

In all seven scenarios the average flow is measured along a certain measurement line (see figs. D.1 to D.4) during a certain measurement period according to Equation 5.1. The flow is chosen as a macroscopic metric to check how well the model is capable of reproducing the throughput in different situations. The reason for using the average flow instead of the flow over the time is twofold. Firstly, after the warming-up period the flow in the reference data is relatively stable. Secondly, it is easier to compare two average flows than two timeseries of flows. The average flow is calculated as follows:

$$\bar{q}_i = \frac{N_i}{\Delta t * l} \quad \text{[ped/s/m]} \tag{5.1}$$

Where N_i is the number of unique pedestrians with main travel direction i that passed the line in the direction equal to the main travel direction and during the duration of the measurement period (Δt). The flow in normalized to a flow per meter of measurement line whereby l is the length of the measurement line. This is done to ease to comparison between scenarios.

As mentioned N_i is the number of *unique* pedestrians that passed the line. This means that a pedestrian is only counted once, namely the first time it passes the line. This is done to prevent artificially high flow values which could occur if, for example, pedestrians pass the line twice because they were pushed back. This would primarily be a possible problem in the bidirectional high density scenario where either:

a An unresolved head-on interaction near the flow line causes a collision in which one of the pedestrian bounces back, or gets pushed back, over the flow line and afterwards passes the flow line a second time. Or,

b The flow has broken down preventing pedestrians to move forward (i.e. towards their destination) whereby the possibility exists that pedestrians standing closely to the flow line are still moving back and forth over the flow line without actually moving forward in a significant way.

Distribution over space

The distribution over space measures how the pedestrians are distributed over the measurement area. A grid of 0.4 x 0.4 m overlays the measurement area and for every cell the percentage of the time it is occupied is determined as follows:

$$F_i = \frac{N_{occ}}{N_{steps}} \quad \text{[-]} \tag{5.2}$$

Where N_{occ} is the number of time steps cell i is occupied by one or more pedestrians (based on the centre point of the pedestrians) and N_{steps} is the number of time steps taken into account. Duives (2016) showed that the model, used in that particular study, was not able to accurately reproduce the spatial distribution patterns. Hence, it is interesting to see if this is also the case in this research given the use of another model and other data sets.

Effort

The effort metric captures how much effort it takes a pedestrian to traverse the measurement area.

This metric is primarily used to check how well the model is capable of reproducing the underlying behaviour of the pedestrians. The better the model is able to reproduce the effort it takes a pedestrian to walk from A to B the more likely it is that the model captures the underlying behaviour well.

The effort is defined as the average change in velocity per time step. So, the more a pedestrian has to change its velocity during the traversal of the measurement area the more effort it thus takes. The effort is calculated as follows:

$$eff_i = \frac{\sum_{n-1}(|v_{j;x} - v_{j-1;x}| + |v_{j;y} - v_{j-1;y}|)}{n-1} \quad [\mathrm{m/s}] \tag{5.3}$$

Where $v_{j;x}$ and $v_{i;y}$ are respectively the speed in the x and y-direction at time step j and n the number of time steps. The speeds are obtained by differentiating the positions:

$$v_{j;x} = \frac{x_j - x_{j-1}}{t} \quad [\mathrm{m/s}] \tag{5.4}$$

Where x_j is the x-position at time step j and t is the duration of the time step. Because the reference data and the simulation have different time steps the trajectories in the reference data are mapped to the time steps of the simulation data such that the time steps match. The choice to interpolate the reference data instead of the simulation data is twofold. Firstly, because the simulation data contains about 1.27 million data sets and the reference data only 7 it is far more practical from a computational point of view to interpolate the reference data. Secondly, the time step of the reference data is smaller than the time step of the simulation data and it is considered better to lose some accuracy than to gain it artificially. The scipy CubicSpline[1] method is used for the interpolation.

For both the reference data as the simulation data the individual effort measurements are combined into a distribution whereby in the case of the simulation data the distribution includes the data from all individual measurement from all replications. The effort of a pedestrian is only included if the pedestrian entered the measurement area between the start of the measurement period and the end of the measurement period minus a buffer. This buffer is the average time it took a pedestrian in the experiments to traverse the measurement area. The use of the buffer prevents measurements to be taken into account from pedestrian who only just entered the measurement area. These effort measurement are discarded because they might not be representative for the effort it takes to traverse the whole measurement area. The buffer is not applied in the case of the bottleneck as is explained in more detail in appendix E.

Travel time

The travel time is the time it takes a pedestrian to traverse the measurement area. It is included because, as Table A.1 shows, it is clearly the most commonly used mesoscopic metric. Hence, it is good to get insight into how well the model can reproduce the travel time found in the data. The travel time of a single pedestrian is determined as follows:

$$TT_i = \frac{t_{end} - t_{start}}{l_{ref}} \quad [\mathrm{s/m}] \tag{5.5}$$

Where t_{start} and t_{end} are respectively the time the pedestrian first entered the measurement area and time the pedestrian left the area. l_{ref} is the average length of the path in the measurement area, as obtained from the reference data. By dividing the travel time by this path length one normalizes the travel time to a travel time per meter. This makes a comparison between different scenarios with different average path lengths easier.

Again, all measurements are combined into a distribution whereby only the travel time of those pedestrians who successfully traversed the whole measurement area during the measurement period are taken into account. The decision of only taking into account those pedestrian who successfully traversed the whole measurement area could potentially skew the results. This could primarily happen in the case of the bidirectional high density scenario where, at some point in the simulation, the flow could break down. This could cause many pedestrians to get stuck in the measurement area till the end of the measurement period. This, in turn, could potentially lead to a gross underestimation of the travel time given that the distribution would primarily be made up of those pedestrian who traversed

[1] https://docs.scipy.org/doc/scipy-0.19.0/reference/generated/scipy.interpolate.CubicSpline.html

the area before the breakdown of the flow. Hence, it should be checked if the problem described above does influence the results.

5.1.3 Objective functions and comparison methods

This subsection answer two questions, namely: 1) What objective function to use per metric?, and 2) How to determine a optimal parameter set from two or more objective functions? As the next subsection (5.1.4) will show, the chosen optimization algorithm is not aimed at finding the set of Pareto optimal solutions and hence the objectives have to be combined into a single objective. This changes the second question to: How to combine the results of two or more objective functions[2] into a single objective function.

In this research the multiple objectives are combined into a single objective using the weighted sum method (Zak & Chong, 2013). This is in line with research by Duives (2016), the only example in the literature using both multiple metrics and multiple scenarios to calibrate a pedestrian model. The choice of the method leads to two other choices which need to be made. Namely, which weights to use and how to normalize the objectives of the different metrics such that they can be added up in a meaningful way. The normalization of the errors is necessary because the metrics have different units and different orders of magnitude.

The normalization method used in this research is based on the ratios between the different metrics in the reference data. For a detailed explanation of this method and the choice to use it, the reader is referred to appendix F. In line with previous research the distance between the reference data and the simulation results is given by a squared error. So, the objective function for a given metric and scenario is given by the normalized Squared Error (SE) and is determined as follows:

$$SE_{norm}(\theta) = \left(\frac{M_{sim}(\theta) - M_{ref}}{M_{norm}}\right)^2 \tag{5.6}$$

Table 5.2: Normalization values

	$\mathbf{M_{norm}}$
Flow	1.0
Spatial distribution	0.18994
Effort - mean	0.99107
Effort - std.	0.20728
Travel time - mean	0.04345
Travel time - std.	0.00953

Where $M_{sim}(\theta)$ is the metric's value obtained from the simulation using parameter set θ, M_{ref} the metric's reference value according to the data (See appendix G) and M_{norm} the value used for the normalization as found in Table 5.2. In the case of the macroscopic metrics, M_{sim} is the mean flow or cell occupation for the given number of replications. In the case of the mesoscopic metrics M_{sim} is either the mean or the standard deviation of the distribution which combines the measurements of all replications. Squaring the errors makes sure that they are all positive and hence makes it easier to add them up when multiple objectives are combined. Below it is explained in more detail how the normalized SE is computed per type of metric.

The objective of the two macroscopic metrics is determined as follows:

$$SE_{norm;macro}(\theta) = \frac{1}{m}\sum_j \left(\frac{\frac{\sum_i M_{sim;i;j}(\theta)}{n} - M_{ref;j}}{M_{norm}}\right)^2 \tag{5.7}$$

Where n is the number of replications and θ the parameter set. In the case of the flow m is the number of main travel directions ($m = 2$ in the case of the two bidirectional scenarios and 1 in all other scenarios). In the case of the spatial distribution m is the number of cells.

The objective of the two mesoscopic metrics is determined as follows[3]:

$$SE_{norm;meso}(\theta) = \frac{1}{2}\left(\frac{M_{sim;\mu}(\theta) - M_{ref;\mu}}{M_{norm;\mu}}\right)^2 + \frac{1}{2}\left(\frac{M_{sim;\sigma}(\theta) - M_{ref;\sigma}}{M_{norm;\sigma}}\right)^2 \tag{5.8}$$

[2]In this research an individual objective is defined by a unique combination of a scenario and a metric and, given the definition in subsection 2.2.4, an individual objective function thus describes the difference between the simulation and reference data given a certain parameter set for that given unique combination of a scenario and a metric

[3]The use of a combination of the mean and standard deviation to describe the distributions of the mesoscopic metrics was chosen over more advanced methods, such as the Kolmogorov-Smirnov (KS) test, because of two practical reasons. Firstly, performing the KS-test is computationally more expensive than simply calculating the mean and standard deviation and comparing it to the reference values. Secondly, using the KS test statistic is not compatible with the used normalization method.

The objective functions for a given set of metrics and scenarios are combined into a single objective function as follows:

$$O(\theta) = \frac{1}{N_s * N_m} \sum_s \sum_m SE_{norm;s;m}(\theta) \tag{5.9}$$

Where $SE_{norm;s;m}(\theta)$ is the value of the objective function of scenarios s and metric m for the parameter set θ and N_s and N_m are, respectively, the number of scenarios and metrics in the set. Due to the use of the SE, all values are positive and hence all resulting objective functions, both the individual and combined ones, only contain values larger or equal to zero. So, the smaller the value of the objective function the smaller the error and hence the better the Goodness-of-Fit (GoF) of the model to the data. Hence to find the optimal parameter set for a given set of scenarios and metrics one has to minimize the combined objective function.

5.1.4 Optimization method, stopping criteria and search space definition

In this research a grid search will be used to obtain the optimal parameter set. The use of the grid search has four advantages over more complex methods (i.e. all other methods mentioned in Table 2.2). Firstly, it is the only method that is certain to find the global optimum (given the level of precision defined by the grid). Secondly, one does not have to determine, beforehand, which metrics, objective functions and combinations of objective functions one want to use. Only the scenarios and search space have to be determined beforehand. This provides much more flexibility because one can research the use of additional metrics, objective functions and combinations based on the results without having to run more simulations. Thirdly, it is certain to cover the whole search space and hence insight into the shape of the surface of the objective space. And fourthly, the choice for the grid search also means that no stopping criteria is required. The only disadvantage of using a grid-search that all other method can potentially be faster assuming the same level of precision.

As chapter 4 concluded, the relaxation time and the viewing angle are the parameters to which the model is most sensitive and hence the most important parameters to take into account during the calibration. So, initially the search space consist of these two parameters. However, whilst testing the implementations of the scenarios it became apparent that, in the case of the bidirectional high density scenario, the default radius was problematic. It was problematic because, when using the default value for the radius, the flow almost immediately went into a grid-lock situation whereby only very few pedestrians, or even none at all, were able to move in the direction of their destination. This remained the case even when lowering the values of the relaxation time and viewing angle, which, as the sensitivity analysis showed, has a positive effect on the efficiency of the flow. The only way to prevent this was to lower the radius of the circles representing the pedestrians. So, besides these the relaxation time and the viewing angle, the radius will also be included in the search space.

With these three parameters the search space is defined as follows:

- The upper and lower limits are determined by a deviation of $\pm 24\%$. In the case of the radius a deviation of -40% is used and the upper boundary is equal to the default value.
- The step size is 3% of the default value (in the case of the radius it is 4% of the default value).

The combination of the step size and the limits was chosen taking the following four considerations into account:

1. The smaller the step size the higher the precision level of the calibration
2. The larger the limits the less likely it is that the optimal value falls outside of the limits of the search space (taking into account a maximum deviation of 25% as used in the sensitivity analysis)
3. The larger the search space the more simulations need to be run during the grid search (i.e. the longer it takes to finish the grid search)
4. The combination of the time it takes the computer, available to this research, to run all simulations for a single point in the search space (i.e. all replications for all seven scenarios) and the total time available to this research to run all simulations.

The reason for using other limits and another step-size for the radius is as follows. The radius is not considered a parameter directly involved in the operational behaviour of the agents but rather as a property of the agents that indirectly influence the behaviour. Hence, as explained above, it is only added for practical reasons whereby the -40% deviation was based on the requirement that, while using the default values (other than the radius), grid-lock would not occur during the simulation of the bidirectional high scenario. The 4% step is based on the practical considerations described above.

Table 5.3: Search space definition

	Default value	Lower limit	Upper limit	Step size
Relaxation time [1/s]	0.5	0.380	0.620	0.015
Viewing angle [degree]	75	57	93	2.25
Radius [m]	0.239	0.14340	0.239	0.00956

5.1.5 Stochasticities

Based on the findings in section 3.4 it is decided that, during the calibration, a fixed set of seeds and hence a fixed number of replications per scenarios will be used. This decision is made because, as section 3.4 showed, even with a high number of replications it remains questionable if small differences are caused by differences in the parameter values or a difference in the exact order of the seeds. This is undesirable during the calibration given that small difference can potentially lead to very different optimal parameter sets.

So, for every scenario a fixed number of replications is used whereby the exact number of replications depends on the scenario. The number of replications was determined using the following methodology: For every scenarios 200 replications[4] are run for 18 different parameter sets. These 18 different parameter set are the unique combinations of the upper limit, the lower limit and the default values of all three parameters included in the search space. In all cases, the same order of seeds was used. This same seed order is also used during the calibration. Based on the convergence of the speed distribution, the number of replications was determined. The number of replications for every scenario can be found in Table 5.4.

Table 5.4: Nr. of replications per scenarios

	N
Bidirectional - high	100
Bidirectional - low	100
Bottleneck	30
Corner - high	50
Corner - low	40
T-junction - high	40
T-junction - low	40

All in all, this method assures that, within a scenario, the differences in the objective functions are not caused by the stochastic nature of the model. However, due to the use of the fixed order of seeds and the fixed number of replications the possibility remains that the results of this research would be slightly different if another order of seeds would have been used.

5.1.6 Running the calibration simulations

Using the procedures described above, more than 1.27 million simulation are run spanning the 7 different scenarios, the 3179 points of the search space and, depending on the scenario, 30 to 100 replications per point in the search space. All simulations have a duration of 1.5 minutes which, based on the measurement periods determined in appendix E, should provide enough simulated time to include both the measurement time and a warming-up period to fill the infrastructure. In the case of the simulations of the bottleneck scenario a duration of 2 minutes is used to assure that all pedestrians have passed the bottleneck before the end of the simulation. The results of these simulations and the subsequent calculation of the different individual and combined objectives are discussed in the next two sections.

[4] Based on the results of section 3.4 whereby is was decided to use 400 in the case of the bidirectional high scenario

5.2 Analysis of the individual objectives

In this section the results of the individual objectives (a combination of a single scenario and a single metric) are discussed. In total there are 42 individual objective spaces (7 scenarios and 4 metrics whereby in the case of the mesoscopic metrics the mean and the standard deviation are discussed separately). The goal of the analysis is to gain insight into how well the model can reproduce the reference data, within the given search space, for a given metric and scenarios. To also gain insight into whether the simulation, for the given search space, generally over or underestimates the results, the non-squared and non-normalized error is used (i.e. $E = M_{sim} - M_{ref}$). Furthermore, for this analysis it is not of interest how the errors are distributed over the objective space. Hence, as Figure 5.2 shows, the distributions of the errors are represented by box plots whereby all box plots are made up of 3179 points (i.e. one for each unique parameter sets). The following subsections will discuss the result per metric.

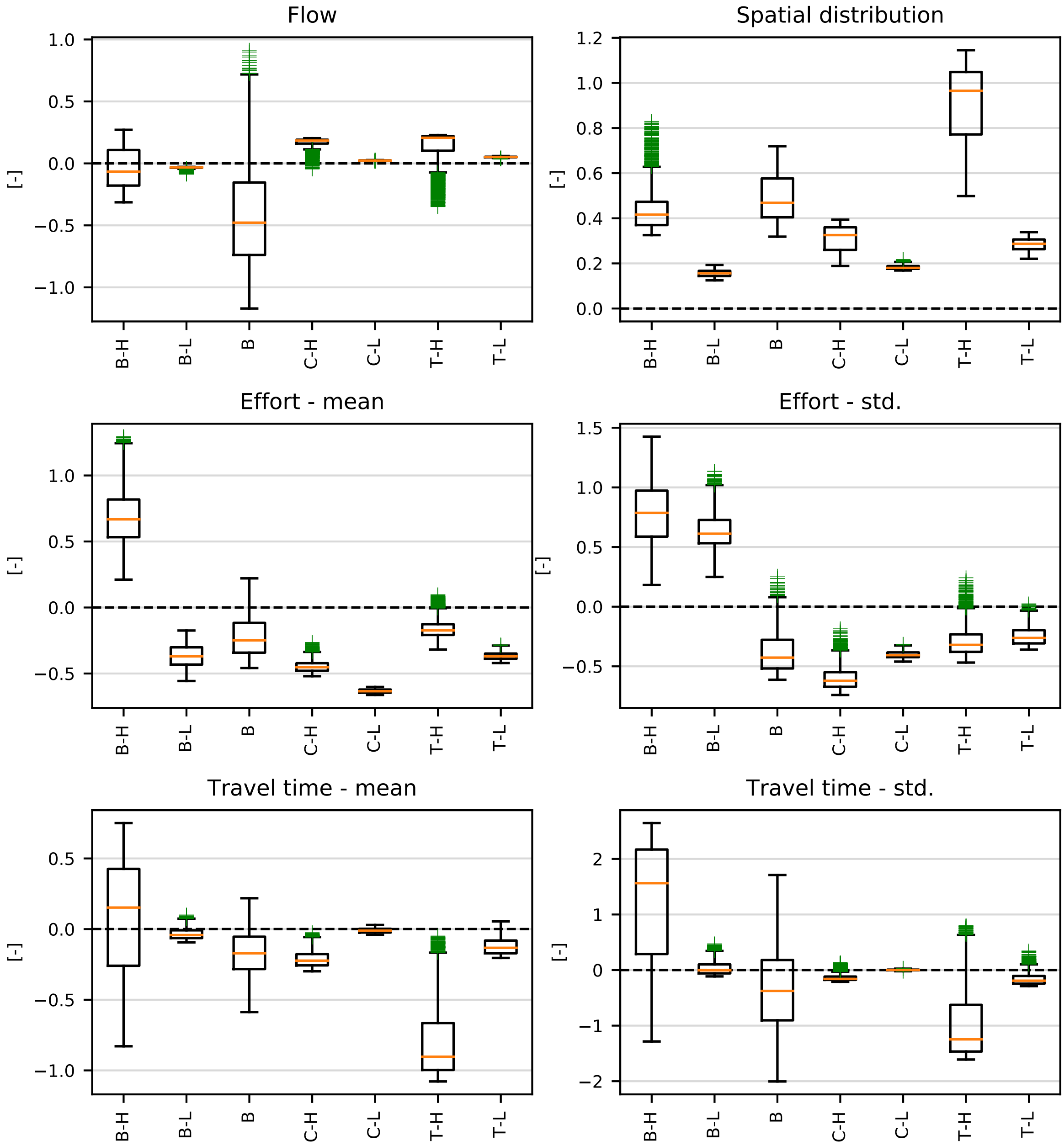


Figure 5.2: Box plots of the non-squared, non-normalized errors per scenario and metric. The acronyms identify the scenarios (i.e. B-H = bidirectional high, B-L = bidirectional low, B = bottleneck, C-H = corner high, C-L = corner low, T-H = t-junction high, T-L = t-junction low).

5.2.1 Flow

Figure 5.2 shows that in the case of the bidirectional low density scenario the flow is always slightly underestimated (less than 0.10 ped/m/s) regardless of the parameter set. In the case of the corner and t-junction low density scenarios it is the opposite because in those two cases the flow is always slightly overestimated. In all other cases (the four high density cases), the flow is both under and overestimated depending on the parameter set.

What is also noticeable is the difference between the low and high density scenarios. In the case of the high density scenarios the errors vary much more (i.e. the objective space is less flat) compared to the low density scenarios. Furthermore, there is also a difference between the movement base cases. In the bidirectional and bottleneck cases, the flow is generally underestimated whilst in the case of a corner or t-junction the flow is generally overestimated.

So, for the given parameter sets, the model should, per individual scenario, be capable of reproducing the flows found in the data within a range of 0.04 ped/s/m.

5.2.2 Spatial distribution

The errors of the spatial distribution, illustrated in Figure 5.2, are the average absolute errors per cell (hence they are always positive). The figure shows that the errors are generally smaller in the low density scenarios and that especially in the case of the t-junction high density scenario the error is generally large compared to the rest.

So, the best possible fit of the model to the data always results in a small average error per cell. However, it is also clear that the size of this smallest possible error depends on lot on the scenario.

5.2.3 Effort

Figure 5.2 shows that the mean effort is generally underestimated with the exception of the bidirectional high density scenario where it is always overestimated. Furthermore, in all low density scenarios and the corner high density scenarios it is always the case that the mean effort is underestimated.

Regarding the standard deviation of the distribution of the effort (over the pedestrians not the objective space), another pattern is found. In the case of the two scenarios of the bidirectional movement base case, the model always overestimates the standard deviation of the distribution. In all other five cases the standard deviation is generally underestimated and in the two scenarios of the corner movement base case this is always the case.

So, in most cases the mean effort will be over or underestimated by at least 15% (compared to the values in Table G.2). And, taking into account the standard deviation, only in the cases of the bottleneck and t-junction high density scenarios is it possible that the distribution of the simulations fits the data closely. Combining the relative errors of both the mean and standard deviation shows this indeed to be the case. The smallest combined relative error, in the case of the two aforementioned scenarios, is about 2% whilst in the other five scenarios it is larger than 13%.

5.2.4 Travel time

The mean travel time is, as Figure 5.2 illustrates, generally underestimated with the exception of the bidirectional high density scenario which is more likely to overestimate the travel time. The figure also illustrates that, in general, the surfaces of the objective spaces of the high density scenarios are less flat given the larger variance in the errors. This same pattern can be found in the case of the standard deviation with the exception of the corner high density scenario. Combining the relative error of both the mean and standard deviation shows that the model can reproduce the travel time distributions of the bottleneck and the corner low density scenarios very well (smallest combined error below 0.2%). In the other five scenarios the smallest combined error is somewhere between 1% and 10%.

So, depending on the scenario, the model is capable of reproducing the travel time distributions with varying levels of accuracy. As subsection 5.1.2 states, the travel time could potentially be skewed, especially in the case of the bidirectional high density scenario. To investigate if this is the case the error of the mean travel time and the error in the number of pedestrians that make up the distribution are compared. If the results are not skewed one would expect a strong negative correlation (i.e. if the distribution of the simulation contains less pedestrians one would expect a higher mean travel time given that the lower number of pedestrians (given the same inflow) indicates a lower throughput and hence a lower speed). Figure H.1 shows this is indeed the case for the bidirectional high density scenario. The same check was performed for the other scenarios and it was found that also for these scenarios there was a very strong negative correlation. Given that the bottleneck scenario has a fixed number of pedestrians this is not relevant for this scenario.

5.2.5 Conclusions on the analysis of the individual objectives

So, as the data shows, the model is very well capable of fitting the flow to the data regardless of the scenario. For the other three metrics the degree to which the model is able to fit to the data depends on the scenario. One has to note that all findings above are based the individual objective spaces and that it remains the question if the model is capable of reproducing the data well if multiple objectives are combined. The next section will investigate different combinations of objectives and discuss how different choices of scenarios and metrics influence the calibration results.

5.3 Analysis of the combined objectives

In this section the results of different combinations of objectives will be discussed. The main goal of comparing the results of these different combinations is to answer the following questions:

1. How does the choice of movement base cases influence the calibration results? To recap the findings of section 2.1, two previous studies (Campanella et al., 2011; Duives, 2016) found that a). Calibrating using a single movement base case, compared to using multiple base cases, will result in better performance on that particular base case. And b). For general usage one does need to calibrate using multiple movement bases to capture all relevant behaviour.
2. How does the choice of level of density influence the calibration results?
3. How does the choice of metrics influence the calibration results? Duives (2016) found that different combinations of metrics clearly lead to different calibration results.

First, an overview will be given of the combinations used during the analysis. Secondly, a general analysis of the results is performed based on the obtained optimal parameter sets for all of the combinations. After this the results of different combinations will be compared to get insight into the three questions posed above. Lastly, the results will be reflected upon by comparing the findings with those of previous studies and the findings of the sensitivity analysis.

5.3.1 Combinations of objectives used during the analysis

Table 5.5 introduces the 16 different combinations of objectives which are used to answer the questions posed in the introduction of this section. The table shows there are five different types of combinations: 1) Combinations using a single scenario and all metrics. 2) Combinations using a single metric and all scenarios. 3) Combinations of all scenarios of the same density level using all four metrics. 4) Combinations of all metrics of the same aggregation level using all seven scenarios. And finally, 5) a combination of all metrics and scenarios.

Depending on the question that is tackled in the following subsections, different sets of these combinations will be used to answer the question. Which combinations are compared, is explained in every subsection.

Table 5.5: Tested combinations of objectives. The acronyms identify the metrics (i.e. Q = flow, SD = spatial distribution, Eff = effort, TT = travel time) and the scenarios (i.e. B-H = bidirectional high, B-L = bidirectional low, B = bottleneck, C-H = corner high, C-L = corner low, T-H = t-junction high, T-L = t-junction low).

Combination	Metrics				Scenarios						
	Q	**SD**	**Eff**	**TT**	**B-H**	**B-L**	**B**	**C-H**	**C-L**	**T-H**	**T-L**
Individual scenario - All metrics											
1. Bidirectional - high	x	x	x	x	x						
2. Bidirectional - low	x	x	x	x		x					
3. Bottleneck	x	x	x	x			x				
4. Corner - high	x	x	x	x				x			
5. Corner - low	x	x	x	x					x		
6. T-junction - high	x	x	x	x						x	
7. T-junction - low	x	x	x	x							x
Individual metrics - All scenarios											
8. Flow	x				x	x	x	x	x	x	x
9. Spatial distribution		x			x	x	x	x	x	x	x
10 Effort			x		x	x	x	x	x	x	x
11 Travel time				x	x	x	x	x	x	x	x
Combinations of scenarios - All metrics											
12 High density scenarios	x	x	x	x	x		x	x		x	
13 Low density scenarios	x	x	x	x		x			x		x
Combinations of metrics - All scenarios											
14 All scenarios - Macro	x	x			x	x	x	x	x	x	x
15 All scenarios - Meso			x	x	x	x	x	x	x	x	x
Combination of all scenarios and metrics											
16 All combined	x	x	x	x	x	x	x	x	x	x	x

5.3.2 General analysis of the results

Table 5.6 presents the optimal parameter sets for all 16 combinations. The results in the table show three notable things. Firstly, given the large variance in optimal parameter sets, it is clear that the choice of scenarios and metrics does affect the results of the calibration. Secondly, in all 16 combinations, the optimal viewing angle is smaller than the default and in many cases equal to the lower limit (57 degrees). Given that PD only takes into account the four closest pedestrians, it indicates that it is more important to take those pedestrians into account who are in front rather than those who are more to the side. This begs the question what would happen if the model would take into account more than four pedestrians. Furthermore, it begs the question, if one would expand the search space, would the viewing angle become significantly smaller than the current lower limit? And, how would this affect the fit of the model and the optimal value of the other two parameters? Thirdly, there are also multiple cases where the relaxation time takes the same value as the either the upper or lower boundary. So, again this begs the question if the values would change significantly if the search space would be expanded. The many cases where the optimal parameter set lies on a boundary of the search space can possibly be explained by a difference between the used combinations and the scenarios and metric used for the basic calibration of the model. This is especially the case for those combinations which only use one scenario or one metric. In the case of the viewing angle, the lack of crossing movements might also explain the fact that the viewing angle is equal to the lower boundary even when multiple scenarios are combined.

Table 5.6 also shows the minimal value of the objective function per combination of objectives. These values give insight into how well the model is capable of reproducing the reference data for the given combination of objectives whereby the lower the value the better the fit. The question if a certain minimal value of the objective function indicates a good fit or not is outside the scope of this research for the following reason: If a fit is good or not depends on the intended application of the model (i.e. what is an acceptable error for the model to make given the intended application) and this research does not assume a specific application for the model.

Table 5.6: The optimal parameter set per combination of objectives whereby $O(\theta^*)$ represents the minimal value of the objective function. The cells in red indicate that the value is at the upper or lower boundary of the search space

	Combination	$O(\theta^*)$	Relaxation time [1/s]	Viewing angle [degree]	Radius [m]
	Individual scenario - All metrics				
1.	**Bidirectional - high (B-H)**	0.1329	0.620	57.00	0.15296
2.	**Bidirectional - low (B-L)**	0.0588	0.620	57.00	0.19120
3.	**Bottleneck (B)**	0.1093	0.395	68.25	0.20076
4.	**Corner - high (C-H)**	0.0561	0.395	57.00	0.23900
5.	**Corner - low (C-L)**	0.0742	0.380	61.50	0.23900
6.	**T-junction - high (T-H)**	0.1190	0.590	57.00	0.21988
7.	**T-junction - low (T-L)**	0.0468	0.380	68.25	0.23900
	Individual metrics - All scenarios				
8.	**Flow (Q)**	0.0146	0.380	59.25	0.20076
9.	**Spatial distribution (SD)**	0.2015	0.575	59.25	0.21988
10.	**Effort (Eff)**	0.1814	0.620	59.25	0.15296
11.	**Travel time (TT)**	0.1798	0.500	57.00	0.23900
	Combinations of scenarios - All metrics				
12.	**High density scenarios (H-D)**	0.2647	0.575	57.00	0.21032
13.	**Low density scenarios (L-D)**	0.0722	0.500	57.00	0.21032
	Combinations of metrics - All scenarios				
14.	**All scenarios - Macro (Macro)**	0.1444	0.545	59.25	0.21988
15.	**All scenarios - Meso (Meso)**	0.2012	0.620	59.25	0.15296
	Combination of all scenarios and metrics				
16.	**All combined (All)**	0.1841	0.575	57.00	0.21032

In the following three subsections the results of different combinations will be compared to answer each of the three questions posed in the previous section. The comparison is based on how much the GoF of combination A decreases when, instead of the optimal parameter set obtained using combination A[5], the optimal parameter set of another combination, combination B, is used (i.e.: How much does the performance of the model, given the scenarios and metrics of combination A, decrease if the model would have been calibrated using the scenario(s) and metric(s) of combination B?). The decrease in GoF is determined as follows:

$$\Delta \text{GoF}_{A;B} = -(O_A(\theta_B^*) - O_A(\theta_A^*)) \quad [-] \tag{5.10}$$

Where $O_A(\theta_A^*)$ is the value of the objective function of combination A when its optimal parameter set θ_A^* is used. $O_A(\theta_B^*)$ is the value of the objective function of combination A if the optimal parameter set of the combination B is used. As is stated in subsection 5.1.3, an increase in the value of the objective function means a decrease in the GoF, hence the minus sign in the equation. So the larger the decrease in GoF, the worse the fit of the model to the data becomes if the given parameter set is used instead of the optimal parameter set. To put the decreases in the GoF into context, section H.2 presents a detailed analysis of how the decreases in the GoF, presented in the following three subsections, are related to the changes in the errors for the individual objectives.

The following two subsections will go into more detail regarding the influence of the two properties of the scenarios that are within the scope of this research. The third subsection will investigate the influence the choice of metrics has on the calibration results.

[5]The optimal parameter found when calibrating based the scenarios and metric of combination A

How to interpret the data in tables 5.7 to 5.10

Tables 5.7 to 5.10 and tables H.15 to H.25 display the results of comparisons between different combinations of objectives. In every table the value in the cell, determined by Equation 5.10, indicates how much the GoF of the combination, in the given column, decreases if, instead of its optimal parameter set, the optimal parameter set of another combination, given by the row, is used. So, for example, if one takes the cell in the second row and fifth column of Table 5.7, the value -0.4743 indicates the difference between the GoF of the t-junction high combination obtained when using its own optimal parameter and the GoF obtained when using the optimal parameter set of the bidirectional high density combination (i.e. $-(O_{T-H}(\theta^*_{B-H}) - O_{T-H}(\theta^*_{T-H})) = -0.4743$ where $O_{T-H}(\theta)$ is the objective space of the t-junction high density combination, θ^*_{B-H} the optimal parameter set of the bidirectional high density combination and θ^*_{T-H} the optimal parameter set of the t-junction high density combination). The number -0.4743 has no direct interpretation other than the larger the number (i.e. the larger the decrease in GoF) the worse the model performs on the predicted combination for the given parameter set. The tables in section H.2 are used to obtain insight into what the number -0.4743 means in terms of the increase in the errors.

Also, all cells in the tables are shaded whereby the darker the shade the larger the difference in GoF.

5.3.3 Movement base case

This subsection answers the question: How does the choice of movement base cases influence the calibration results? Table 5.7 presents the results of the comparisons between the different combinations. All comparisons are made between (combinations of) scenarios of the same density level to exclude the possibility that (part of) the differences are caused by a difference in the level of density and not by a difference in movement base case.

Table 5.7: Difference in the GoF, as determined by Equation 5.10, for the comparisons between different movement base cases. The combinations are identified by their acronyms as listed in Table 5.6. Every column shows the decrease in GoF, compared to the optimal GoF, of that specific combination if the optimal parameter set of the combination, defined by the row, is used. The darker the shading of the cell the larger the decrease in GoF.

Used parameter set	Predicted combination								
	H-D	**B-H**	**B**	**C-H**	**T-H**	**L-D**	**B-L**	**C-L**	**T-L**
H-D	X	-0.3528	-0.1937	-0.0402	-0.0548				
B-H	-0.0136	X	-0.1223	-0.0992	-0.4743				
B	-0.1257	-0.4084	X	-0.0501	-0.6858				
C-H	-0.1013	-0.3289	-0.1533	X	-0.5646				
T-H	-0.0135	-0.3907	-0.2679	-0.0369	X				
L-D						X	-0.0093	-0.0110	-0.0164
B-L						-0.0081	X	-0.0245	-0.0366
C-L						-0.0207	-0.0978	X	-0.0011
T-L						-0.0187	-0.0924	-0.0003	X

First, comparisons are made between the combinations of all movement base cases of the same density level and the individual movement base cases. This analysis will show whether a parameter set, based on calibrating the model using multiple movement base cases, does decrease the GoF of the individual movement base cases, by how much and if the level of density influences this.

The first and sixth row of the table show that, for both the high and low density cases, it is indeed the case that the GoF of the individual movement base cases decrease when the parameter set is used that is based on the combination of the movement base cases. It is also clear that the density level influences the size of the decrease of the GoF of the individual movement base cases and how much the decreases in GoF varies between the individual movement base cases. In the low density case, using the optimal parameter set of the combined movement base cases results in far smaller decreases in the GoF of the individual movement base cases compared to the high density case (between -0.0164 and -0.0093 in the low density case compared to between -0.3528 and -0.0402 in the high density case). Furthermore, in the low density case the difference between how much the GoFs of the individual movement base cases decreases is much smaller compared to the high density case. Overall, the parameter set resulting from calibrating the model using the combination of low density cases leads to a reasonably good fit on the individual low density scenarios given that, on average, the errors only increase by between 0.46% and 2.98% of the normalization values (for a more detailed explanation see section H.2). This cannot be said for the high density case where the errors increase by, again on average, between 4.78% and 23.18% of the normalization values.

So, the data shows that when the model is calibrated using multiple movement base cases this does decrease the GoF of the model to the individual movement base cases and that this is especially the case for the high density scenarios.

To investigate if part of the difference between the low and high density cases is caused by the fact that the high density case includes two additional movement base cases (entering and exiting in the bottleneck scenario), the model is also calibrated using the high density scenarios except for the bottleneck. The results showed that, without the bottleneck scenario, the differences become even larger. So, the difference between the low and high density scenarios cannot be explained by the additional movement base cases in the high density scenarios.

The second set of comparisons are used to answer the question - How does the use of a single movement base case to calibrate the model influence the GoF of the other movement base cases and the combinations of movement base cases? The data shows that in all cases the GoF of the individual movement base cases decreases when the optimal parameter set of another movement base case is used. However, the data also shows that the decrease in the case of using the optimal parameter set of the corner low density scenario in the t-junction low density scenario is very small, certainly compared to the rest, and that this is also the case vice versa. Table 5.6 does also show that the parameter sets are fairly similar given that only the viewing angle differs slightly. This is not a surprise given that both involve the pedestrians rounding a corner and in the t-junction low density case the influence of merging is probably limited given the low densities. Furthermore, the data shows that, again, the differences are much larger in the case of high densities. The fact that, in most cases, the GoF of other movement case cases decrease significantly when the optimal parameter set of another movement base case, shows that it is necessary to calibrate using multiple movement base cases to capture all relevant behaviour, especially in the case of high density levels. However, as the analysis also shows, at the low density level it is likely sufficient to include either the t-junction or the corner but not necessarily both. This is also supported by the data in section H.2. Because, as the data shows, the decreases of respectively -0.0011 and -0.0003 are correlated with very small increases of the errors (respectively 0.46% and 0.04% on average).

Besides the observations discussed above, Table 5.7 shows two other notable patterns in the case of the high density scenarios. Namely, regardless of the used parameter set, the corner high density scenarios always has smallest decrease in the GoF out of all four individual high density scenarios. And, both the bidirectional and t-junction high density scenarios shows large decrease in the GoF when the optimal parameter set of another scenario is used. However, when the scenarios are combined the decrease in GoF is relatively small for the t-junction scenario whilst for bidirectional scenario it remains high. The first observation can be explained using Figure 5.3 which shows how the objective values of the four high density scenarios are distributed. The figure illustrates that the objective values of the corner scenario are generally smaller than those of the other three scenarios and that they also vary less. The curves also explain part of the second observation because most of their objective values are larger than those of the corner and bottleneck scenarios. However, the difference in the decrease in GoF when the optimal parameter set of the combined scenarios is used isn't. A possible explanation for this is that the area(s) in the objective space where the bidirectional scenario has relatively low objectives values does/do not coincide with the area(s) where the t-junction scenarios has relatively low objectives values. If this is the case the calibration method would not be able to obtain a good fit on both of them simultaneously. Furthermore, as Table 5.6 the optimal parameter set of the bidirectional scenario has a larger distance to the optimal parameter sets of the bottleneck and corner scenarios than the t-junction scenario. Hence, it is logical that, when all high density scenarios are combined, and, when the method cannot obtain a good fit on both the bidirectional scenario and the t-junction scenario simultaneously, the calibration method results in a good fit for the t-junction scenario and not for the bidirectional scenario.

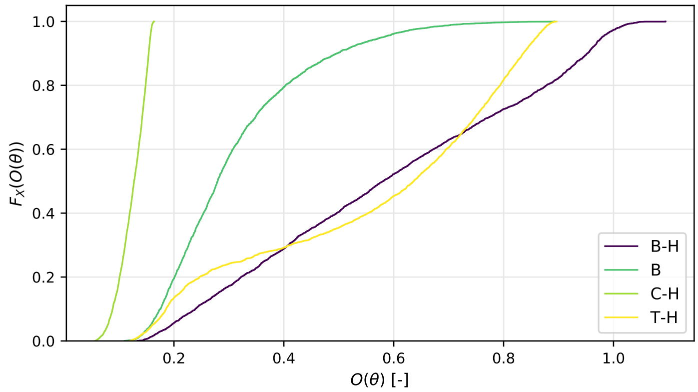


Figure 5.3: The cumulative distribution functions (CDFs) showing how the objective values are distributed for the four high density scenarios. The curves thus show the likelihood (y-axis) of encountering an objective value smaller than a certain value (x-axis) in the objective space of the given combination. The further to the left the curve generally is the smaller the objective values (and thus the errors) and the larger the spread over the x-axis the larger the difference between objective values in the same objective space (i.e. the less flat the objective space is).

So, overall it is clear that the choice of movement base case(s) does influence the calibration results. None of the movement base cases is able to capture all relevant behaviour and hence one does need a combination of multiple movement base cases to capture all relevant behaviour, especially at high density levels. In some cases though, certain movement base cases might be interchangeable (e.g. the low density corner and t-junction scenarios). And although one needs a combination of multiple movement base cases to capture all relevant behaviour, this practice does lead to a decrease in the performance on the individual movement base case.

5.3.4 Density level

The second question posed at the start of the section is: How does the choice of level of density influence the calibration results? In order to answer this question two comparisons will be made. Firstly, the three scenarios which have both a low and high density case will be compared. Secondly, a comparison is made using the combinations of high and low density scenarios.

The data in Table 5.8 shows the following regarding the influence of the density level on the individual movement base cases. Firstly, in all three cases the decrease in the GoF is far smaller when the optimal parameter set of the high density case is used in the low density case than vice versa. Especially in the cases of the bidirectional and t-junction scenarios the differences between the two different density level cases are large compared to the corner scenarios. Secondly, for both the high and the low density levels the decrease in GoF is largest in the case of the t-junction scenario. This indicates that, for the t-junction scenario, the difference in behaviour between a low density case and a high density case is larger than is the case for the other two scenarios. This is also consistent with the fact that, as Table 5.6 shows, the difference in parameter sets between the two t-junction scenarios is larger than is the case for the other two pairs of scenarios.

When the decreases in the GoF depicted in Table 5.8 are put into the context of the change in the errors the following can be concluded. The decreases in the GoF of the bidirectional low density scenario and both corner scenarios correlate with very minor increases in the average error ($< 1\%$) and increases in the standard deviation of the errors. Hence for these three scenarios, using the optimal parameter set obtained using the other density level leads to a very similar GoF and sizes of the errors. This is not the case for the other three scenarios where the errors and/or the standard deviations of the errors increase a lot more, especially in the case of the t-junction high density scenario.

When the two combinations of scenarios of the same density level are compared the following can be concluded. The first two columns of Table 5.9 show that using the parameter set, obtained using

Table 5.8: Difference in the GoF, as determined by Equation 5.10, for the comparisons between different density levels for every movement base case. The combinations are identified by their acronyms as listed in Table 5.6. Every column shows the decrease in GoF, compared to the optimal GoF, of that specific combination if the optimal parameter set of the combination, defined by the row, is used. The darker the shading of the cell the larger the decrease in GoF.

Used parameter set	Predicted combination: B-H	B-L	C-H	C-L	T-H	T-L
B-H		-0.0026				
B-L	-0.3149					
C-H				-0.0019		
C-L			-0.0032			
T-H						-0.0258
T-L					-0.5869	

the combination of high density scenarios, results is less of a decrease of the GoF of the combination of low density scenarios than vice versa. This is consistent with the previous conclusion that for all three scenarios the usage of the high density's optimal parameter set on the low density case leads to a smaller decrease in the GoF than vice versa. When the decreases of the GoFs are put into the context of the change in the errors it is clear that using the parameter set obtained using the combination of high density scenarios on the combination of low density scenarios leads to only minor changes in the errors as the mean error increases from $\pm 19.38\%$ to $\pm 19.49\%$ and the standard deviation of the errors from 19.12 to 19.73. This is not the case the other way round where the increases in both the mean error and the standard deviation of the errors are many times larger.

Table 5.9: Difference in the GoF, as determined by Equation 5.10, for the comparisons between different density levels for combinations of movement base cases. The combinations are identified by their acronyms as listed in Table 5.6. Every column shows the decrease in GoF, compared to the optimal GoF, of that specific combination if the optimal parameter set of the combination, defined by the row, is used. The darker the shading of the cell the larger the decrease in GoF. (U.p.s = Used parameter set)

U.p.s.	Predicted combination: H-D	L-D	B-H	B	C-H	T-H	B-L	C-L	T-L
H-D	X	-0.0044	-0.3528	-0.1937	-0.0402	-0.0548	-0.0055	-0.0194	-0.0250
L-D	-0.0655	X	-0.3154	-0.1262	-0.0433	-0.4185	-0.0093	-0.0110	-0.0164

Table 5.9 also shows how the decreases in the GoF for all individual scenario when either of the parameter sets obtained using the combinations of scenario are used. Surprisingly, the decrease in GoF of both the high density bidirectional scenario and the bottleneck scenario is smaller when the optimal parameter of the low density combination is used. This might be explained by:

a) As shown in the previous subsection the location of the optimal parameter set obtained using the combination of the high density scenarios is dominated by the t-junction scenario. And,

b) In this subsection is was found that the density level has the largest impact on the difference between the t-junction scenarios.

So, given the two findings above it might be expected that the combination of low density scenario results in a parameter set that balances the errors differently over the scenarios than the combination of high density scenarios would and that this has the largest impact on the t-junction high density scenario.

Also surprisingly, the fit of the bidirectional low density scenario is better when the optimal parameter set of the high density combination is used. This whilst it was not included in the objectives based on which the optimal parameter set was determined and the fact that the high density bidirectional scenario has the worst fit compared to the other scenarios. This can be explained by the fact that the bottleneck movement base case does not include a low density case. Tests show that if the bottleneck scenario is omitted from the combination of high density cases, the resulting parameter set does indeed show a worse fit for the bidirectional low density scenario compared to the fit using the parameter set obtained using the low density combination.

The last observation that is made is the fact that Table 5.6 shows that the optimal parameter sets of the combination of high density scenario and the combination of all seven scenarios are the same. This can possibly be explained by the fact that, as the CDFs in Figure 5.4 clearly show, the objective values of the low density cases are almost always smaller than the objective values of the high density cases regardless of the parameter set. Hence, in this case the low density scenarios won't have any influence on the location of the optimal parameter set when they are combined with the high density scenarios.

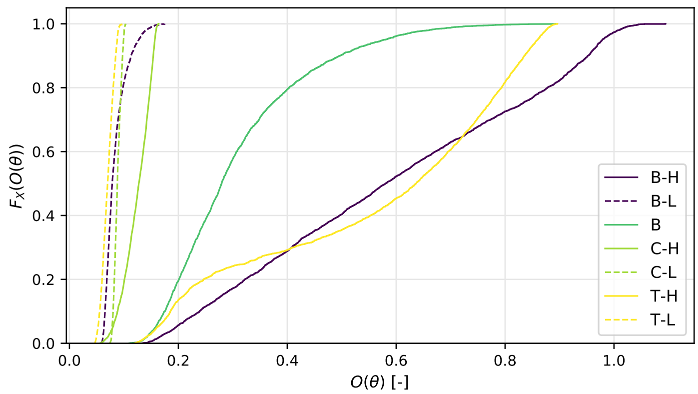


Figure 5.4: The CDFs showing how the objective values are distributed for the all seven scenarios

So, generally a calibration using solely low density scenarios cannot capture the behaviour found in the high density scenarios. This is not the case vice versa. The exception seems to be the corner movement base case. Hence, it is clearly important to, at least, include scenarios with the highest density levels, one wants the model to be able to reproduce, in the set of scenarios one uses for the calibration. Is also has to be noted that, in this particular case, the low density scenarios could even be omitted given that the same optimal parameter set is found regardless if they are included or not. However, the question remains if this would also be the case if another combination of scenarios is used or another combination of metrics. Hence, it would be advisable to at least check the performance of the model on the low density cases during the validation.

5.3.5 Metrics

The question - How does the choice of metrics influence the calibration results? - is investigated using the results from combinations 8 - 11 and 14 - 16. Table 5.6 clearly shows that the choice of metric or combination of metrics influences the resulting optimal parameter set. The comparisons, whose results are presented in Table 5.10, give a more detailed picture from which a number of observations can be made.

Firstly, there seems to be a correlation between the distribution of the effort and the spatial distribution. When the model is calibrated using only one of them, the decrease in the GoF of the other is small, certainly when compared to the flow or travel time distribution. The correlation is also apparent when the model is calibrated using the two macroscopic metrics.

Secondly, when the optimal parameter set obtained using solely the travel times is used the decrease in the GoF is very small. As subsection H.2.3 shows the decrease of -0.0079 correlates with an average increase of just 2% of the normalization value and the standard deviation also increases only slightly. This in contrast to when the optimal parameter set of either of the other two metrics is used. These lead to increases of the around 7.5% and large increases in the standard deviation.

Thirdly, it is clear that if combinations of metrics are used and both macroscopic metrics are included, the resulting optimal parameter set results in only a small decrease in the GoF of the spatial distribution but a very large decrease in the GoF of the flow. Figure 5.5 gives insight why this might be the case. As the figure illustrates, the objective values of the flow are small compared to the other three metrics and hence it is thus not surprising that when the flow is used in combination with any of the other three metrics it weighs less heavily when determining the optimal parameter set. The po-

Table 5.10: Difference in the GoF, as determined by Equation 5.10, for the comparisons between different metrics. The combinations are identified by their acronyms as listed in Table 5.6. Every column shows the decrease in GoF, compared to the optimal GoF, of that specific combination if the optimal parameter set of the combination, defined by the row, is used. The darker the shading of the cell the larger the decrease in GoF.

		Predicted combination			
		Q	**SD**	**Eff**	**TT**
Used parameter set	**Q**	X	-0.1281	-0.1213	-0.1635
	SD	-0.0844	X	-0.0228	-0.1454
	Eff	-0.0902	-0.0198	X	-0.0596
	TT	-0.0079	-0.1235	-0.0412	X
	Macro	-0.0697	-0.0029	-0.0261	-0.1416
	Meso	-0.0079	-0.1235	-0.0412	0.0000
	All	-0.0697	-0.0120	-0.0548	-0.0223

tential influence of the choice of normalization values on this observation is discussed in more detail in section 5.4.

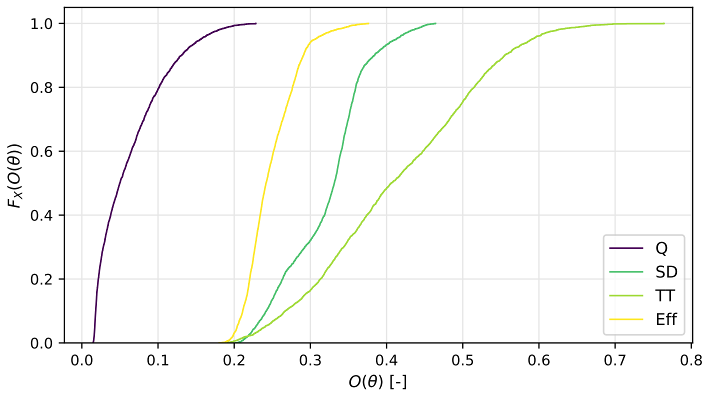


Figure 5.5: The CDFs showing how the objective values are distributed for the four metrics

Fourthly, when only the two mesoscopic parameters are used when calibrating the model this leads to the same optimal parameter set as when only the travel time is used. Figure 5.5 shows that this can be explained by the objective values of the effort are generally smaller than those of the travel time and hence, comparable to the macroscopic metrics where the spatial distribution is dominant, the travel time is in this case dominant.

Lastly, Table 5.10 shows clearly that whatever combinations of metric is used, the model cannot be calibrated such that is results in a good fit on all four of them.

Overall it is clear that the choice of metrics influences the results of the calibration. Depending on the choice of metric or combination of metrics, different optimal parameter sets are found which in turn lead to different results regarding the GoF to the individual metrics. Furthermore, the method used to determine the objective function does seem to influence the results as well. Especially the normalization could be a large influence. The results also show that the model is not capable of simultaneously obtaining accurate results for the different metrics using a single parameter set.

5.3.6 Reflection on the results

The previous three subsections list a number of findings. The question is, how do these finding compare to the findings of previous studies? And, how do the finding compare to the findings of the sensitivity analysis?

The findings of this research regarding the influence of the movement base cases are found to be consistent with both (Campanella et al., 2011) and (Duives, 2016). Similar to those studies, this research finds that:

- It is indeed the case that the GoF of the individual movement base cases decreases when the parameter set based on multiple movement base cases is used.
- It is indeed necessary to use multiple movement base cases, when calibrating a model, to capture all relevant behaviour.

Also, comparable to the finding in (Duives, 2016), this research finds that the choice of metric does clearly influence the results of the calibration and that the model cannot obtain a good fit on all of them simultaneously. Furthermore, there are also indications that the choice of objective function and normalization method influence the results when more than one metric is used.

Furthermore, the finding that the high density scenarios influence the calibration more heavily than the low density scenarios, is consistent with the findings of the sensitivity analysis. Namely, the sensitivity analysis (see chapter 4) shows that the model is more sensitive to changes in the parameters in the case of the high density scenarios.

The fact that the results of this study are in line with the previous studies, whilst using a different model than the two previous studies, raises an important question. Namely, is the fact that the models cannot obtain a good fit on different scenarios using only a single parameter set caused by the fact that:

a.) the models are simplifications of the behaviour of pedestrians and the models are too simple to capture the behaviour of pedestrian in different flow situations well using only a single parameter set. Or,

b.) the behaviour of pedestrians in different flow situations is so different that it might not be a valid approach to try to capture this using a single model (i.e. the assumption that the behaviour of the pedestrians is independent of the flow situation is not valid).

The results of this research cannot answer this question. However, it is important to answer this question given that, in the case it would be the second cause listed above, it would fundamentally change the way in which we need to model pedestrian behaviour.

5.4 Discussion on the calibration results and the methodology

In this section the used methodology is systematically discussed in order to obtain insight into how changes in the methodology (e.g. using another objective function) could potentially impact the results. All elements that fall outside of the scope of this research will be discussed. So, these are all elements except for the metrics and the scenarios and the accompanying reference data. The three elements that are within the scope of this research are discussed in chapter 6 where the primary question is - How generalizable are the results?.

So, in total six of the nine elements will be discussed in this section whereby, similarly to section 5.1, some elements are discussed simultaneously. Per elements it will be discussed if and how a different choice, regarding that element, could potentially change the results.

5.4.1 Input definition

Four types of input were identified in subsection 5.1.1. Namely, the geometry of the infrastructure, the route choice, the demand patterns and the speed distribution. The geometry of the infrastructure was implemented such that it matched the experiments exactly so there is no reason to assume it has had any influence on the calibration. The need for the waypoints in the bidirectional and t-junction scenarios does raise the question what would happen if one were to omit them or position them slightly different. Would it affect the calibration results significantly or is the model not very sensitive to these changes. To answer this one could test different configurations of waypoints to obtain insight into the effect it has on the flows in these scenarios.

The third input to the model were the demand patterns. The demands were assumed to be constant and tests showed this to be a good approximation. However, would it impact[1] the calibration results if a more precise definition of the demands would have been used. Given that: a) The approximations of the demands are already fairly precise (i.e. the high R^2 values shown in subsection 5.1.1), and b) Due to the stochastic nature of the model there is already variability in exactly where pedestrians enter the simulation and the route they choose to walk from this point onward. It is the question if the slight increase in variability in the demand pattern would have any measurable effect on the simulations.

The last input to the model is the distribution of the preferred speeds. The distribution was fitted on the speed measurements of 54 participants. Figure 5.1 shows that the distribution is a reasonable fit to the measurements however by no means a perfect fit. However, one also has to note that the speed measurements are based on only a sample of the participants and hence the exact distribution of the preferred speeds is to some degree uncertain. So, the main question would be how sensitive the model is to this uncertainty in the distribution of the preferred speeds.

Overall, three of the four inputs used in this research could potentially have some impact on the calibration results. However, to ascertain if this would change the results to such a degree that it would change the conclusions, one would have to test how sensitive the model is to uncertainties in the input.

5.4.2 Objective functions and comparison methods

This subsection discusses if and how:

a.) Using a different objective function could lead to different calibration results.

b.) Using different normalization values could lead to different calibration results.

c.) Using a different method for combining multiple objectives into one could lead to different calibration results.

Would a different objective function lead to different results? Table I.1 in appendix I shows that using another objective function, in this case using the absolute errors instead of the squared errors, does indeed lead to different optimal parameter sets. However, clear differences between the optimal parameter sets of different scenarios and different metrics still clearly exist. So, the fact that differences in the optimal parameter sets also exist when using a different objective function is an indication that the conclusions from the previous section are likely to hold when another objective function is used.

Would different normalization values change the results? The detailed analysis in appendix H shows that, although the results do change here and there, the main conclusions hold when instead of all four metrics only one of the metrics is used. Hence, the usage of different normalization values would not change the main conclusions regarding the movement base cases and the density levels. It would of course have an impact on the conclusion regarding the metrics. However, also in this case slight changes in the normalization values are not expected to change the main conclusions.

Would a different method for combining multiple objectives into one change the results? In many respects this question is similar to the question above regarding the normalization values because it is primarily the question if using different weights for the individual objectives would result in changes. Given that the detailed analyses in section H.3 show that the main conclusions also hold when only single objectives are used, using different weights are not likely to affect the main results.

5.4.3 Optimization method, stopping criteria and search space definition

This subsection discusses how different choices regarding the optimization method, the stopping criteria and the search space definition could potentially affect the calibration results. If one assumes that one uses a different optimization method correctly such that it finds the global optimum for the given level of precision, the question becomes - If one uses a different method and stopping criteria which would change the precision with which the global optimum is determined, how could this potentially affect the results?. This question is discussed in the part of this subsection that discusses the potential affect of changing the precision of the search space definition.

The choice of search space could influence the results in three ways. Firstly, it only takes into account 2 of the 7 parameters which were investigated during the sensitivity analysis. Although the sensitivity

[1]Would it result in changes in the calibration results such that the conclusions, made based on the analyses of the previous section, would not longer be valid?

analysis concluded that the model was not sensitive, or very slightly sensitive, to changes in these five parameters, it did only take into account the first-order effects. So, the question is, would including more parameters significantly change the results? A more extensive sensitivity analysis, which would include the higher-order effects, could give insight into this question. If the parameters, not included into the search space, would not show any significant higher-order effects it would be unlikely that the results of the calibration would change significantly.

Secondly, Table 5.6 in section 5.3 shows that in multiple cases the optimal value of the viewing angle and/or the relaxation time is equal to the upper or lower boundary of the search space. This begs the question if these values are actually the optimal value or if they are the optimal value given the boundaries of the search space. To answer this question the search space would have to be extended, in accordance with section 4.2 it should be checked if the new boundaries still produce realistic results, and additional simulations would have to be run. However, given the large variation in the optimal parameter sets, especially in the optimal value of the relaxation time, it is considered unlikely that expanding the search space will decrease the difference between the optimal parameter sets for the different combinations and hence unlikely that it will significantly change the results.

Thirdly, to keep the number of required simulations at a feasible level a step-size of 3% of the default value (4% in case of the radius) was used in this research. The question is what would happen to the results if a smaller step-size would be chosen? The potential effect of a smaller step-size is explained using the example depicted in Figure 5.6. The figure shows the objective function $O(\theta)$ at two different locations (I and II) in the 1d search space (Θ, $\theta \in \Theta$). The solid lines represent the objective function and the vertical striped lines the location of the points in the grid for a given level of precision. The figure shows that, although the minimum value of the objective function can be found at location I, due to the way the grid overlays the objective space it is found at location II. This example shows that the precision with which the grid is defined matters and that the more precise the grid the more likely the actual location at which the objective is minimized is found.

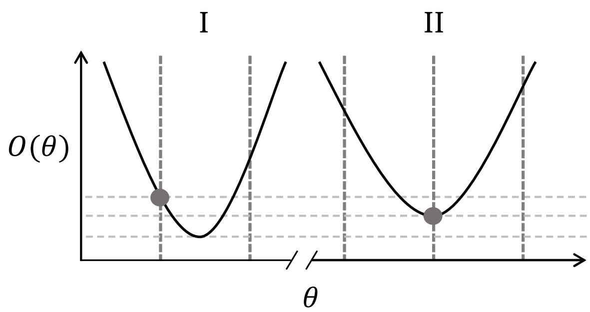


Figure 5.6: Example of the influence of the precision of the search space on the optimal parameter set

So, given the fact that the precision of the grid can influence the results how likely is it that using a more precise grid would significantly change the findings? Based on the analysis of the objective space surfaces in appendix J, it would probably be worthwhile to investigate this given that, for example, in the bidirectional high density scenarios there are points that reside in a very different part of the objective space (i.e. have significantly different parameter sets) than the optimal point but have only slightly higher objective values. However, this does not hold for the other three high density scenarios so, for example, the conclusion that one needs a combination of multiple movement base case will likely still be valid even if the search space is defined more precisely.

5.4.4 Stochasticities

Section 3.4 showed that even at a high number of replications different seed orders can result in significant different results. This means that, if for two sets of replications different seed orders are used, one has to determine if and how much of the difference found between the two sets is caused by the stochastic nature. Subsection 5.1.5 explained why this is problematic during a calibration and hence why the decision was made to use a fixed seed order during the calibration. However, by using a fixed seed order and hence also a fixed number of replications it remains the question if the resulting distribution (for a given metric) is representative of the actual distribution (i.e. the one one would get if an infinite amount of replications were to be used). Hence, the question also remains, would a different order of the seeds or a higher number of replications lead to significantly different results?

Ideally, one would research this by, for example, defining every point in the objective space as a distribution around the current objective value and by using a monte-carlo-like method determining the likelihood that a certain point in the objective space is the optimal point. However, time restrictions prevented this analysis to be performed within this research. However, the analysis in appendix J gives some insight into which cases are probably more sensitive to this than others. Appendix J concludes that the low density scenarios are more likely to be affected by this than high density scenarios and that the bidirectional scenarios are more likely to be affected by this than the other scenarios of the same density level. Furthermore, it also shows that when the flow would be used as the sole metric the calibration results are far more likely to be affected than if any of the other metric or a combination of metrics is used.

Overall, the analysis of appendix J does show that using a different order of the seeds or a higher number of replications could affect the results. However, what the effect exactly would be and if it would result in different conclusions cannot be said based on this analysis. So, further research into how to deal with the stochastic nature of pedestrian models during calibration would be advised. This could both involve coming up with a better (more stable) method to determine the required number of replications or a method aimed at dealing with the uncertainty caused by the stochastic nature during the calibration.

5.5 Practical implications of the results

The findings of the previous section have a number of implication for practice. The main implication of the results for practice is that the ***intended use*** of the model should be taken into account when deciding which scenarios, metrics, objective functions and method for combining multiple objectives one should use. Why this is the case is discussed in the remainder of the section.

As this research confirms, one needs to use multiple movement base cases when calibrating a model intended for general usage. However, when the intended use of the model is more limited (e.g. in its use only some movement base cases occur) it might be better to also use a more limited set of movement base cases during the calibration. This is mainly because, as this research also confirms, the GoF of the individual movement base cases decreases when multiple movement base cases are used during the calibration. The results do also indicate that it is important to take into account the density level during the calibration. Again, depending on the intended use of the model, different density levels should be taken into account during the calibration. Furthermore, as the results show, it is far more important to take the higher levels of density into account.

Given that:

1. There are many possible metrics which one could use to quantify a pedestrian flow (see Table A.1)
2. The results show that the choice of metric or combinations of metrics does influence the results
3. The results show that, in the case of using multiple metrics, the choice of normalization method and the method for combining multiple objectives into a single objective also influence the results

the choice of metrics, accompanying objective functions and method for combining multiple objectives into a single objective depends even more on the intended use than the choice of scenarios. Depending on the intended use of the model, one should decide which metric or metrics are most important (i.e. which metrics are most important and hence need to be reproduced as accurately as possibly). One should also decide how to reflect this in the method for combining multiple objectives into one. Besides using different normalization methods in combination with the weighted sum method, one could also use other approaches. For example, one could use methods based on constraints (e.g. the ϵ-Constraint method) where one or more of the objectives are transformed into constraints (Ehrgott, 2005). An example of this would be that one sets the maximum error one would allow the model to make with regard to the flow as a constraint and then optimize based on the travel time distribution.

To show how the intended use of the model could potentially affect how it should be calibrated two examples are given. The first example is a model that is intended to be used for evacuation studies.

During evacuation the flows are primarily unidirectional and depending on the situations the density can be high (for example a full stadium that is being evacuated). So, this would mean one would like to prioritize the unidirectional movement base cases over the bidirectional and crossing movement base cases either by using different weights or not including them altogether. Furthermore, one would probably only have to include those scenarios with high densities as the results of this research show that including the low density scenarios will probably not affect the calibration results. Regarding the metrics, the distribution of evacuation times could be a good candidate as the most important metric (i.e. the objective functions should be determined such that calibration results in a parameter set that leads to an error that is at least lower than a certain threshold) supplemented with a metric that focusses more on the underlying behaviour such as the effort indicator in this research. Whatever the exact choice of scenarios and metrics is and regardless of how they are prioritized, one should note that it is important to validate the calibrated model using a wider selection of scenarios and metrics to obtain insight into what would happen if, for example, for whatever reason there are bidirectional flows during the simulation of the evacuation.

The second example used is a station hall during off-peak hours. In this case the bidirectional flows and crossing flow become more important and the higher density levels less important. Hence, one would primarily prioritize these movement base cases at a low density level. Regarding the metrics, a metric that captures how comfortable a pedestrian can walk from A to B might become more relevant than performance measures such as the flow or the travel time. So, for example, the distribution of the effort can be taken as the primary metric supplemented with a metric such as the travel time distribution. Again, it is important to validate the calibrated model using a wider selection of scenarios and metrics.

In the examples above, using a parameter set that has been obtained using the specifics of the type of situation that has to modelled into account will probably result in more accurate prediction than if one where to use a parameter set obtained using no specific application in mind.

5.6 Conclusions

In this chapter the model was calibrated using different combinations of objectives to investigate the influence different choices regarding scenarios and metrics have on the calibration results. In section 5.1 the methodology used was explained. In total, seven different scenarios were used and four different metrics. Based on, among other things, the sensitivity analysis, the search space was made up out of three parameters, the relaxation time, the viewing angle and the radius. This lead to a search space of 3179 points. To deal with the stochastic nature of the model every scenario used a fixed number of replications and the same order of seeds. All in all, this lead to over 1.27 million simulations, the results of which were compared to reference data which, in turn, lead to 42 different objective spaces (1 for each individual objectives).

Section 5.2 discussed the results of these 42 individual objectives. The data showed, for the given set of parameters, that the model was capable of fitting the simulated flow to the data. For the other three metrics the degree to which the model was able to fit to the data depended on the scenario.

Section 5.3 discussed the comparisons between different combinations of objectives. In total 16 different combinations of scenarios and metric were used whereby the multiple objectives were combined into a single objective using the weighted sum method. The different comparisons were used to answer the following three questions:

1. How does the choice of movement base cases influence the calibration results?
2. How does the choice of level of density influence the calibration results?
3. How does the choice of metrics influence the calibration results?

The calibration results show that using different combinations of scenarios and metric does lead to different optimal parameter sets and hence the choice of scenarios and metrics clearly influences the calibration results. Regarding the influence of the movement base cases it was found that: ***a)*** The GoF of the individual movement base cases decreases when the parameter set based on multiple movement base cases is used. And ***b)*** It is necessary to use multiple movement base cases, when calibrating a model, to capture all relevant behaviour. These two findings are in line with the finding of the two previous studies which investigated the influence of the movement base cases. A difference in the level of density did also impact the results. High density scenarios have a larger impact on the results which is consistent with the findings of the sensitivity analysis where it was found that the model is more sensitive to changes in the parameters in the high density scenarios. Also, the comparisons showed that, similar to (Duives, 2016), the choice of metric also influence the results. When a combination of metrics is used the results also showed that the choice of normalization method also can influence the results. Lastly, the results point to a fundamental question. Namely, given the differences found between the scenarios, is it a valid approach to use a single model to capture the behaviour of pedestrians?

Section 5.4 discussed the used methodology and how different choices regarding the methodology could potentially affect the results. Elements, such as using a more precise search space grid or using another seed order were found to have the potential to affect the results. However, quantitative analyses, for which there was no time in this research, should be performed to ascertain what the exact effects are and if this would change the main conclusion significantly. The fact that the main conclusions are in line with previous research and the results of the sensitivity analysis, though, gives a strong indication that this is not likely.

The result were also found to have a number of practical implications as is pointed out in section 5.5. The main implication found was that it is important to take the intended use of the model into account when performing a calibration. This is especially the case for the choice of metrics and the method used to combine multiple objectives into a single objective.

6 Conclusions, discussion and recommendations

In this chapter the main conclusions are drawn, the findings and limitation are discussed and recommendations to both science and practice are made. In section 6.1 the main conclusions of this research are discussed in light of the main research question. Section 6.2 will, based on the findings, discuss the main limitations of this research. Lastly, section 6.3 will discuss the recommendations that can be made based on this research.

6.1 Conclusions

This research aims to improve the applicability of the multiple-objective framework for calibrating pedestrian simulation models by gaining improved insights into how the choice of objectives influences the calibration results. To structure the research the following main research question is posed:

> ***How can a microscopic pedestrian model be calibrated, using a multiple-objective approach, given its stochastic nature and differences in behaviour in different flow situations?***

A review of the literature concluded that the transferability of models, which are calibrated using a limited focus, is questionable. Hence, as the literature suggests, a multiple-objective approach should be used when calibrating a pedestrian simulation model. The review of the literature identified nine elements which are necessary to calibrate a pedestrian model using multiple objectives.

As the research question states, one should take into account the stochastic nature of the model when calibrating it. Section 3.4 showed why this is the case whereby it was found that, even with a high number of replications, different seed orders lead to significantly different speed distributions. During a calibration this is problematic given that one assumes that differences in the objective values are caused by differences in the parameter values. If these differences could also be caused by the stochastic nature of the model this assumption would no longer hold. So, to prevent this, this research used a fixed order of seeds during the calibration.

The research question also states that one should take into account the differences in behaviour in different flow situations. Chapter 4 showed why this is not only the case during the calibration but why this is also important during the sensitivity analysis. This research showed that the sensitivity of the model to changes in the parameters differs between the different scenarios. Hence, to get a complete picture of the sensitivity of the model one should use multiple scenarios. This complete picture of the sensitivities is, in turn, necessary because the sensitivities are used, among other things, in determining the search space for the calibration.

Chapter 5 showed why it is indeed important to take into account the differences in behaviour in different flow situations during the calibration. Subsection 5.3.3 showed that in order to capture all relevant behaviour one does need to include multiple movement base cases in the calibration. However, as the subsection also showed at some density levels some movement base cases might be interchangeable. Subsection 5.3.4 showed that one also needs to take into account the density level. The results show that the influence of the high density level scenarios on the calibration is far larger than that of the low density level scenarios. Hence, it is clearly important to, at least, include scenarios with the highest density levels, one wants the model to be able to reproduce, in the set of scenarios one uses for the calibration.

Chapter 5 also showed in subsection 5.3.5 that the choice of metric or combinations of metrics clearly influences what optimal parameter set results from the calibration. When using multiple metrics during the calibration, it also clear that the method used to combine multiple objectives into a single objective influences the results especially the normalization of the metrics. Furthermore, the data also shows that the model does not seem capable of obtaining a good fit on all metrics at the same time. Hence, it is important to prioritize which metrics are most important for the given application such that is can be assured that a good fit (or at least a as good as possible fit) is obtained during the calibration.

All in all, the conclusions above point to one important implication for the use of the multiple-objective approach when calibrating a pedestrian simulation model. One should calibrate a pedestrian simulation model based on the intended application of the model. So, the calibration should include those scenarios that are likely to occur to ensure that all relevant behaviour is captured. However, including scenarios that are unlikely to occur, given the intended application, will likely influence the calibration results negatively and hence also the model's accuracy with which it can predict the traffic state. Furthermore, the metrics and accompanying objective functions and method for combining the multiple objectives into a single objective, should be chosen such that they represent which metrics are most important to the intended application and what level of accuracy one wants the model to have regarding the particular metrics.

Moreover, the conclusions also point to an important question. Is the assumption, underlying most if not all pedestrian simulation models, that the behaviour of the pedestrians is independent of the flow situation valid? This research cannot answer this question. However, if the answer to this question is, no it is not valid, it would fundamentally change the way in which we need to model pedestrian behaviour.

6.2 Discussion

This section discusses the limitations of the research and puts the results into perspective. Four different elements will be discussed. Firstly, subsection 6.2.1 discusses the limitations of the fact that only a number of different scenarios were used during this research. Subsection 6.2.2 discusses the used metrics and especially the choice not to use microscopic metrics. Thirdly, subsection 6.2.3 discusses the lack of a validation. And lastly, subsection 6.2.4 discusses the generalizability of the results (i.e. are the conclusions also likely to hold when another pedestrian simulation model is used?). The limitations of the used calibration methodology have already been discussed and this discussion can be found in section 5.4.

6.2.1 Scenarios

In subsection 2.2.2 five different properties of a scenario were identified. Namely, the infrastructure, the demand patterns, the population composition, the movement base case and the density level. Out of these five properties only two have been researched in this study (the movement base case and the density level) and also for these two properties there were limitations regarding the extent to which they could be researched. This subsection will discuss the limitations per property.

Infrastructure

This research did not vary the geometry of the infrastructure and hence cannot answer the question whether or not one should include different geometries when calibrating a pedestrian simulation model. The review of the literature in subsection 2.2.2 also found mixed evidence for the influence of the geometry on the flow. Hence, this could be a topic for further research.

Demand patterns

Comparably to the infrastructure, the effect of varying the demand patterns was not studied in this research. And, again, the review of the literature did not provide a clear answer as to if varying the demand patterns influences the flow and if so by how much. So, this could also be a topic for further research.

Population composition

The review of the literature in subsection 2.2.2 found that the composition of the population can influence the flow and that it is thus an important element to consider whilst calibrating a pedestrian model. However, section 3.1 showed that the available data did not make it possible to vary this property in this research. Furthermore, the population used in the experiments which provided the reference data for this research was fairly homogeneous. So, this begs how the calibration results, and especially the differences regarding the movement base cases and the density levels, would change if the population composition would be varied and if the main conclusion would still hold.

Movement base case

The influence of using different movement base cases has been studied in this research. However, it lacked two classes of movement base cases, namely the crossing flows and the vertical movements. It would be interesting to study if these movement base cases also would have to be included or if they show similar behaviour to another movement base case. Regardless of this is the case or not, including them would not change the main findings of this research but only expand them.

Density level

This research only included two different levels of density during the calibration and this has given insight into the influence the density level could have on the calibration. However, the insights are limited and research including more levels of density would be necessary to get a more detailed view of what levels of density produce different results and which don't.

6.2.2 Metrics

This research used a total of four metrics to assess the influence of using different metrics whilst calibrating a pedestrian model. The chosen metrics covered two of the three aggregation levels and did not include microscopic metrics. Subsection 5.1.2 gives the reasons why this choice was made. However, this raises the question if the conclusions would also hold if microscopic metrics would be used instead of macroscopic or mesoscopic metrics. The primary concern would be if the macroscopic and mesoscopic capture the behaviour at sufficiently high detail for them to differentiate between solutions with comparable levels of GoF. Chapter J shows that, for example, the flow metric (a macroscopic metric), does have problems with differentiating between the solutions obtained using different parameter sets and that this also holds for the combination of all four metrics in the case of the low density scenarios. So, it would be interesting to study if adding microscopic metrics would make it possible to differentiate between the solutions of different parameter sets in these cases.

However, one has to keep in mind that, as shown in (Campanella, 2016), calibrating a pedestrian model solely using trajectories (i.e. calibrating it only on the microscopic level) does not lead to accurate predictions on the macroscopic level and that pedestrian simulation models are mostly used to make predictions on the macroscopic level (Campanella et al., 2014). So, it might only be fruitful to use microscopic metrics within the domain of the search space where it has been established that the model produces valid outcomes at the macroscopic and mesoscopic levels.

6.2.3 Validation

If one looks at definitions of validation given in the literature (e.g. Department of Defense, 1996; Law, 2008; Knoop & Buisson, 2014) one can see that there are two important elements regarding the definition, namely:

- Firstly, that the goal of the process is to determine to which degree the model is an accurate representation of the real world.
- And secondly, that this should be determined in light of the intended use of the model.

Figure 1.1 also shows it is an important step in the development of the model and especially in determining the quality of the calibrated model and thus in part the quality of the calibration procedure. Due to time constraints, a validation could not be performed within this research. However, it also has to be noted that within this research the goal was not to deliver a calibrated model and that there is no intended usage assumed. So, the validation would have served a slightly different goal within this research. Namely, to obtain some additional insights regarding the effects of different choices of objectives (e.g. by using different/additional metrics during the validation) and to check if the findings would also hold in slightly different situation (e.g. when slightly different levels of density would be used). So, though not strictly necessary for this type of research, performing a validation has the potential to expand the insights into the effects of choosing different objectives and to give better insight into how reliable the conclusions of this research are.

6.2.4 Generalisation of the results to other microscopic models

The results of this research are in line with other studies in the literature that both used another microscopic simulation model (NOMAD). Together with the fact that PD, the model used in this research, and NOMAD use different approaches (see chapter 2 in (Campanella, 2016) for more details), this is a strong indication that the main conclusions of this research do not depend on the specific model used. However, more research using other modelling approaches, for example cellular automata or discrete choice modelling, would give better insight into how generalizable the conclusion actually are.

6.3 Recommendations

Based on this research a number of recommendations can be made to both science and practice. First the recommendations for practice will be discussed in subsection 6.3.1 followed by the recommendations for science discussed in subsection 6.3.2.

6.3.1 Recommendations to practice

Based on this study the following recommendations can be done for practice.

- Calibrate the model using the flow situations that are expected to occur given the intended use and using the metrics which are most important for the given application. Including flow situations that are unlikely to occur in the given application will likely decrease the accuracy of the model. The same holds for the metrics.
- Following from the first recommendation, calibrate the model for one specific application only. Calibrating the model for multiple applications at once will likely decrease the accuracy of the model's predictions if these applications include different flow situations which are likely to occur and/or different metric which are important.
- Validate the model using a wide range of flow situations and metrics such that one has insight into how reliable or unreliable the model's predictions are for those situations which weren't included in the calibration. This way one can determine how detrimental the occurrence of these flow situations would be if they unexpectedly were to occur in the simulation.

6.3.2 Recommendations to science

Based on this study also a number of recommendations can be done for science.

- Based on this research the following suggestions can be done for future research:
  - Research into the effect of the scenario properties that were not part of this research. Namely, in order of relevance, the population composition, the geometry of the infrastructure and the demand patterns.
  - Research into how to deal with the stochastic nature of pedestrian models during calibrating and validating the model or whilst performing a sensitivity analysis.
  - Research into how to combine calibration based on microscopic metrics with calibration using macroscopic and mesoscopic metrics whereby the microscopic metrics are used in those situations where the macroscopic and mesoscopic metrics lack the power to differentiate between different solutions.
  - Research into the question if the behaviour of the pedestrians is independent of the flow situation.

- When doing experiments with pedestrians with the goal of using the data for calibrating and validating the model, take the following two considerations into account:
  - Take into account the composition of experiments used to collect data on pedestrian walking behaviour. When the data set is collected with the possible intention to use it for calibrating a pedestrian model it is good to take into account the different scenario properties described in subsection 2.2.2. For example, if the plan is to include multiple movement base cases try to include at least those who are most likely to show the largest differences in behaviour.
  - The duration of the experiments, short time spans might be a cause for the large number of necessary replications given that stochastic elements (e.g. preferred speed, the initial location of the origin and destination which influence the local route choice) have more impact on the results when short time spans (e.g. the 1.5 minutes used in this research) are simulated. Furthermore, at high density levels the question is if certain flows can be sustained for longer periods at a time and primarily if the model is also capable of reproducing this. For example, in this research it was found that in the case of the high density bidirectional scenario total breakdown of the flow occurred quite frequently. However, it was also found that the exact time when this occurred differed a lot between simulations and sometimes this only occurred at the very end of a simulation (at a moment possibly outside of the measurement period)

Acronyms

A-D test Anderson-Darling test. iii, 34–37, 44, 46

CDF cumulative distribution function. 66, 68, 69

CMA Covariance Matrix Adaptation. 16

ECM Explicit Corridor Map. 29, 30

GA Genetic Algorithm. 16, 17

GoF Goodness-of-Fit. iv, 57, 63–70, 73, 74, 78, 108, 109, 113–120, 122, 123

IID Independent Identical Distributions. 13

IRM Indicative Route Method. 30

MSE Mean-Squared Error. 13, 14

OD-pair Origin-Destination pair. 30, 32, 52

PD Pedestrian Dynamics ®. iii, 3, 29–34, 37, 39, 51–53, 62, 79

RMSE Root-Mean-Squared Error. 13

SA Simulated Annealing. 16

SE Squared Error. 56, 57

T&P Transport & Planning. ii, 19

Bibliography

Abdelghany, A., Abdelghany, K. & Mahmassani, H. (2016). A hybrid simulation-assignment modeling framework for crowd dynamics in large-scale pedestrian facilities. *Transportation Research Part A: Policy and Practice*, *86*, 159–176. (Cit. on pp. 6, 88).

Bandini, S., Gorrini, A. & Vizzari, G. (2014). Towards an integrated approach to crowd analysis and crowd synthesis: A case study and first results. *Pattern Recognition Letters*, *44*, 16–29. (Cit. on p. 88).

Banerjee, B. & Kraemer, L. (2010). Validation of agent based crowd egress simulation. (pp. 1551–1552). Proceedings of the 9th international conference on autonomous agents and multiagent systems: Volume 1 - volume 1. Toronto, Canada: International Foundation for Autonomous Agents and Multiagent Systems. (Cit. on p. 88).

Bauer, D. (2011). Comparing pedestrian movement simulation models for a crossing area based on real world data. In R. D. Peacock, E. D. Kuligowski & J. D. Averill (Eds.), *Pedestrian and evacuation dynamics* (pp. 547–556). Boston, MA: Springer US. (Cit. on pp. 6, 88, 89).

Berrou, J. L., Beecham, J., Quaglia, P., Kagarlis, M. A. & Gerodimos, A. (2007). Calibration and validation of the legion simulation model using empirical data. In *Pedestrian and evacuation dynamics 2005* (pp. 167–181). Springer. (Cit. on pp. 1, 6, 88, 89).

Bohannon, R. W. & Williams Andrews, A. (2011). Normal walking speed: A descriptive meta-analysis. *Physiotherapy*, *97*(3), 182–189. (Cit. on p. 10)

Campanella, M. C. (2016). *Microscopic modelling of walking behaviour* (PhD Dissertation). Retrieved from https://repository.tudelft.nl/islandora/object/uuid%3Ab65e6e12-a85e-4846-9122-0bb9be47a762. (Cit. on pp. ii, 10, 12, 14, 16, 22, 23, 53, 78, 79)

Campanella, M. C., Hoogendoorn, S. P. & Daamen, W. (2009a). Effects of heterogeneity on self-organized pedestrian flows. *Transportation Research Record: Journal of the Transportation Research Board*, *2124*, 148–156. (Cit. on pp. 6, 10).

Campanella, M. C., Hoogendoorn, S. P. & Daamen, W. (2009b). Improving the nomad microscopic walker model. *IFAC Proceedings Volumes*, *42*(15), 12–18. (Cit. on pp. 6, 88, 89).

Campanella, M. C., Hoogendoorn, S. P. & Daamen, W. (2011). A methodology to calibrate pedestrian walker models using multiple-objectives. In *Pedestrian and evacuation dynamics* (pp. 755–759). Springer. (Cit. on pp. 1, 2, 7, 9, 13, 61, 70, 90).

Campanella, M. C., Hoogendoorn, S. P. & Daamen, W. (2014). Quantitative and qualitative validation procedure for general use of pedestrian models. In *Pedestrian and evacuation dynamics 2012* (pp. 891–905). Springer. (Cit. on pp. 2, 3, 9, 40, 78, 88–90).

Chattaraj, U., Seyfried, A. & Chakroborty, P. (2009). Comparison of pedestrian fundamental diagram across cultures. *Advances in complex systems*, *12*(03), 393–405. (Cit. on p. 10).

Chattaraj, U., Seyfried, A., Chakroborty, P. & Biswal, M. K. (2013). Modelling single file pedestrian motion across cultures. *Procedia - Social and Behavioral Sciences*, *104*, 698–707. (Cit. on p. 89).

Chraibi, M., Ensslen, T., Gottschalk, H., Saadi, M. & Seyfried, A. (2016). Assessment of models for pedestrian dynamics with functional principal component analysis. *Physica A: Statistical Mechanics and its Applications*, *451*, 475–489. (Cit. on p. 90).

Chraibi, M., Seyfried, A. & Schadschneider, A. (2014). Quantitative validation of the generalized centrifugal force model. In U. Weidmann, U. Kirsch & M. Schreckenberg (Eds.), *Pedestrian and evacuation dynamics 2012* (pp. 603–613). Cham: Springer International Publishing. (Cit. on p. 88).

Daamen, W. (2004). *Modelling passenger flows in public transport facilities* (PhD Thesis). (Cit. on pp. 9, 15, 19, 20, 88, 89).

Daamen, W. & Hoogendoorn, S. P. (2012). Calibration of pedestrian simulation model for emergency doors by pedestrian type. *Transportation Research Record: Journal of the Transportation Research Board*, *2316*, 69–75. (Cit. on p. 90)

Davidich, M. & Köster, G. (2012). Towards automatic and robust adjustment of human behavioral parameters in a pedestrian stream model to measured data. *Safety science*, *50*(5), 1253–1260. (Cit. on pp. 6, 88).

Davidich, M. & Köster, G. (2013). Predicting pedestrian flow: A methodology and a proof of concept based on real-life data. *PLoS ONE*, *8*(12), e83355. (Cit. on pp. 1, 6, 88, 89).

Department of Defense. (1996). *Verification, validation and accreditation (VV&A) Recommended practices guide*. Retrieved from https://acc.dau.mil/adl/en-US/649735/file/71953/Verification,%20validation,%20_%20Accreditation%20(VV_A).pdf. (Cit. on p. 78)

Duives, D. C. (2016). *Analysis and modelling of pedestrian movement dynamics at large-scale events* (PhD Dissertation). Retrieved from https://repository.tudelft.nl/islandora/object/uuid%3A08831f69-9b8e-44cf-8afe-f4a3e7bc9a9c?collection=research. (Cit. on pp. ii, 1, 2, 7, 9, 10, 12–17, 38, 41, 54, 56, 61, 70, 75, 104)

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. (Cit. on pp. 1, 10).

Duives, D. C., Daamen, W. & Hoogendoorn, S. P. (2014). Influence of group size and group composition on the adhered distance headway. *Transportation Research Procedia*, *2*(0), 183–188. (Cit. on p. 10)

Duives, D. C., Daamen, W. & Hoogendoorn, S. P. (2016). Continuum modelling of pedestrian flows—part 2: Sensitivity analysis featuring crowd movement phenomena. *Physica A: Statistical Mechanics and its Applications*, *447*, 36–48. (Cit. on p. 1).

Duives, D. C., Daamen, W. & Hoogendoorn, S. P. (2015). Quantification of the level of crowdedness for pedestrian movements. *Physica A: Statistical Mechanics and its Applications*, *427*, 162–180. (Cit. on pp. 24, 33)

Ehrgott, M. (2005). Scalarization techniques. In *Multicriteria optimization* (pp. 97–126). Berlin, Heidelberg: Springer Berlin Heidelberg. (Cit. on p. 73)

Galea, E. R. (1998). A general approach to validating evacuation models with an application to exodus. *Journal of Fire Sciences*, *16*(6), 414–436. (Cit. on p. 89).

Galea, E. R., Deere, S., Brown, R. & Filippidis, L. (2014). An evacuation validation data set for large passenger ships. In U. Weidmann, U. Kirsch & M. Schreckenberg (Eds.), *Pedestrian and evacuation dynamics 2012* (pp. 109–123). Cham: Springer International Publishing. (Cit. on p. 89).

Geraerts, R. (2010). Planning short paths with clearance using explicit corridors. In *Robotics and automation (icra), 2010 ieee international conference on* (pp. 1997–2004). IEEE. (Cit. on p. 29).

Gorrini, A., Bandini, S. & Sarvi, M. (2014). Group dynamics in pedestrian crowds. *Transportation Research Record: Journal of the Transportation Research Board*, *2421*, 51–56. (Cit. on p. 10)

Gorrini, A., Bandini, S., Sarvi, M., Dias, C. & Shiwakoti, N. (2013). An empirical study of crowd and pedestrian dynamics: The impact of different angle paths and grouping. *Transportation Research Record*, *41*, 42. (Cit. on p. 10).

Helbing, D. & Molnar, P. (1995). Social force model for pedestrian dynamics. *Physical review E*, *51*(5), 4282. (Cit. on p. 14).

Hoogendoorn, S. P. (2001). *Normative pedestrian flow behavior: Theory and applications* (Report No. Vk2001.002). Delft University of Technology. (Cit. on p. 2).

Hoogendoorn, S. P. & Daamen, W. (2007). Microscopic calibration and validation of pedestrian models: Cross-comparison of models using experimental data. In *Traffic and granular flow'05* (pp. 329–340). Springer. (Cit. on pp. 6, 16, 90).

INCONTROL Simulation Solutions. (2016). Pedestrian dynamics: Solutions. Web Page. Web Page. Accessed 16-08-2016. Retrieved from http://www.pedestrian-dynamics.com/crowd-management/blog.html. (Cit. on pp. 1, 29)

Karamouzas, I., Geraerts, R. & Overmars, M. (2009). Indicative routes for path planning and crowd simulation. In *Proceedings of the 4th international conference on foundations of digital games* (pp. 113–120). ACM. (Cit. on p. 30).

Keip, C. & Ries, K. (2009). *Dokumentation von versuchen zur personenstromdynamik - projekt "hermes"*. Retrieved from http://ped.fz-juelich.de/experiments/2009.05.12_Duesseldorf_Messe_Hermes/docu/VersuchsdokumentationHERMES.pdf. (Cit. on pp. 20, 22, 24, 52)

Klein, W., Köster, G. & Meister, A. (2010). Towards the calibration of pedestrian stream models. In R. Wyrzykowski, J. Dongarra, K. Karczewski & J. Wasniewski (Eds.), *Parallel processing and applied mathematics: 8th international conference, ppam 2009, wroclaw, poland, september 13-16, 2009, revised selected papers, part ii* (pp. 521–528). Berlin, Heidelberg: Springer Berlin Heidelberg. (Cit. on pp. 6, 88).

Klügl, F. (2008). A validation methodology for agent-based simulations. In *Proceedings of the 2008 acm symposium on applied computing* (pp. 39–43). ACM. (Cit. on p. 1).

Knoop, V. L. & Buisson, C. (2014). The value of calibration and validation of probabilistic discretionary lane-change models. In *93rd annual meeting transportation research board, washington, usa, 12-16 january 2014; authors version*. TRB. (Cit. on p. 78).

Ko, M., Kim, T. & Sohn, K. (2013). Calibrating a social-force-based pedestrian walking model based on maximum likelihood estimation. *Transportation*, *40*(1), 91–107. (Cit. on p. 90)

Köster, G., Treml, F., Seitz, M. & Klein, W. (2014). Validation of crowd models including social groups. In *Pedestrian and evacuation dynamics 2012* (pp. 1051–1063). Springer. (Cit. on pp. 88–90).

Kretz, T., Grünebohm, A., Kaufman, M., Mazur, F. & Schreckenberg, M. (2006). Experimental study of pedestrian counterflow in a corridor. *Journal of Statistical Mechanics: Theory and Experiment*, *2006*(10), P10001. (Cit. on p. 11).

Kretz, T., Hengst, S. & Vortisch, P. (2008). Pedestrian flow at bottlenecks-validation and calibration of vissim's social force model of pedestrian traffic and its empirical foundations. *arXiv preprint arXiv:0805.1788*. (Cit. on pp. 88, 89).

Law, A. M. (2008). How to build valid and credible simulation models. In *Proceedings of the 40th conference on winter simulation* (pp. 39–47). Winter Simulation Conference. (Cit. on p. 78).

Moussaïd, M., Helbing, D., Garnier, S., Johansson, A., Combe, M. & Theraulaz, G. (2009). Experimental study of the behavioural mechanisms underlying self-organization in human crowds. *Proceedings of the Royal Society B: Biological Sciences*. (Cit. on p. 90).

Moussaïd, M., Helbing, D. & Theraulaz, G. (2011). How simple rules determine pedestrian behavior and crowd disasters. *Proceedings of the National Academy of Sciences*, *108*(17), 6884–6888. (Cit. on pp. 30, 88, 90).

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*(4), e10047. (Cit. on pp. 10, 89).

NIST/SEMATECH. (2012a). E-handbook of statistical methods - 1.3.5.14 anderson-darling test. Retrieved from http://www.itl.nist.gov/div898/handbook/eda/section3/eda35e.htm. (Cit. on p. 34)

NIST/SEMATECH. (2012b). E-handbook of statistical methods - 1.3.5.16 kolmogorov-smirnov goodness-of-fit test. Retrieved from http://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm. (Cit. on p. 34)

Papadimitriou, E., Yannis, G. & Golias, J. (2009). A critical assessment of pedestrian behaviour models. *Transportation Research Part F: Traffic Psychology and Behaviour*, *12*(3), 242–255. (Cit. on p. 1)

PTV Group. (2016). Ptv viswalk use cases. Web Page. Accessed 16-08-2016. Retrieved from http://vision-traffic.ptvgroup.com/en-uk/products/ptv-viswalk/use-cases/. (Cit. on p. 1)

Punzo, V., Montanino, M. & Ciuffo, B. (2015). Do we really need to calibrate all the parameters? variance-based sensitivity analysis to simplify microscopic traffic flow models. *IEEE Transactions on Intelligent Transportation Systems*, *16*(1), 184–193. (Cit. on p. 38).

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*(1), 36–56. (Cit. on pp. 6, 89, 90).

Ronchi, E., Kuligowski, E. D., Reneke, P. A., Peacock, R. D. & Nilsson, D. (2013). *The process of verification and validation of building fire evacuation models* (Report No. Technical Note (NIST TN) - 1822). NIST. (Cit. on pp. 14, 15, 34).

Rudloff, C., Bauer, D., Matyus, T. & Seer, S. (2011). Mind the gap: Boarding and alighting processes using the social force paradigm calibrated on experimental data. In *2011 14th international ieee conference on intelligent transportation systems (itsc)* (pp. 353–358). (Cit. on pp. 1, 89, 90).

Rudloff, C., Matyus, T. & Seer, S. (2014). Comparison of different calibration techniques on simulated data. In U. Weidmann, U. Kirsch & M. Schreckenberg (Eds.), *Pedestrian and evacuation dynamics 2012* (pp. 657–672). Cham: Springer International Publishing. (Cit. on pp. 16, 17).

Rudloff, C., Matyus, T., Seer, S. & Bauer, D. (2011). Can walking behavior be predicted? analysis of calibration and fit of pedestrian models. *Transportation Research Record: Journal of the Transportation Research Board*, (2264), 101–109. (Cit. on pp. 6, 88–90).

Sano, T., Yoshida, Y., Takeichi, N., Kimura, T. & Minegishi, Y. (2011). Experimental study of crowd flow passing through simple-shaped room and validation for an evacuation simulator. In R. D. Peacock, E. D. Kuligowski & J. D. Averill (Eds.), *Pedestrian and evacuation dynamics* (pp. 587–599). Boston, MA: Springer US. (Cit. on p. 88).

Savannah Simulations AG. (2016). Simwalk: Case studies. Web Page. Accessed 16-08-2016. Retrieved from http://www.simwalk.com/downloads/index.html. (Cit. on p. 1)

Schadschneider, A., Eilhardt, C., Nowak, S. & Will, R. (2011). Towards a calibration of the floor field cellular automaton. In *Pedestrian and evacuation dynamics* (pp. 557–566). Springer. (Cit. on pp. 88, 90).

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*, 124–154. (Cit. on p. 1).

Schadschneider, A. & Seyfried, A. (2009). Validation of ca models of pedestrian dynamics with fundamental diagrams. *Cybernetics and Systems: An International Journal*, *40*(5), 367–389. (Cit. on p. 88).

Seer, S., Brändle, N. & Ratti, C. (2014). Kinects and human kinetics: A new approach for studying pedestrian behavior. *Transportation research part C: emerging technologies*, *48*, 212–228. (Cit. on pp. 1, 6, 89, 90).

Seer, S., Rudloff, C., Matyus, T. & Brändle, N. (2014). Validating social force based models with comprehensive real world motion data. In W. Daamen, D. Duives & S. P. Hoogendoorn (Eds.), *The conference on pedestrian and evacuation dynamics 2014 (ped 2014), 22-24 october 2014, delft, the netherlands* (Vol. 2, pp. 724–732). Transportation Research Procedia. Elsevier. (Cit. on p. 90)

Shiwakoti, N., Gong, Y., Shi, X. & Ye, Z. (2015). Examining influence of merging architectural features on pedestrian crowd movement. *Safety Science*, *75*, 15–22. (Cit. on p. 9)

Tang, M. & Jia, H. (2011). An approach for calibration and validation of the social force pedestrian model. In *Transportation, mechanical, and electrical engineering (tmee), 2011 international conference on* (pp. 2026–2031). (Cit. on pp. 6, 90).

Toledo, T., Ben-Akiva, M., Darda, D., Jha, M. & Koutsopoulos, H. (2004). Calibration of microscopic traffic simulation models with aggregate data. *Transportation Research Record: Journal of the Transportation Research Board*, (1876), 10–19. (Cit. on p. 16).

Toledo, T. & Koutsopoulos, H. (2004). Statistical validation of traffic simulation models. *Transportation Research Record: Journal of the Transportation Research Board*, *1876*, 142–150. (Cit. on pp. 13–15, 34).

Wagoum, A. U. K. & Seyfried, A. (2013). Conception, development, installation and evaluation of a real time evacuation assistant for complex buildings. *Procedia - Social and Behavioral Sciences*, *104*, 728–736. (Cit. on p. 88).

Weichen, L., Chraibi, M., Seyfried, A., Zhang, J., Xiaoping, Z. & Ying, Z. (2014). Validation of fds+evac for pedestrian simulations in wide bottlenecks. In *17th international ieee conference on intelligent transportation systems (itsc)* (pp. 554–559). (Cit. on pp. 6, 88–90).

Wolinski, D., J Guy, S., Olivier, A.-H., Lin, M., Manocha, D. & Pettré, J. (2014). Parameter estimation and comparative evaluation of crowd simulations. In *Computer graphics forum* (Vol. 33, pp. 303–312). Wiley Online Library. (Cit. on pp. 7, 16, 17).

Yang, X., Daamen, W., Hoogendoorn, S. P., Chen, Y. & Dong, H. (2014). Breakdown phenomenon study in the bidirectional pedestrian flow. *Transportation Research Procedia*, *2*, 456–461. (Cit. on p. 10)

Zak, S. H. & Chong, E. K. P. (2013). Multiobjective optimization. In *An introduction to optimization* (Chap. 24, pp. 577–598). Somerset, Unites States: Wiley. (Cit. on pp. 13, 56).

Zhang, J., Klingsch, W., Rupprecht, T., Schadschneider, A. & Seyfried, A. (2012). Empirical study of turning and merging of pedestrian streams in t-junction. In *Fourth international symposium on agent-based modeling and simulation*. (Cit. on p. 42).

Zhang, J., Klingsch, W., Schadschneider, A. & Seyfried, A. (2011). Transitions in pedestrian fundamental diagrams of straight corridors and t-junctions. *Journal of Statistical Mechanics: Theory and Experiment*, *2011*(06), P06004. (Cit. on p. 9).

Zhang, J., Klingsch, W., Schadschneider, A. & Seyfried, A. (2012). Ordering in bidirectional pedestrian flows and its influence on the fundamental diagram. *Journal of Statistical Mechanics: Theory and Experiment*, *2012*(02), P02002. (Cit. on pp. 10, 20).

Appendix

A Overview metrics 87

B Scenarios sensitivity analysis 91

C Results quantitative sensitivity analysis 93
C.1 Bidirectional straight - high . . . 93
C.2 Bidirectional straight - low . . . 94
C.3 Bottleneck . . . 95
C.4 Corner - high . . . 96
C.5 Corner - low . . . 97
C.6 T-junction - high . . . 98
C.7 T-junction- low . . . 99

D Scenarios calibration 100

E Measurement periods calibration 102

F Objective functions and normalization method 104

G Reference data - Metrics 106

H Detailed calibration results 108
H.1 Correlation between mean travel time error and pedestrian count error . . . 108
H.2 How is the decrease in GoF related to the changes in the errors? . . . 108
H.3 Tables split out per individual objective . . . 113

I Calibration - different objective functions 121

J Decrease in GoF versus the distance from the optimal parameter set 122

A | Overview metrics

The table below contains an overview of metrics, found in the literature, which are used to evaluate the performance of a pedestrian simulation model. As subsection 2.2.3 describes, the metrics are categorized based on three properties, namely:

1. Aggregation level: Is the metric macroscopic, mesoscopic or microscopic?
2. Quantitative or qualitative: Is the metric obtained quantitatively or qualitatively?
3. General or specific: Can the metric be used independent of the flow situation or is it used only in specific flow situations?

Table A.1: Overview of metrics, found in the literature, used to evaluate the performance of a pedestrian simulation model. The metrics are categorized by: 1) Aggregation level (Macro, Meso or Micro). 2) Whether they are obtained quantitatively (Quan.) or qualitatively (Qual.). and 3) If they are used only in a specific flow situation (Spec.) or independent of the flow situation (Gen.).

Metric	**Macro**	**Meso**	**Micro**	**Quan.**	**Qual.**	**Spec.**	**Gen.**	**Studies**
Formation of lanes	x				x	x		Campanella et al. (2014)
Funnel shape upstream of bottleneck	x				x	x		Campanella et al. (2014), Berrou et al. (2007)
Merging patterns	x				x	x		Wagoum and Seyfried (2013)
Stop & Go waves	x				x	x		Moussaïd et al. (2011)
Turbulence	x				x	x		Moussaïd et al. (2011)
Zipper effect	x				x	x		Campanella et al. (2014)
Density	x			x			x	Weichen et al. (2014), Berrou et al. (2007), Davidich and Köster (2013), Daamen (2004), Abdelghany et al. (2016)
FD (Flow-density)	x			x			x	Bandini et al. (2014), Berrou et al. (2007), Schadschneider and Seyfried (2009), Klein et al. (2010), Davidich and Köster (2012), Davidich and Köster (2013)
FD (Speed-density)	x			x			x	Bandini et al. (2014), Chraibi et al. (2014), Köster et al. (2014), Moussaïd et al. (2011), Rudloff, Matyus, Seer and Bauer (2011), Schadschneider et al. (2011, 2011), Campanella et al. (2014, 2009b)
Flow	x			x			x	Berrou et al. (2007), Kretz et al. (2008), Sano et al. (2011), Abdelghany et al. (2016)
Ped count per region per step vs. the capacity of the region	x			x			x	Banerjee and Kraemer (2010)
Pedestrian within area over time	x			x			x	Daamen (2004), Bauer (2011)
Speed over space	x			x			x	Weichen et al. (2014)

Continued on next page

Table A.1 – continued from previous page

Metric	Macro	Meso	Micro	Quan.	Qual.	Spec.	Gen.	Studies
Speed vs distance headway	x			x			x	Chattaraj et al. (2013)
Bottleneck capacity	x			x		x		Campanella et al. (2014), Weichen et al. (2014), Campanella et al. (2009b)
Crossing time (Boarding and Alighting)	x			x		x		Rudloff, Bauer et al. (2011)
Egress speed vs. group size	x			x		x		Köster et al. (2014)
Lane formation	x			x		x		Campanella et al. (2009b)
Occupants out vs. time	x			x		x		Galea (1998), Galea et al. (2014)
Speed vs. group size	x			x		x		Moussaïd et al. (2010)
Time vs. nr. of passengers alighted or boarded	x			x		x		Berrou et al. (2007)
Total evacuation time	x			x		x		Galea et al. (2014), Kretz et al. (2008)
Distribution ped trajectories		x		x			x	Davidich and Köster (2013)
Probability of choosing an alternative		x		x			x	Robin et al. (2009)
Speed		x		x			x	Davidich and Köster (2013)
Travel time		x		x			x	Campanella et al. (2014), Bauer (2011), Rudloff, Matyus, Seer and Bauer (2011), Seer, Brändle and Ratti (2014), Bauer (2011)
Walking speed alighting or boarding (Mean, variance and vs distance)		x		x		x		Daamen (2004)
Walking speed stairs up or down (Mean, variance and distribution)		x		x		x		Daamen (2004)
Average distance between pedestrians			x		x		x	Campanella et al. (2014)

Continued on next page

Table A.1 – continued from previous page

Metric	Macro	Meso	Micro	Quan.	Qual.	Spec.	Gen.	Studies
Group cohesion			x		x		x	Köster et al. (2014)
Number of collisions with nearby pedestrians			x		x		x	Campanella et al. (2014)
Staying in simulated area			x		x		x	Campanella et al. (2014)
Trajectories			x		x		x	Moussaïd et al. (2011), Seer, Brändle and Ratti (2014), Weichen et al. (2014)
Number of collisions straight on			x		x	x		Campanella et al. (2014)
Pushed back pedestrians			x		x	x		Campanella et al. (2014)
Pushed towards the obstacles			x		x	x		Campanella et al. (2014)
Trapped pedestrians			x		x	x		Campanella et al. (2014)
Acceleration curve			x	x			x	Moussaïd et al. (2009)
Choice sets			x	x			x	Robin et al. (2009)
Trajectories (Acceleration)			x	x			x	Campanella et al. (2011), Hoogendoorn and Daamen (2007), Daamen and Hoogendoorn (2012), Ko et al. (2013), Seer, Rudloff et al. (2014), Tang and Jia (2011)
Trajectories (Position)			x	x			x	Seer, Brändle and Ratti (2014), Rudloff, Matyus, Seer and Bauer (2011), Chraibi et al. (2016), Tang and Jia (2011), Schadschneider et al. (2011)
Wall and pedestrian overlap			x	x			x	Rudloff, Bauer et al. (2011)

B | Scenarios sensitivity analysis

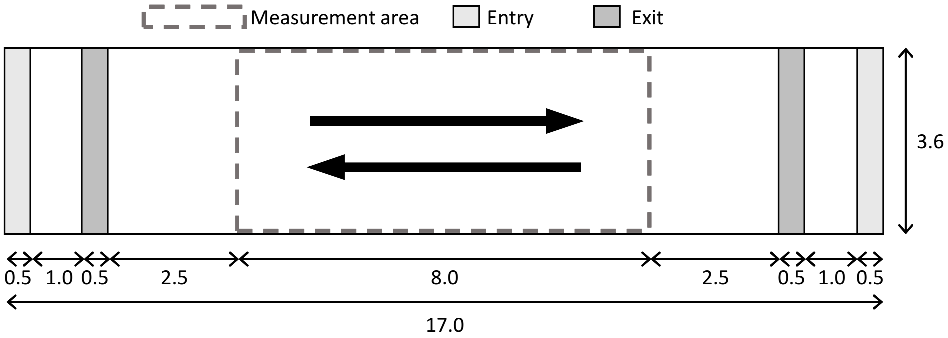


Figure B.1: Overview bidirectional scenario

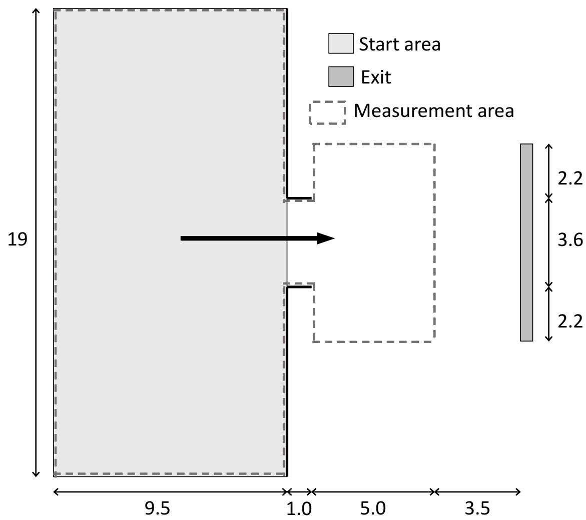


Figure B.2: Overview bottleneck scenario

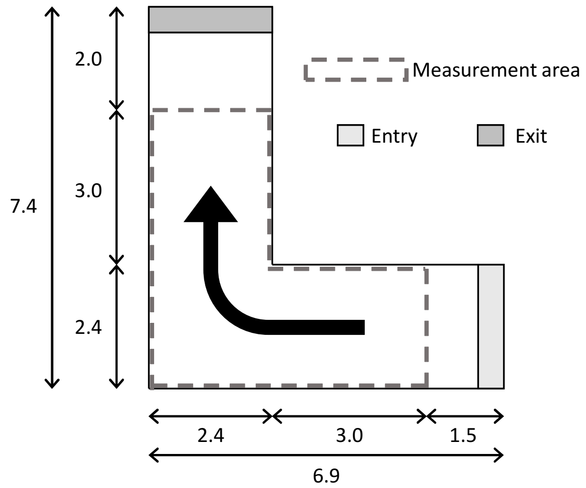


Figure B.3: Overview corner scenario

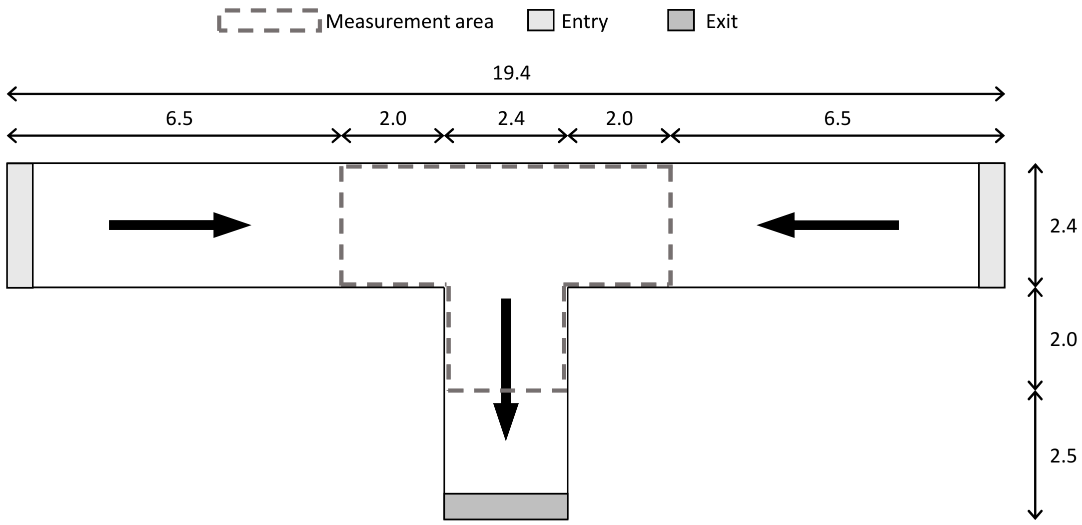


Figure B.4: Overview T-junction scenario

Table B.1: Inflows per scenario in ped/s/m/entry

	Low	High
Bidirectional	0.2	0.4
Bottleneck	*	
Corner	0.5	1.0
T-junction	0.4	0.8

* Start area filled with a density of 3 $\mathrm{ped/m^2}$

C Results quantitative sensitivity analysis

C.1 Bidirectional straight - high

Table C.1: Deviation from the default value [%] - Bidirectional straight high

	Mean	Std.	N
Personal distance -25%	1.07	-1.13	131
Personal distance +25%	0.33	0.00	130
Side pref. update factor -25%	0.33	-0.19	109
Side pref. update factor +25%	0.79	-0.94	122
Relaxation time -25%	9.43	-9.74	100
Relaxation time +25%	-21.12	21.41	262
Min. desired speed -25%	0.50	-0.04	107
Min. desired speed +25%	0.96	-0.82	110
Preferred clearance -25%	0.44	-1.04	115
Preferred clearance +25%	1.04	-0.98	100
FoV avoidance range -25%	0.76	0.02	118
Viewing angle -25%	3.14	-4.13	100
Viewing angle +25%	-1.52	3.79	121

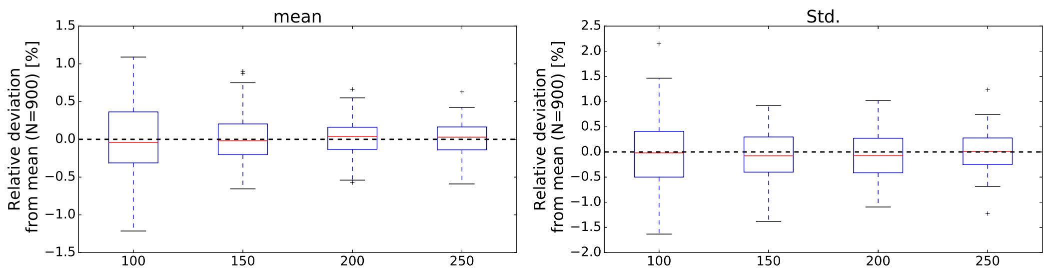


Figure C.1: Influence of seeds versus the number of replications - Bidirectional straight high

C.2 Bidirectional straight - low

Table C.2: Deviation from the default value [%] - Bidirectional straight low

	Mean	Std.	N
Personal distance -25%	-0.23	1.80	50
Personal distance +25%	-0.36	2.47	50
Side pref. update factor -25%	0.10	0.77	50
Side pref. update factor +25%	-0.50	1.89	50
Relaxation time -25%	0.66	-0.89	50
Relaxation time +25%	-2.37	5.11	71
Preferred clearance -25%	0.21	0.38	50
Preferred clearance +25%	-0.08	0.41	59
FoV avoidance range -25%	0.13	0.71	50
FoV avoidance range +25%	0.09	0.16	52
Viewing angle -25%	-0.32	-0.13	50
Viewing angle +25%	-0.86	2.31	50

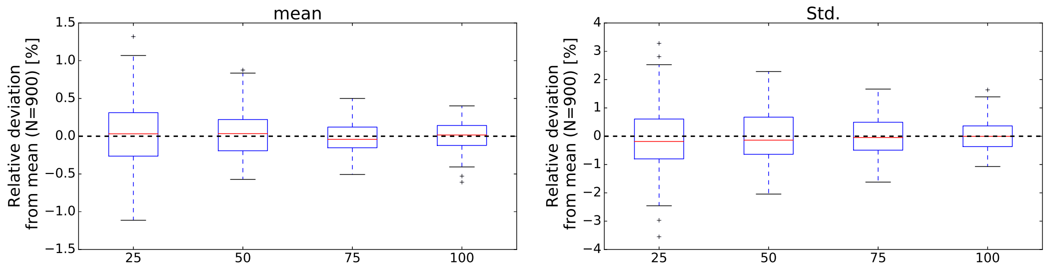


Figure C.2: Influence of seeds versus the number of replications - Bidirectional straight low

C.3 Bottleneck

Table C.3: Deviation from the default value [%] - Bottleneck

	Mean	Std.	N
Personal distance -25%	1.13	-1.35	50
Personal distance +25%	-1.43	0.94	50
Side pref. update factor -25%	-0.57	-0.21	50
Side pref. update factor +25%	-0.16	-0.17	50
Relaxation time -25%	17.48	8.63	50
Relaxation time +25%	-13.06	-6.96	53
Min. desired speed -25%	-0.47	-0.69	61
Min. desired speed +25%	-0.17	0.25	64
Preferred clearance -25%	-0.48	-0.32	56
Preferred clearance +25%	0.43	0.21	66
FoV avoidance range -25%	-0.57	0.26	53
Viewing angle -25%	7.59	2.03	54
Viewing angle +25%	-3.44	0.31	50

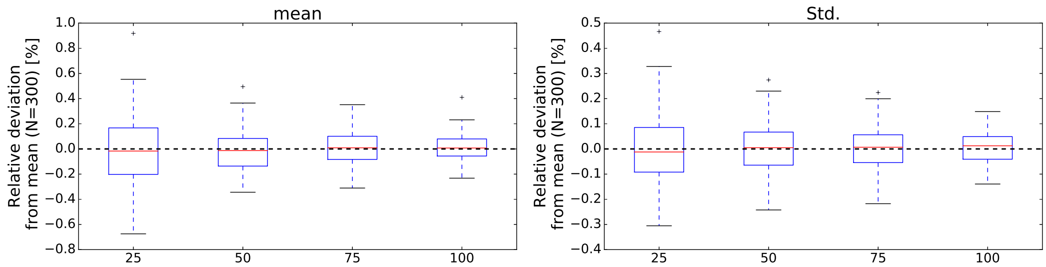


Figure C.3: Influence of seeds versus the number of replications - Bottleneck

C.4 Corner - high

Table C.4: Deviation from the default value [%] - Corner high

	Mean	Std.	N
Personal distance -25%	0.28	-2.26	50
Personal distance +25%	-0.66	0.71	64
Side pref. update factor -25%	-0.20	-0.26	67
Side pref. update factor +25%	-0.03	-0.77	65
Relaxation time -25%	5.81	-5.29	50
Relaxation time +25%	-10.98	10.32	64
Min. desired speed -25%	-0.46	-0.35	54
Min. desired speed +25%	0.24	-1.31	63
Preferred clearance -25%	0.74	0.59	50
Preferred clearance +25%	-1.05	0.56	75
FoV avoidance range -25%	0.29	-1.56	53
Viewing angle -25%	-1.90	3.66	50
Viewing angle +25%	0.27	-1.83	50

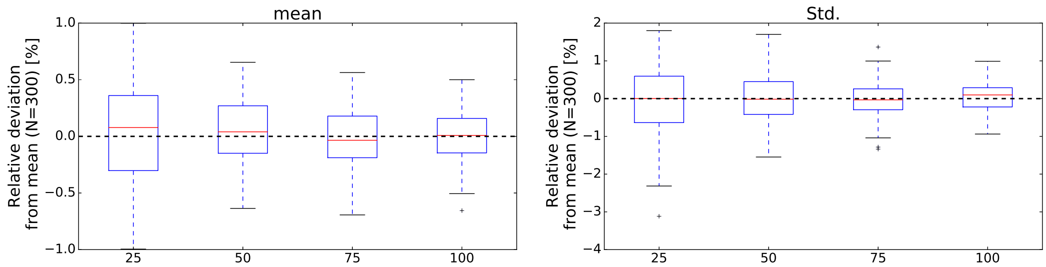


Figure C.4: Influence of seeds versus the number of replications - Corner high

C.5 Corner - low

Table C.5: Deviation from the default value [%] - Corner low

	Mean	Std.	N
Personal distance -25%	-0.58	-0.23	52
Personal distance +25%	-0.40	0.31	50
Side pref. update factor -25%	-0.29	-0.76	52
Side pref. update factor +25%	0.10	-0.08	57
Relaxation time -25%	1.79	1.17	50
Relaxation time +25%	-2.04	1.27	50
Preferred clearance -25%	0.28	-0.49	50
Preferred clearance +25%	-0.28	-0.32	50
FoV avoidance range -25%	0.77	-2.10	50
FoV avoidance range +25%	0.18	-0.36	50
Viewing angle -25%	-0.25	0.43	50
Viewing angle +25%	0.65	-1.10	58

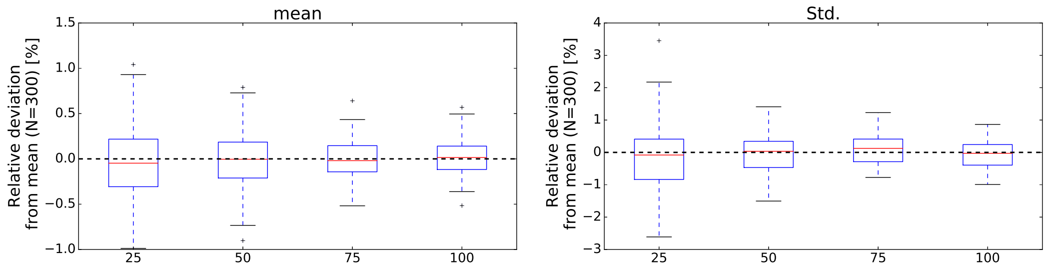


Figure C.5: Influence of seeds versus the number of replications - Corner low

C.6 T-junction - high

Table C.6: Deviation from the default value [%] - T-junction high

	Mean	Std.	N
Personal distance -25%	0.43	-1.52	75
Personal distance +25%	-0.48	1.40	75
Side pref. update factor -25%	-1.54	-0.82	75
Side pref. update factor +25%	0.95	0.46	75
Relaxation time -25%	72.87	-8.03	88
Relaxation time +25%	-17.35	-7.16	75
Min. desired speed -25%	0.18	-0.60	75
Min. desired speed +25%	0.80	0.54	75
Preferred clearance -25%	-0.50	-0.22	75
Preferred clearance +25%	-0.39	-0.44	75
FoV avoidance range -25%	-0.68	-0.02	75
Viewing angle -25%	7.92	-0.90	75
Viewing angle +25%	-3.69	1.55	75

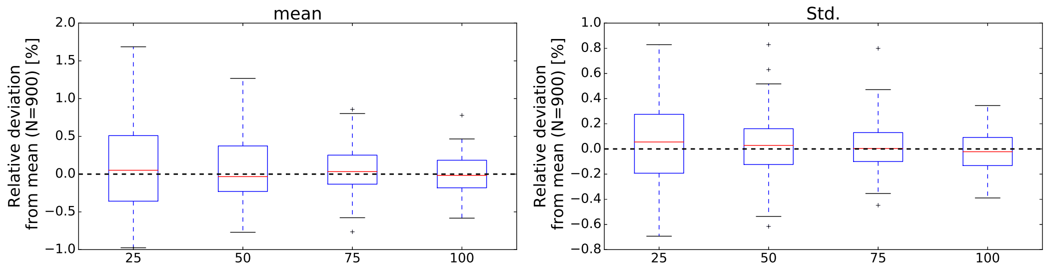


Figure C.6: Influence of seeds versus the number of replications - T-junction high

C.7 T-junction- low

Table C.7: Deviation from the default value [%] - T-junction low

	Mean	Std.	N
Personal distance -25%	0.02	-0.78	75
Personal distance +25%	-0.31	0.55	75
Side pref. update factor -25%	-0.29	0.88	75
Side pref. update factor +25%	-0.25	-0.09	75
Relaxation time -25%	6.44	-11.37	75
Relaxation time +25%	-9.28	16.33	75
Preferred clearance -25%	0.33	-0.11	75
Preferred clearance +25%	-0.41	-0.23	75
FoV avoidance range -25%	-0.20	-0.13	75
FoV avoidance range +25%	-0.22	-0.43	75
Viewing angle -25%	0.66	-1.36	75
Viewing angle +25%	-0.22	1.18	75

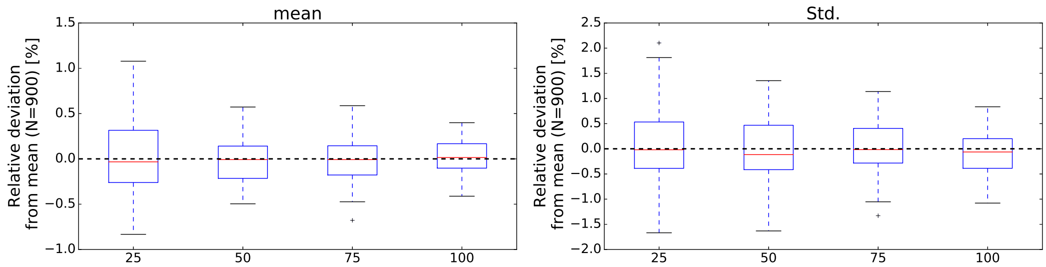


Figure C.7: Influence of seeds versus the number of replications - T-junction low

D | Scenarios calibration

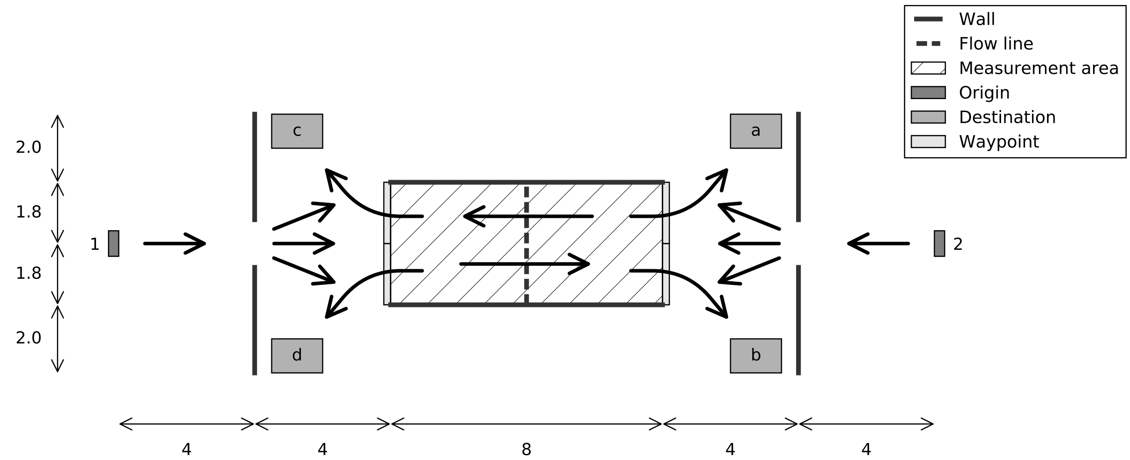


Figure D.1: Overview bidirectional scenario

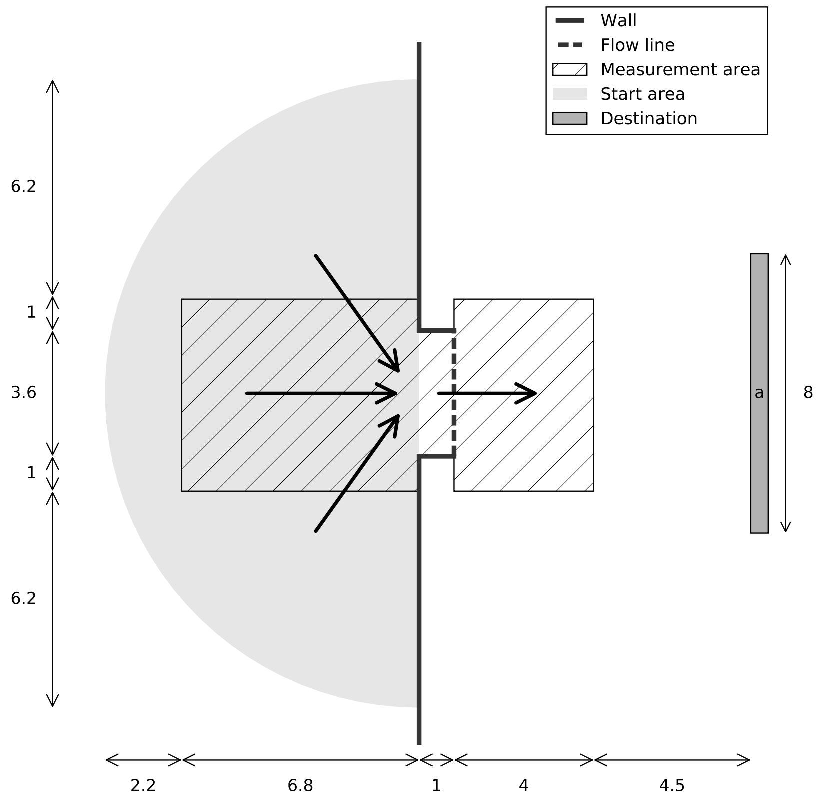


Figure D.2: Overview bottleneck scenario

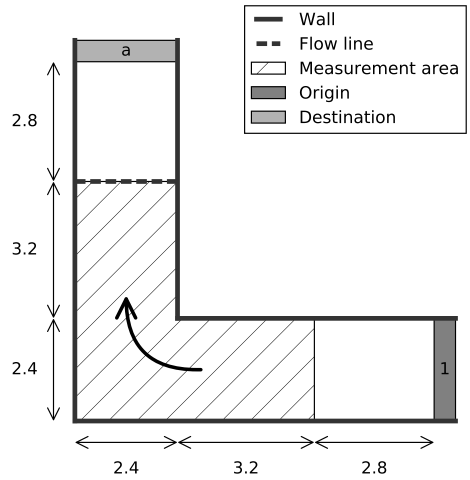


Figure D.3: Overview corner scenario

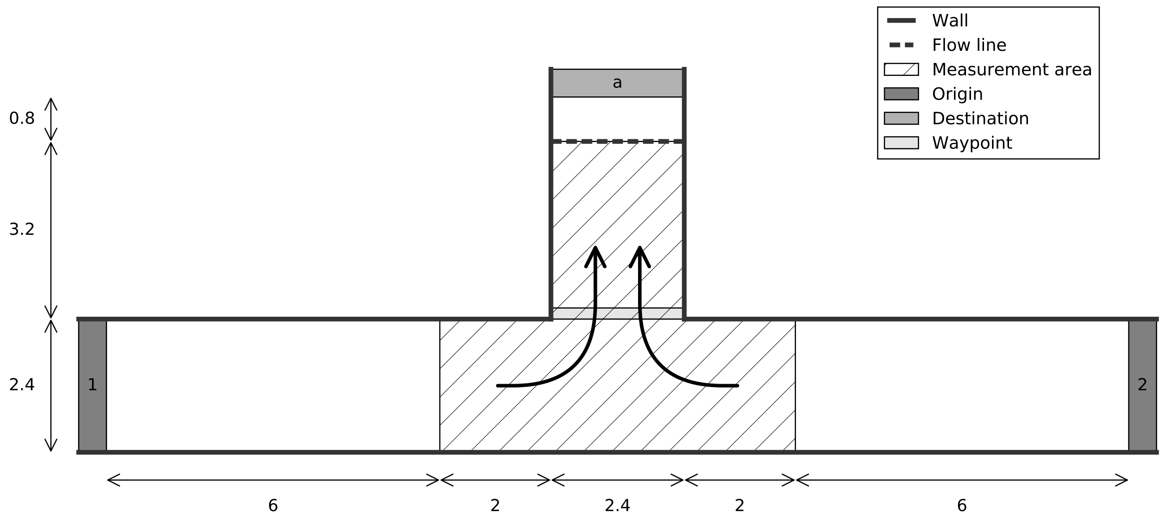


Figure D.4: Overview T-junction scenario

E | Measurement periods calibration

The measurement period determines the time frame within which the different metrics are measured. As this appendix will show, there is a difference between the scenarios and the metrics regarding how the measurement period is determined. In the case of the scenarios there is a difference between the bottleneck scenarios and all other scenarios. This is due to the fact that the bottleneck scenario is fully populated at the start of the simulation whilst all other scenarios need a warming up period to populate the infrastructure with the desired density level. The difference between the metrics is due to how and where they are measured.

The duration of the measurement period for the simulations is determined based on the reference data. This is the case for all scenarios except the bottleneck due to the fact that the bottleneck scenario is the only scenario based on a fixed number of pedestrians. The start time of the measurement period is determined per individual simulation and in the case of the simulations the end time is the start time plus the duration. Table E.1 shows, for every metric, an overview of how the start and end of the measurement period is determined. Table E.2 shows the durations of the measurement periods, determined based on the reference data, per scenario and metric. In the remainder of this appendix it is explained in more detail why the measurement periods are determined as shown in Table E.2.

Table E.1: Measurement periods

	Bottleneck scenario		**Other scenarios**	
	Start	*End*	*Start*	*End*
Flow	First pedestrian passes flow line	Last pedestrian passes flow line	First pedestrian passes flow line + free travel time	Last pedestrian enters the measurement area
Spatial distribution	First pedestrian passes flow line	Last pedestrian exits the measurement area	First pedestrian exits the measurement area	Last pedestrian enters the measurement area
Effort	First pedestrian passes flow line	Last pedestrian exits the measurement area	First pedestrian exits the measurement area	Last pedestrian enters the measurement area
Travel time	First pedestrian passes flow line	Last pedestrian exits the measurement area	First pedestrian exits the measurement area	Last pedestrian enters the measurement area

Table E.2: Measurement period duration in seconds

	Flow	**Spatial Distr.**	**Effort**	**Travel time**
Bidirectional - high	43.3750	40.8125	46.0000	46.0000
Bidirectional - low	27.2500	24.6875	29.8750	29.8750
Corner - high	23.3750	23.3750	28.6875	28.6875
Corner - low	29.5625	29.5625	34.8750	34.8750
T-junction - high	54.1875	54.1875	58.6875	58.6875
T-junction - low	36.6250	36.6250	41.1250	41.1250

Flow

Start: Time the first pedestrian passes flow line + free travel time.

End: Time the last pedestrian enters the measurement area.

Free travel time: The time it takes a pedestrian to travel from the start of the measurement area to the flow line when walking at the average preferred speed.

Why: This should ensure that a pedestrian passing the flow line during the measurement period experiences the desired density level when passing the flow line.

Notes: In the case of the bidirectional scenarios the start time is not determined based on the first pedestrian that passes the flow line. However, for both travel directions, the time the first pedestrian passes the line is taken and the maximum of those two is used. In the case of the end time the minimum time of the two directions is used.

Bottleneck: In the case of the bottleneck scenario, the average flow is determined by the time it takes all 349 pedestrian to pass the bottleneck. Hence, the measurement period is simply the time between the first pedestrian passing the flow line and the last pedestrian passing the flow line.

Spatial distribution

Start: Time the first pedestrian exists the measurement area + free travel time.

End: Time the last pedestrian enters the measurement area.

Free travel time: The time it takes a pedestrian to travel from the start of the measurement area to the end of the measurement area when walking at the average preferred speed.

Why: This should ensure that all steps of all pedestrians, taken into account, happen whilst the density is at the desired level.

Note: In the case of the bidirectional scenarios the start time is not determined based on the first pedestrian that exists the measurement area. However, for both travel directions, the time the first pedestrian exists the measurement area is taken and the maximum of those two is used. In the case of the end time the minimum time of the two directions is used.

Bottleneck: In the case of the bottleneck scenario the spatial distribution is based during the period the measurement area contains at least one pedestrian.

Effort and travel time

Start: Time the first pedestrian exists the measurement area.

End: Time the last pedestrian enters the measurement area.

Why: This should ensure that only the effort or travel time of those pedestrians that experienced the desired level of density is taken into account.

Note: In the case of the bidirectional scenarios the start time is not determined based on the first pedestrian that exists the measurement area. However, for both travel directions, the time the first pedestrian exists the measurement area is taken and the maximum of those two is used. In the case of the end time the minimum time of the two directions is used.

Bottleneck: In the case of the bottleneck scenario the effort and travel time of all pedestrians is taken into account.

F Objective functions and normalization method

As subsection 5.1.3 states, the question is: How to combine the results of two or more objective functions into a single objective function? As the subsection also states, the choice is made to use the weighted sum method to combine the results of the individual objective into a single objective function. However, in order to combine the results of different metrics in a meaningful way they need to be normalized given the differences in units and orders of magnitude. This appendix will explain the method which is used and the rationale behind it.

Roughly speaking one has two options when determining the error between the simulation results and the reference data:

1. Absolute error $(M_{sim} - M_{ref})$
2. Relative error $\left(\frac{M_{sim}}{M_{ref}} - 1\right)$

The choice between these two options is the choice between two different sets of assumptions.

The first set of assumptions is related to how the size of the deviation $(M_{sim} - M_{ref})$ is related to the size of the error given the reference value (M_{ref}). In the case of the absolute error one assumes that the relationship between the size of the deviation and the size of the error is not dependent on the size of the reference value (i.e. a deviation of 1 unit, for example 1 ped/s/m in the case of the flow, is equally wrong regardless of the reference value). In the case of the relative error one assumes that the same deviation is less wrong in the case of a higher reference value compared with a lower reference value.

The second set of assumptions is related to how the size of the deviation of one metric is related to the size of the deviation in another metric. In the case of the relative error one assumes that the deviation of one times the reference value of one metric is equally wrong as the deviation of one times the reference value of another metric. In the case of the absolute error one assumes that the deviation of 1 unit of one metric is equally wrong as a deviation of 1 unit of another metric (e.g. a deviation of 1 ped/s/m is equally wrong as a deviation of 1 m/s in the mean effort).

On top of the two sets of assumptions presented above, one has to also take the following practical point into account when making a choice between the two options. In the case of the relative error a problem can occur when the reference value is zero given that the error will become infinite. This situation will occur in the case of the spatial distribution metrics where there are cells which are never occupied. Unless one is of the opinion that any deviation from a reference value of zero constitutes an infinitely large error one has come up with another method to determine the size of the error.

Taking the assumptions and the practical point presented above into account the choice is made to use the absolute error for the following reasons:

1. It avoids the practical problem described above.
2. The assumption that, for an equally large deviation, the error is smaller in the case of a scenario with a large reference value compared to a scenario with a smaller reference value, assumes different levels of accuracy between the different scenarios. Given the goal of calibrating a model for general use, this is not considered to be a be a very good assumption. Certainly, given that it might even be preferable to gain more accurate predictions near the capacity point than a point somewhere in the free-flow branch given the large impact small deviations can have within this area of the fundamental diagram.
3. The problem with the different scales and units can be solved by the normalizing.

However, this choice leads to another problem. How to normalize the errors? The only example found in the pedestrian modelling literature was that of (Duives, 2016). In the case of Duives (2016), the normalization is done based on the maximum error for a given scenario and metric.

$$E_{norm;s;m}(\theta) = \frac{E_{s;m}(\theta)}{\max\limits_{\theta \in \Theta} E_{s;m}} \tag{F.1}$$

Where $E_{norm;s;m}(\theta))$ is the normalized error for parameters set θ, scenario s and metric m. Θ is the set of parameter sets. This method assumes that the maximum error of a scenario and metric is equally wrong as every other maximum error of every other combination of scenario and metric. So, for example, an error in the flow of 1 ped/s/m in one scenario is considered equally wrong as an error of 1.5 ped/s/m in another scenario. Or, a larger error in the flow is considered equally wrong as a small error in the mean effort. This assumption is very similar to the assumption made when using the relative error as hence is problematic for the same reason.

To take away a part of the problem, namely the assumption that for an equally large deviation the normalized error can differ depending on the size of the maximum error in the given scenario, one could normalize based on the maximum value per metric over all scenarios:

$$E_{norm;s;m}(\theta) = \frac{E_{s;m}(\theta)}{\max\limits_{\theta \in \Theta} \max\limits_{s \in S} E_{s;m}} \tag{F.2}$$

Where S is the set of all scenarios.

However, this method still has one problem: The maximum error in metric A is equally wrong as the maximum error of metric B whilst the maximum error of metric A can be, relatively speaking, many times larger than the maximum error of metric B. To prevent this problem one could come up with error values for every metric which one considers to be 100% wrong. So, for example, one could consider an error in the flow of 1 ped/s/m as equally wrong as an error of 1 s/m in the travel time. However, on what should these values be based?

An option would be to base it on the ratios between the values of the metrics from the reference data. In Table F.1 the ratios per scenario can be found whereby the ratios are calculated based on equating the flow to 1 (So in the table below, an error of 1 ped/s/m is considered equal to an error of 0.8783 s/m in the travel time mean in the case of the bidirectional low density scenario).

Table F.1: Ratios between metric values per scenario

	B-H	**B-L**	**B**	**C-H**	**C-L**	**T-H**	**T-L**
Flow	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Spatial distr.	0.3110	0.1250	0.1267	0.0499	0.3184	0.0630	0.3356
TT mean	1.2581	0.8783	0.7783	0.8377	1.2732	1.0581	0.8537
TT std.	0.3006	0.1006	0.3872	0.1043	0.1522	0.2606	0.1454
Effort mean	0.0342	0.0507	0.0191	0.0399	0.0900	0.0256	0.0446
Effort std.	0.0089	0.0128	0.0048	0.0101	0.0154	0.0063	0.0083

However, as one can see these ratios differ per scenarios. So, there are two options:

1. Use different values for different scenarios
2. Take the mean ratios and base the normalization values on that

The first option would involve running into the same problem as using the relative errors, namely: the size of the error depends on the reference value so, for example, an error of 1 ped/s/m in on scenario equals an error of 1.5 ped/s/m in another. Hence, this is not considered a good option. The second option would lead to the normalization values displayed in Table 5.2. Overall, this method (using the second option) is considered to be a better method than using the maximum values given that is also, approximately, takes into account the relative differences between the metrics.

G | Reference data - Metrics

Acronyms used in the tables for the data sets:

	Name	Data set
B-H	Bidirectional high	BOT-360-200-200
B-L	Bidirectional low	BOT-360-075-075
B	Bottleneck	AO-360-400
C-H	Corner - high	EO-240-150-240
C-L	Corner - low	EO-240-060-240
T-H	T-junction - high	KO-240-150-240
T-L	T-junction - low	KO-240-060-240

Table G.1: Measured flows [ped/s/m] per data set

	Total	**l2r**	**r2l**
B-H	1.27	0.63	0.64
B-L	0.87	0.48	0.39
B	2.32		
C-H	1.16		
C-L	0.58		
T-H	1.62		
T-L	1.07		

Table G.2: Measured effort [m/s] per data set

	Mean	**Std.**	**N**
B-H	0.0434	0.0113	182
B-L	0.0439	0.0111	74
B	0.0422	0.0112	349
C-H	0.0462	0.0117	67
C-L	0.0520	0.0089	41
T-H	0.0415	0.0102	217
T-L	0.0477	0.0089	95

Table G.3: Measured travel time per data set

	Mean	**Std.**	**N**	**Mean**	**Std.**	**Length**
	[s/m]			**[s]**		**[m]**
B-H	1.59	0.381	150	13.81	3.300	8.66
B-L	0.76	0.087	73	6.17	0.707	8.12
B	1.80	0.896	349	17.00	8.455	9.43
C-H	0.97	0.121	63	7.75	0.966	7.99
C-L	0.74	0.088	40	5.49	0.656	7.46
T-H	1.72	0.423	204	12.00	2.956	6.99
T-L	0.91	0.155	93	5.66	0.964	6.20

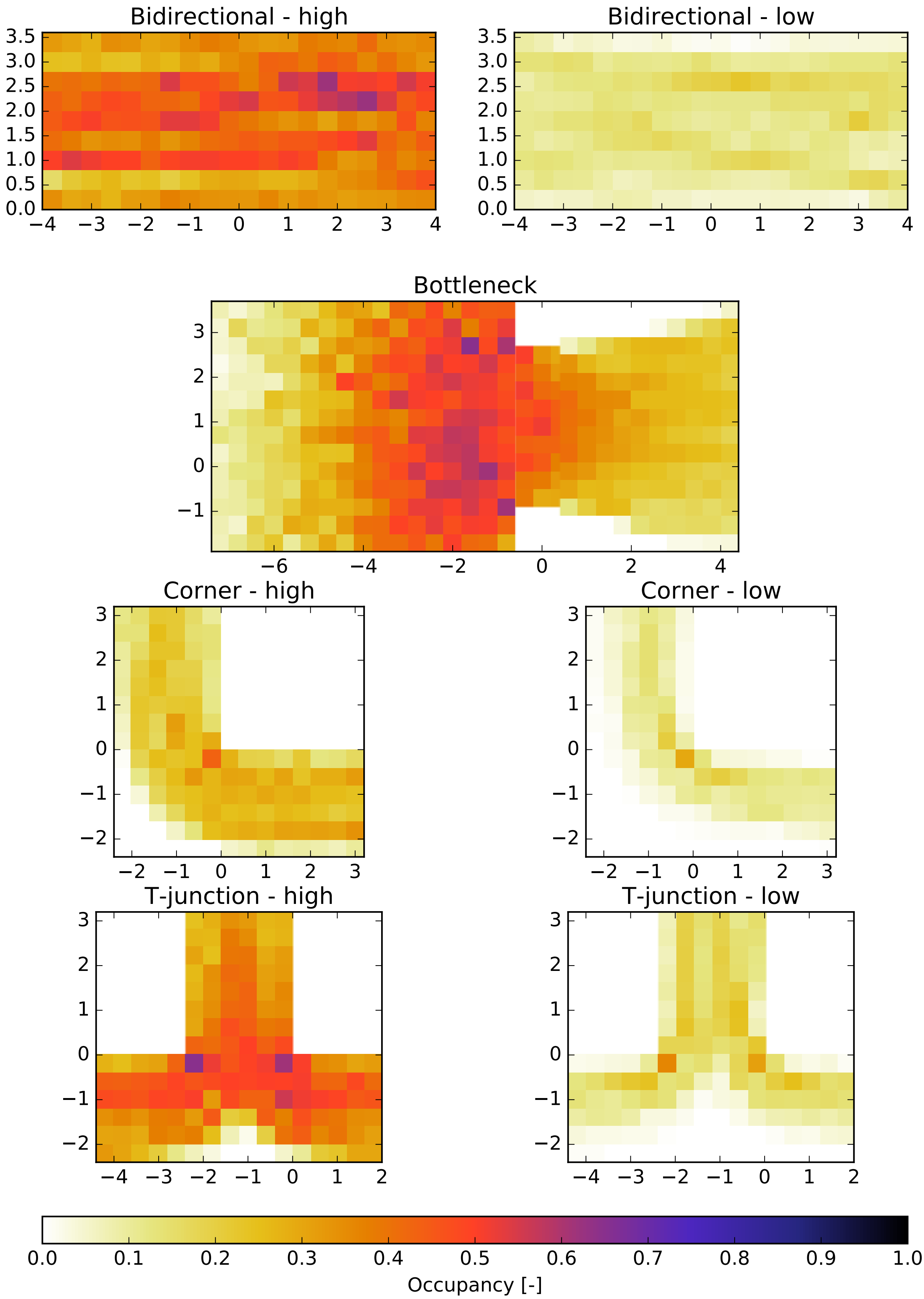


Figure G.1: Spatial distributions

H Detailed calibration results

In this appendix a more detailed picture of the calibration results is presented. Section H.1 presents how the mean travel time error and the pedestrian count error are correlated in the case of the bidirectional high density scenario. Section H.2 discusses how the decrease in GoF is related to the average normalized error. And finally, section H.3 presents the calibration results split into their individual objectives.

H.1 Correlation between mean travel time error and pedestrian count error

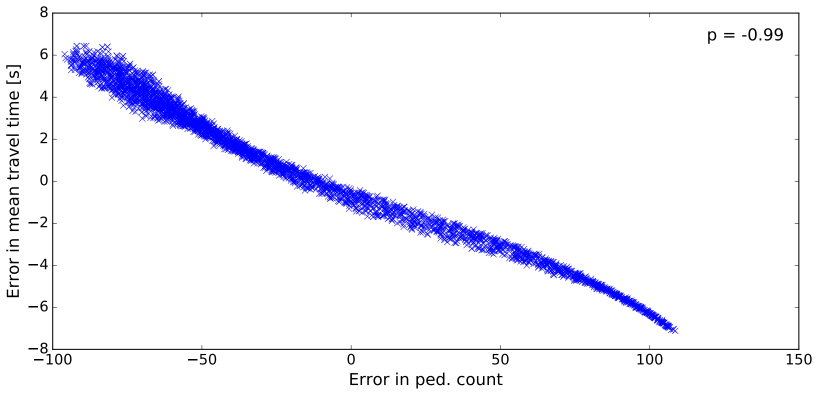


Figure H.1: Correlation between mean travel time error and pedestrian count error - Bidirectional high density scenario (p is the Pearson correlation coefficient)

H.2 How is the decrease in GoF related to the changes in the errors?

This section presents how the decreases in GoF, shown in tables 5.7 to 5.10, are related to the changes in the errors of the individual objectives (a combination of a single metric and scenario). The changes in the errors by means of the changes in the average error and the change in the standard deviation of the errors. How these are computed will be shown by the following example:

Table H.1: An overview of the changes in errors of the combination of low density cases when, instead of its own optimal parameter set, the optimal parameter set of the bidirectional low density scenario is used

			Metrics						
			Q		SD	Eff$_\mu$	Eff$_\sigma$	TT$_\mu$	TT$_\sigma$
Scenarios	**B-L**	***Base***	***-7.83***	***0.99***	***15.81***	***-40.28***	***54.68***	***-3.53***	***1.82***
		B-L	-8.29	1.06	14.44	-42.54	45.49	-0.91	8.95
	C-L	***Base***		***1.96***	***17.48***	***-62.80***	***-38.49***	***-0.81***	***-0.67***
		B-L		2.33	20.73	-65.47	-42.77	1.84	1.95
	T-L	***Base***		***4.81***	***27.90***	***-35.91***	***-20.08***	***-12.89***	***-19.44***
		B-L		4.61	32.53	-40.21	-30.81	-3.13	-3.12

The table above shows the normalized errors, expressed as percentages of the normalization values, for the combination of the low density scenarios. The ***base*** rows show the errors made when the optimal parameter set obtained using the combination of the low density scenarios is used. The ***B-L*** rows show the errors made when the optimal parameter set obtained using the bidirectional low density scenario is used. When the optimal parameter, obtained using the combination of the low density scenarios, is used the average error is $\pm 19.54\%$ of the normalization values. This number is calculated as follows:

$$\bar{E}_{A:B} = \frac{1}{N_s + N_m} \sum_s \sum_m \left| \frac{M_{sim;s;m}(\theta^*_B) - M_{ref;s;m}}{M_{norm;m}} * 100\% \right| \tag{H.1}$$

Where $M_{sim;s;m}(\theta^*_B)$ is the simulated value of scenario s and metric m for the optimal parameter set of B. $M_{ref;s;m}$ is the reference value of scenario s and metric m and $M_{norm;m}$ is the normalization value of metric m. N_s and N_m are respectively the number of scenarios and metrics. The absolute values of the errors are used because the interest is in the average size of the errors. Furthermore, as the table shows, the metrics are not combined as is the case in the calibration but are treated separately (e.g. the mean and the standard deviation of the effort are not combined into a single value). Similarly, the standard deviation of the errors is also based on the absolute values.

So, the errors shown in the table above result in the following average errors and standard deviations:

		Error stats.		**Deviation from base**	
	ΔGoF	**Mean**	**Std.**	**Mean**	**Std.**
Base		***19.38***	***19.12***		
B-L	-0.0081	19.54	20.01	0.16	0.89

The remainder of this appendix will present the tables showing how the decreases in the GoF, presented in tables 5.7 to 5.10, are related to the change in the average size of the errors and the variance of the errors.

H.2.1 Combined scenarios

Table H.2: An overview of the decreases in GoF and the accompanying descriptive statistics of the errors whereby the base case is the combination of the high density scenarios

		Error stats.		**Deviation from base**	
	ΔGoF	**Mean**	**Std.**	**Mean**	**Std.**
Base		***37.78***	***33.64***		
B-H	-0.0136	41.41	27.45	3.62	-6.19
B	-0.1257	44.31	44.67	6.53	11.03
C-H	-0.1013	43.99	40.69	6.21	7.05
T-H	-0.0135	37.40	35.85	-0.38	2.21

Table H.3: An overview of the decreases in GoF and the accompanying descriptive statistics of the errors whereby the base case is the combination of the low density scenarios

		Error stats.		**Deviation from base**	
	ΔGoF	**Mean**	**Std.**	**Mean**	**Std.**
Base		***19.38***	***19.12***		
B-L	-0.0081	19.54	20.01	0.16	0.89
C-L	-0.0207	20.10	25.67	0.72	6.55
T-L	-0.0187	19.82	25.39	0.44	6.27

H.2.2 Individual scenarios

Table H.4: An overview of the decreases in GoF and the accompanying descriptive statistics of the errors whereby the base case is the high density bidirectional scenario

		Error stats.		**Deviation from base**	
	ΔGoF	**Mean**	**Std.**	**Mean**	**Std.**
Base		***29.19***	***17.08***		
H-D	-0.3528	52.38	53.07	23.18	35.99
B	-0.4084	54.60	56.55	25.41	39.47
C-H	-0.3289	53.83	48.11	24.64	31.03
T-H	-0.3907	54.34	55.28	25.15	38.20
L-D	-0.3907	50.95	49.63	21.76	32.55
B-L	-0.3149	49.14	51.45	21.50	34.37

Table H.5: An overview of the decreases in GoF and the accompanying descriptive statistics of the errors whereby the base case is the bottleneck scenario

		Error stats.		**Deviation from base**	
	ΔGoF	**Mean**	**Std.**	**Mean**	**Std.**
Base		***22.38***	***13.30***		
H-D	-0.1937	39.16	25.03	16.78	11.73
B-H	-0.1223	42.89	21.05	20.51	7.75
C-H	-0.1533	35.21	23.53	12.83	10.23
T-H	-0.2679	44.67	27.18	22.29	13.88
L-D	-0.1262	37.27	15.21	5.02	-1.87

Table H.6: An overview of the decreases in GoF and the accompanying descriptive statistics of the errors whereby the base case is the high density corner scenario

		Error stats.		**Deviation from base**	
	ΔGoF	**Mean**	**Std.**	**Mean**	**Std.**
Base		***21.65***	***3.68***		
H-D	-0.0402	26.43	21.19	4.78	17.51
B-H	-0.0992	33.99	22.55	12.34	18.87
B	-0.0501	30.30	10.37	8.65	6.69
T-H	-0.0369	24.19	23.69	2.54	20.01
L-D	-0.0433	28.91	15.33	7.26	11.65
C-L	-0.0032	22.40	3.64	0.75	-0.04

Table H.7: An overview of the decreases in GoF and the accompanying descriptive statistics of the errors whereby the base case is the high density t-junction scenario

		Error stats.		**Deviation from base**	
	ΔGoF	**Mean**	**Std.**	**Mean**	**Std.**
Base		***23.60***	***17.23***		
H-D	-0.0548	30.73	21.47	7.14	4.24
B-H	-0.4743	61.59	39.34	38.00	22.10
B	-0.6858	68.24	61.70	44.65	44.47
C-H	-0.5646	63.65	57.55	40.05	40.31
L-D	-0.4185	56.14	48.60	32.54	31.37
T-L	-0.5869	64.75	58.45	41.15	41.22

Table H.8: An overview of the decreases in GoF and the accompanying descriptive statistics of the errors whereby the base case is the combination of the low density bidirectional scenario

		Error stats.		Deviation from base	
	ΔGoF	**Mean**	**Std.**	**Mean**	**Std.**
Base		***17.39***	***18.81***		
L-D	-0.0093	17.85	21.25	0.46	2.44
C-L	-0.0978	22.70	37.47	5.31	18.66
T-L	-0.0924	22.30	36.81	4.91	18.00
H-D	-0.0055	18.30	19.94	0.92	1.13
B-H	-0.0026	17.25	19.09	-0.14	0.28

Table H.9: An overview of the decreases in GoF and the accompanying descriptive statistics of the errors whereby the base case is the combination of the low density corner scenario

		Error stats.		Deviation from base	
	ΔGoF	**Mean**	**Std.**	**Mean**	**Std.**
Base		***19.53***	***23.26***		
L-D	-0.0110	20.37	25.48	0.84	2.22
B-L	-0.0245	22.52	26.52	2.99	3.26
T-L	-0.0003	19.57	23.37	0.04	0.11
H-D	-0.0194	21.43	26.41	1.90	3.15
C-H	-0.0019	19.75	23.58	0.22	0.32

Table H.10: An overview of the decreases in GoF and the accompanying descriptive statistics of the errors whereby the base case is the combination of the low density t-junction scenario

		Error stats.		Deviation from base	
	ΔGoF	**Mean**	**Std.**	**Mean**	**Std.**
Base		***17.19***	***11.47***		
L-D	-0.0164	20.17	10.92	2.98	-0.55
B-L	-0.0366	19.07	17.23	1.88	5.75
C-L	-0.0011	17.65	11.38	0.46	-0.10
H-D	-0.0250	18.95	14.82	1.76	3.35
T-H	-0.0258	17.67	16.47	0.48	4.99

H.2.3 Metrics

Table H.11: An overview of the decreases in GoF and the accompanying descriptive statistics of the errors whereby the base case is the flow metric

		Error stats.		Deviation from base	
	ΔGoF	**Mean**	**Std.**	**Mean**	**Std.**
Base		***9.10***	***6.91***		
SD	-0.0844	16.61	24.62	7.51	17.70
Eff	-0.0902	16.75	25.54	7.65	18.63
TT	-0.0079	11.24	8.47	2.14	1.56
Macro	-0.0697	15.09	22.89	5.99	15.98
Meso	-0.0079	11.24	8.47	2.14	1.56
All	-0.0697	14.77	22.85	5.68	15.94

Table H.12: An overview of the decreases in GoF and the accompanying descriptive statistics of the errors whereby the base case is the spatial distribution metric

		Error stats.		Deviation from base	
	ΔGoF	**Mean**	**Std.**	**Mean**	**Std.**
Base		***32.53***	***16.81***		
Q	-0.1281	39.68	31.92	7.15	15.11
Eff	-0.0198	33.65	20.05	1.12	3.24
TT	-0.1235	39.68	26.00	7.14	9.20
Macro	-0.0029	32.92	17.37	0.38	0.56
Meso	-0.1235	39.68	26.00	7.14	9.20
All	-0.0120	33.40	17.06	0.86	0.25

Table H.13: An overview of the decreases in GoF and the accompanying descriptive statistics of the errors whereby the base case is the effort metric

		Error stats.		Deviation from base	
	ΔGoF	**Mean**	**Std.**	**Mean**	**Std.**
Base		***37.89***	***19.76***		
Q	-0.1213	41.62	37.12	3.72	17.36
SD	-0.0228	42.58	15.16	4.69	-4.59
TT	-0.0412	45.25	13.24	7.35	-6.52
Macro	-0.0261	42.21	17.30	4.32	-2.46
Meso	-0.0412	45.25	13.24	7.35	-6.52
All	-0.0548	42.86	14.70	4.96	-5.06

Table H.14: An overview of the decreases in GoF and the accompanying descriptive statistics of the errors whereby the base case is the travel time metric

		Error stats.		Deviation from base	
	ΔGoF	**Mean**	**Std.**	**Mean**	**Std.**
Base		***27.15***	***34.05***		
Q	-0.1635	39.59	45.01	12.44	10.96
SD	-0.1454	27.06	52.26	-0.10	18.20
Eff	-0.0596	26.65	42.78	-0.50	8.72
Macro	-0.1416	30.20	49.96	3.05	15.91
Meso	0.0000	27.15	34.05	0.00	0.00
All	-0.0223	24.87	43.14	-2.28	9.08

H.3 Tables split out per individual objective

The results, found in tables 5.7 to 5.10 are split-out by either metric or scenario and checks are performed if the main conclusions of section 5.3 still hold. The results are discussed in the same order as they are in section 5.3. So, subsection H.3.1 discusses if the results of subsection 5.3.3 still hold, subsection H.3.2 discusses if the results of subsection 5.3.4 still hold and subsection H.3.3 discusses if the results of subsection 5.3.5 still hold.

H.3.1 Movement base cases - detailed results

The main conclusions of subsection 5.3.3 were as follows:

1. The GoF of the individual scenarios decreases, compared to the optimal GoF, when the parameter set of the combined set is used.
2. The effect, described above, is larger in the case of the high density levels compared to the low density cases.
3. The GoF of the individual scenarios decreases, compared to the optimal GoF, when the parameter set of another scenario of the same density level is used.
4. The effect, described above, is larger in the case of the high density levels compared to the low density cases.
5. At low densities the corner and t-junction scenarios are interchangeable.
6. No good fit can be obtained on simultaneously the t-junction and bidirectional high density scenarios.

Tables H.15 to H.18 show the results split-out by metric. The data in the tables shows that the conclusions, listed above, do indeed hold when just one of the metrics is used except for the last two conclusions. The data shows that, although using the optimal parameter set of the low density t-junction scenario does produces only a very small decrease in the GoF of the corner scenario (compared to the bidirectional low scenario) this is not the case the other way round except when the flow in used as the only metric. This could be explained by the fact that as Figure 5.2 clearly shows, the errors for the individual metrics in the case of the corner low density scenario are generally smaller and vary less compared to the low density t-junction scenario with the exception of the effort. Whether or not good fit can be obtained on simultaneously the t-junction and bidirectional high density also depends on the used metric. The model primarily seems to have a problem with obtaining the simultaneously good fit in the case of the travel time. When one of the other three metrics is used is does not seem to be that much of a problem.

Two other notable things that the four tables below show are: 1) When the travel time is used as the sole metric, the t-junction and corner high density scenarios seem interchangeable. And 2) he fit of the model to the travel time in the case of the bidirectional high density scenario is very bad when the optimal parameter set of any of the other scenarios is used.

So, the major patterns do not depend on the specific metric used (i.e. the first for conclusions listed above). However, the more specific patterns, the last two conclusions, do seem to depend on the specific (combination of) metric(s) used.

Table H.15: Difference in the GoF, as determined by Equation 5.10, for the comparisons between different movement base cases using the **flow** as the only metric. The combinations are identified by their acronyms as listed in Table 5.6. Every column shows the decrease in GoF, compared to the optimal GoF, of that specific combination if the optimal parameter set of the combination, defined by the row, is used.

Used parameter set	Predicted combination: H-D	B-H	B	C-H	T-H	L-D	B-L	C-L	T-L
H-D	X	-0.0074	-0.0066	-0.0327	-0.0496				
B-H	-0.0142	X	-0.0696	-0.0368	-0.0467				
B	-0.0046	-0.0285	X	-0.0375	-0.0487				
C-H	-0.2507	-0.0598	-0.9772	X	-0.0623				
T-H	-0.1326	-0.0312	-0.5783	-0.0175	X				
L-D						X	-0.0004	-0.0001	-0.0006
B-L						-0.0004	X	-0.0376	-0.0487
C-L						-0.0003	-0.0005	X	-0.0014
T-L						-0.0019	-0.0065	-0.0002	X

Table H.16: Difference in the GoF, as determined by Equation 5.10, for the comparisons between different movement base cases using the **spatial distribution** as the only metric. The combinations are identified by their acronyms as listed in Table 5.6. Every column shows the decrease in GoF, compared to the optimal GoF, of that specific combination if the optimal parameter set of the combination, defined by the row, is used.

Used parameter set	Predicted combination: H-D	B-H	B	C-H	T-H	L-D	B-L	C-L	T-L
H-D	X	-0.0280	-0.2821	-0.0182	-0.0202				
B-H	-0.5287	X	-0.1856	-0.0966	-1.0501				
B	-0.6224	-0.3634	X	-0.1338	-1.2101				
C-H	-0.3124	-0.0395	-0.4074	X	-0.0205				
T-H	-0.2993	-0.0320	-0.3728	-0.0099	X				
L-D						X	-0.0122	-0.0012	0.0000
B-L						-0.0984	X	-0.0253	-0.0990
C-L						-0.0821	-0.0383	X	-0.0369
T-L						-0.0616	-0.0126	-0.0010	X

Table H.17: Difference in the GoF, as determined by Equation 5.10, for the comparisons between different movement base cases using the **effort** as the only metric. The combinations are identified by their acronyms as listed in Table 5.6. Every column shows the decrease in GoF, compared to the optimal GoF, of that specific combination if the optimal parameter set of the combination, defined by the row, is used.

Used parameter set	Predicted combination: H-D	B-H	B	C-H	T-H	L-D	B-L	C-L	T-L
H-D	X	-0.1886	-0.0995	-0.1335	-0.0149				
B-H	-0.0345	X	-0.2146	-0.2790	-0.0810				
B	-0.0892	-0.5282	X	-0.2093	-0.0557				
C-H	-0.1077	-0.8370	-0.0053	X	-0.0250				
T-H	-0.0577	-0.5378	-0.0375	-0.0919	X				
L-D						X	-0.0435	-0.0288	-0.0457
B-L						-0.0213	X	-0.0897	-0.0923
C-L						-0.0369	-0.2120	X	-0.0167
T-L						-0.0793	-0.3504	-0.0054	X

Table H.18: Difference in the GoF, as determined by Equation 5.10, for the comparisons between different movement base cases using the **travel time** as the only metric. The combinations are identified by their acronyms as listed in Table 5.6. Every column shows the decrease in GoF, compared to the optimal GoF, of that specific combination if the optimal parameter set of the combination, defined by the row, is used.

		Predicted combination								
		H-D	B-H	B	C-H	T-H	L-D	B-L	C-L	T-L
Used parameter set	H-D	X	-0.0344	-0.2147	-0.0248	-0.9209				
	B-H	-0.0526	X	-0.2874	-0.0302	-1.0877				
	B	-0.4665	-2.7176	X	-0.0205	-0.3226				
	C-H	-0.5022	-2.3073	-0.8963	X	0.0000				
	T-H	-0.5022	-2.3073	-0.8963	0.0000	X				
	L-D						X	-0.0007	-0.0004	-0.0006
	B-L						-0.0051	X	-0.0001	-0.0168
	C-L						-0.0077	-0.0082	X	-0.0168
	T-L						-0.0108	-0.0341	-0.0002	X

H.3.2 Density levels - detailed results

The main conclusions of subsection 5.3.4 were as follows:

1. In the case of the individual scenarios, the decrease in GoF is far smaller in the case that the optimal parameter set of the high density scenario is used in the low density scenario (of the same movement base case) than vice versa.

2. The difference in GoF are largest for the t-junction scenarios.

3. The use of the optimal parameter set obtained using the combination of high density scenarios in the combination of low density scenario leads to a small decrease in the GoF. This is not the case vice versa.

Tables H.19 to H.25 show the results split-out by metric. Regarding the first conclusion listed above, it holds for all metrics except the effort. In the case of the effort the decrease in GoF of the bidirectional low density scenario is larger than the decrease in the case of the high density scenario. Furthermore, the difference between the t-junction scenarios is also very small.

The second conclusion listed above holds when the macroscopic metrics are use but not when the mesoscopic metrics are used. In the case of the effort the t-junction low density scenario has indeed the largest decrease in GoF of the low density scenarios. However, in the case of the high density scenarios the largest decrease in GoF is found in the case of the corner scenario. In the case of the travel time the largest decreases in GoF are found for the bidirectional scenarios.

The last conclusion listed above holds for the flow and the travel time. However, in the case of the spatial distribution and the effort the decrease in the GoF of the combination of low density scenarios is not small. It is however clearly smaller than the decrease in the GoF of the combination of high density scenarios.

The main conclusion that it is more important to include the high density scenarios does seem to hold regardless of the metric or metrics used certainly when multiple movement base cases are used. However, the data also shows that it does not necessarily holds for every combination of movement base case and metric.

Table H.19: Difference in the GoF, as determined by Equation 5.10, for the comparisons between different density levels for every movement base case using the **flow** as the only metric. The combinations are identified by their acronyms as listed in Table 5.6. Every column shows the decrease in GoF, compared to the optimal GoF, of that specific combination if the optimal parameter set of the combination, defined by the row, is used.

Used parameter set	Predicted combination: B-H	B-L	C-H	C-L	T-H	T-L
B-H		-0.0004				
B-L	-0.0721					
C-H				-0.0002		
C-L			-0.0255			
T-H						-0.0013
T-L					-0.1191	

Table H.20: Difference in the GoF, as determined by Equation 5.10, for the comparisons between different density levels for combinations of movement base cases using the **flow** as the only metric. The combinations are identified by their acronyms as listed in Table 5.6. Every column shows the decrease in GoF, compared to the optimal GoF, of that specific combination if the optimal parameter set of the combination, defined by the row, is used. (U.p.s = Used parameter set)

U.p.s.	Predicted combination: H-D	L-D	B-H	B	C-H	T-H	B-L	C-L	T-L
H-D	X	-0.0001	-0.0074	-0.0066	-0.0327	-0.0496	-0.0002	-0.0002	-0.0011
L-D	-0.0545	X	-0.0002	-0.2519	-0.0278	-0.0346	-0.0004	-0.0001	-0.0006

Table H.21: Difference in the GoF, as determined by Equation 5.10, for the comparisons between different density levels for every movement base case using the **spatial distribution** as the only metric. The combinations are identified by their acronyms as listed in Table 5.6. Every column shows the decrease in GoF, compared to the optimal GoF, of that specific combination if the optimal parameter set of the combination, defined by the row, is used.

Used parameter set	Predicted combination: B-H	B-L	C-H	C-L	T-H	T-L
B-H		-0.0192				
B-L	-0.0585					
C-H				-0.0137		
C-L			-0.1500			
T-H						-0.0961
T-L					-0.8790	

Table H.22: Difference in the GoF, as determined by Equation 5.10, for the comparisons between different density levels for combinations of movement base cases using the **spatial distribution** as the only metric. The combinations are identified by their acronyms as listed in Table 5.6. Every column shows the decrease in GoF, compared to the optimal GoF, of that specific combination if the optimal parameter set of the combination, defined by the row, is used. (U.p.s = Used parameter set)

U.p.s.	Predicted combination: H-D	L-D	B-H	B	C-H	T-H	B-L	C-L	T-L
H-D	X	-0.0404	-0.0280	-0.2821	-0.0182	-0.0202	-0.0055	-0.0259	-0.1032
L-D	-0.2231	X	-0.0213	-0.3184	-0.0358	-0.8655	-0.0122	-0.0012	0.0000

Table H.23: Difference in the GoF, as determined by Equation 5.10, for the comparisons between different density levels for every movement base case using the **effort** as the only metric. The combinations are identified by their acronyms as listed in Table 5.6. Every column shows the decrease in GoF, compared to the optimal GoF, of that specific combination if the optimal parameter set of the combination, defined by the row, is used.

		Predicted combination					
		B-H	**B-L**	**C-H**	**C-L**	**T-H**	**T-L**
Used parameter set	**B-H**		-0.0196				
	B-L	-0.0133					
	C-H				-0.0078		
	C-L			-0.0948			
	T-H						-0.0203
	T-L					-0.0223	

Table H.24: Difference in the GoF, as determined by Equation 5.10, for the comparisons between different density levels for combinations of movement base cases using the **effort** as the only metric. The combinations are identified by their acronyms as listed in Table 5.6. Every column shows the decrease in GoF, compared to the optimal GoF, of that specific combination if the optimal parameter set of the combination, defined by the row, is used. (U.p.s = Used parameter set)

		Predicted combination								
		H-D	**L-D**	**B-H**	**B**	**C-H**	**T-H**	**B-L**	**C-L**	**T-L**
U.p.s.	**H-D**	X	-0.0256	-0.1886	-0.0995	-0.1335	-0.0149	-0.1033	-0.0497	-0.0416
	L-D	-0.0892	X	-0.5282	0.0000	-0.2093	-0.0557	-0.0435	-0.0288	-0.0457

Table H.25: Difference in the GoF, as determined by Equation 5.10, for the comparisons between different density levels for every movement base case using the **travel time** as the only metric. The combinations are identified by their acronyms as listed in Table 5.6. Every column shows the decrease in GoF, compared to the optimal GoF, of that specific combination if the optimal parameter set of the combination, defined by the row, is used.

		Predicted combination					
		B-H	**B-L**	**C-H**	**C-L**	**T-H**	**T-L**
Used parameter set	**B-H**		-0.0023				
	B-L	-1.5712					
	C-H				-0.0001		
	C-L			-0.0108			
	T-H						-0.0020
	T-L					-0.0334	

Table H.26: Difference in the GoF, as determined by Equation 5.10, for the comparisons between different density levels for combinations of movement base cases using the **travel time** as the only metric. The combinations are identified by their acronyms as listed in Table 5.6. Every column shows the decrease in GoF, compared to the optimal GoF, of that specific combination if the optimal parameter set of the combination, defined by the row, is used. (U.p.s = Used parameter set)

		Predicted combination								
		H-D	**L-D**	**B-H**	**B**	**C-H**	**T-H**	**B-L**	**C-L**	**T-L**
U.p.s.	**H-D**	X	-0.0017	-0.0344	-0.2147	-0.0248	-0.9209	-0.0008	0.0000	-0.0059
	L-D	-0.1216	X	-1.0595	-0.0411	-0.0160	-0.5648	-0.0007	-0.0004	-0.0006

H.3.3 Metrics - detailed results

The main conclusions of subsection 5.3.5 were as follows:

1. There seems to be some correlation between the spatial distribution and effort.
2. The spatial distribution dominates the flow when both are used.
3. The travel time dominates the effort when both are used.
4. The model cannot obtain a good fit on all four metrics simultaneously.

Tables H.27 to H.33 show the results split-out by scenario. The first conclusion listed above does hold for the low density scenario but not for the high density scenarios. This begs the question if the correlation found when combining all scenarios is due to the inclusion of the low density scenarios or that is would also be apparent if one were to combine the four high density scenarios.

In none of the individual cases the spatial distribution seems do dominate the flow as the decrease in the GoF of the flow is generally not much larger when the optimal parameter set obtained using the two macroscopic metrics is used. The travel time also does not seem to dominate the effort in most of the cases. Only in case of the bidirectional scenarios and the t-junctions low density scenario this is somewhat the case. Hence, both seem to be the result of combining multiple movement base cases.

The last conclusion does hold for all seven scenarios. In none of the cases the model is capable of simultaneously obtaining a good fit on all four metrics. This does indicate that it is important to prioritize which metrics are most important given the application of the model as one cannot necessarily obtain a good fit on all metrics.

So, overall the most important conclusion that the model cannot obtain a good fit on all four metrics simultaneously does not seem to be dependent on the used scenario or combination of them. The other conclusion, however, do seem to depend on the exact combination of scenarios used.

Table H.27: Difference in the GoF, as determined by Equation 5.10, for the comparisons between different metrics using the **bidirectional high density** scenario as the sole scenarios. The combinations are identified by their acronyms as listed in Table 5.6. Every column shows the decrease in GoF, compared to the optimal GoF, of that specific combination if the optimal parameter set of the combination, defined by the row, is used.

		Predicted combination			
		Q	**SD**	**Eff**	**TT**
Used parameter set	**Q**	X	-0.0811	-0.7682	-0.6197
	SD	-0.0031	X	-0.7488	-0.1252
	Eff	-0.0642	-0.1221	X	-1.8546
	TT	-0.0167	-0.0656	-0.1434	X
	Macro	-0.0007	-0.0020	-1.1050	-0.2234
	Meso	-0.0251	-0.1001	-0.0650	-0.0171
	All	-0.0129	-0.0527	-0.0959	-0.0193

Table H.28: Difference in the GoF, as determined by Equation 5.10, for the comparisons between different metrics using the **bidirectional low density** scenario as the sole scenarios. The combinations are identified by their acronyms as listed in Table 5.6. Every column shows the decrease in GoF, compared to the optimal GoF, of that specific combination if the optimal parameter set of the combination, defined by the row, is used.

		Predicted combination			
		Q	**SD**	**Eff**	**TT**
Used parameter set	**Q**	X	-0.0417	-0.1226	-0.0092
	SD	-0.0060	X	-0.0018	-0.0864
	Eff	-0.0070	-0.0018	X	-0.1010
	TT	-0.0002	-0.0119	-0.1054	X
	Macro	-0.0014	-0.0013	-0.0410	-0.0357
	Meso	-0.0002	-0.0298	-0.0164	-0.0040
	All	-0.0083	-0.1908	-0.1236	-1.2741

Table H.29: Difference in the GoF, as determined by Equation 5.10, for the comparisons between different metrics using the **bottleneck** scenario as the sole scenarios. The combinations are identified by their acronyms as listed in Table 5.6. Every column shows the decrease in GoF, compared to the optimal GoF, of that specific combination if the optimal parameter set of the combination, defined by the row, is used.

		Predicted combination			
		Q	**SD**	**Eff**	**TT**
Used parameter set	**Q**	X	-0.0601	-0.0620	-0.5983
	SD	-0.1644	X	-0.0053	-0.9130
	Eff	-0.3284	-0.0838	X	-1.7291
	TT	-0.5000	-0.2620	-0.2375	X
	Macro	-0.0001	-0.0365	-0.1144	-0.4472
	Meso	-0.4709	-0.4320	-0.0047	-0.0184
	All	-0.0443	-0.1301	-0.0109	-0.0654

Table H.30: Difference in the GoF, as determined by Equation 5.10, for the comparisons between different metrics using the **corner high density** scenario as the sole scenarios. The combinations are identified by their acronyms as listed in Table 5.6. Every column shows the decrease in GoF, compared to the optimal GoF, of that specific combination if the optimal parameter set of the combination, defined by the row, is used.

		Predicted combination			
		Q	**SD**	**Eff**	**TT**
Used parameter set	**Q**	X	-0.0139	-0.2822	-0.0036
	SD	-0.0062	X	-0.2197	-0.0021
	Eff	-0.0295	-0.0410	X	-0.0427
	TT	-0.0001	-0.0209	-0.2480	X
	Macro	-0.0062	0.0000	-0.2197	-0.0021
	Meso	-0.0295	-0.0410	0.0000	-0.0427
	All	-0.0278	-0.0318	-0.0083	-0.0378

Table H.31: Difference in the GoF, as determined by Equation 5.10, for the comparisons between different metrics using the **corner low density** scenario as the sole scenarios. The combinations are identified by their acronyms as listed in Table 5.6. Every column shows the decrease in GoF, compared to the optimal GoF, of that specific combination if the optimal parameter set of the combination, defined by the row, is used.

		Predicted combination			
		Q	**SD**	**Eff**	**TT**
Used parameter set	**Q**	X	-0.0124	-0.0598	0.0000
	SD	-0.0003	X	-0.0183	-0.0007
	Eff	-0.0002	-0.0072	X	-0.0005
	TT	-0.0001	-0.0151	-0.0545	X
	Macro	-0.0003	0.0000	-0.0183	-0.0007
	Meso	-0.0002	-0.0072	0.0000	-0.0005
	All	-0.0004	-0.0011	-0.0024	-0.0006

Table H.32: Difference in the GoF, as determined by Equation 5.10, for the comparisons between different metrics using the **t-junction high density** scenario as the sole scenarios. The combinations are identified by their acronyms as listed in Table 5.6. Every column shows the decrease in GoF, compared to the optimal GoF, of that specific combination if the optimal parameter set of the combination, defined by the row, is used.

		Predicted combination			
		Q	**SD**	**Eff**	**TT**
Used parameter set	**Q**	X	-0.1549	-0.0757	-0.0643
	SD	-0.0162	X	-0.0687	-0.0310
	Eff	-0.0366	-0.7107	X	-1.3202
	TT	-0.0522	-0.1295	-0.0702	X
	Macro	-0.0070	-0.0007	-0.0784	-0.0326
	Meso	-0.0363	-0.0717	-0.0423	-0.0079
	All	-0.0084	-0.0015	-0.0623	-0.0201

Table H.33: Difference in the GoF, as determined by Equation 5.10, for the comparisons between different metrics using the **t-junction low density** scenario as the sole scenarios. The combinations are identified by their acronyms as listed in Table 5.6. Every column shows the decrease in GoF, compared to the optimal GoF, of that specific combination if the optimal parameter set of the combination, defined by the row, is used.

		Predicted combination			
		Q	**SD**	**Eff**	**TT**
Used parameter set	**Q**	X	-0.1200	-0.0882	-0.0399
	SD	-0.0012	X	-0.0010	-0.0579
	Eff	-0.0013	-0.0047	X	-0.0545
	TT	-0.0010	-0.0931	-0.0788	X
	Macro	-0.0012	0.0000	-0.0010	-0.0579
	Meso	-0.0014	-0.0080	-0.0038	-0.0485
	All	-0.0012	-0.0043	-0.0017	-0.0526

I Calibration - different objective functions

This appendix shows how the optimal parameter sets of all 16 combinations would change if instead of the squared error the absolute of the error would have been used in the objective functions. The three equations below show how the equations 5.6 to 5.8 would be if the absolute error is used instead of the squared error.

$$E_{norm}(\theta) = \left| \frac{M_{sim}(\theta) - M_{ref}}{M_{norm}} \right| \tag{I.1}$$

$$E_{norm;macro}(\theta) = \frac{1}{m} \sum_{j} \left| \frac{\frac{\sum_i M_{sim;i;j}(\theta)}{n} - M_{ref;j}}{M_{norm}} \right| \tag{I.2}$$

$$E_{norm;meso}(\theta) = \frac{1}{2} \left| \frac{M_{sim;\mu}(\theta) - M_{ref;\mu}}{M_{norm;\mu}} \right| + \frac{1}{2} \left| \frac{M_{sim;\sigma}(\theta) - M_{ref;\sigma}}{M_{norm;\sigma}} \right| \tag{I.3}$$

Table I.1: Comparison of the optimal parameter sets resulting from either using a squared objective function or an absolute error objective function. The red shading in the Abs columns indicates how large the difference is between the two parameter value (the darker the shading the larger the difference)

		Relaxation time [1/s]		Viewing angle [degree]		Radius [m]	
		MSE	*Abs*	*MSE*	*Abs*	*MSE*	*Abs*
1.	**B-H**	0.620	0.590	57.00	57.00	0.15296	0.15296
2.	**B-L**	0.620	0.620	57.00	66.00	0.19120	0.16252
3.	**B**	0.395	0.380	68.25	68.25	0.20076	0.19120
4.	**C-H**	0.395	0.515	57.00	57.00	0.23900	0.23900
5.	**C-L**	0.380	0.380	61.50	59.25	0.23900	0.20076
6.	**T-H**	0.590	0.575	57.00	59.25	0.21988	0.21988
7.	**T-L**	0.380	0.380	68.25	66.00	0.23900	0.23900
8.	**Q**	0.380	0.380	59.25	81.75	0.20076	0.18164
9.	**SD**	0.575	0.590	59.25	59.25	0.21988	0.21988
10.	**Eff**	0.500	0.410	57.00	57.00	0.23900	0.23900
11.	**TT**	0.620	0.605	59.25	57.00	0.15296	0.21032
12.	**H-D**	0.575	0.605	57.00	57.00	0.21032	0.21032
13.	**L-D**	0.500	0.410	57.00	57.00	0.21032	0.23900
14.	**Macro**	0.545	0.380	59.25	68.25	0.21988	0.19120
15.	**Meso**	0.620	0.515	59.25	57.00	0.15296	0.23900
16.	**All**	0.575	0.605	57.00	57.00	0.21032	0.21032

J | Decrease in GoF versus the distance from the optimal parameter set

The relationship between the decreases in the GoF and the distance between the accompanying parameter sets and the optimal parameter set gives insight into how the objective values are distributed in relation to the optimal parameter set and the lowest objective value.

The distance between a parameter set and the optimal parameter set is determined as follows:

$$D_{(\theta,\theta^*)} = \sqrt{N_\tau(\theta,\theta^*)^2 + N_\phi(\theta,\theta^*)^2 + N_r(\theta,\theta^*)^2} \tag{J.1}$$

Where $N_\tau(\theta,\theta^*)$, $N_\phi(\theta,\theta^*)$ and $N_r(\theta,\theta^*)$ are respectively the number of step-sizes between the optimal parameter set θ^* and the parameter set θ for the relaxation time, the viewing angle and the radius. So $N_\tau(\theta,\theta^*) = 2$ means that the relaxation time of parameter set θ is 2 step-sizes removed from the relaxation time of the optimal parameter set. The maximum distance is 24.74 ($\sqrt{16^2 + 16^2 + 10^2}$) which is the distance between one corner of the search space to the opposite corner of the search space.

The relationship is of interest because it can give insight into the question if the objective space contains local minima which have only a slightly higher objective value than the minimal/optimal objective value but are in a totally different area of the search space. This is of interest because, as subsection 5.4.3 shows, slight changes in, for example the search-grid, might cause slight changes in the objective space. The more local minima with slightly higher objective values than the optimal value the objective space contains which are in a totally different area of the search space, the more likely it is that small changes in the methodology can cause the optimal parameter set to change significantly.

The graphs in figures J.1 to J.3 show the relationships between the decrease in GoF and the distance to the optimal parameter set for all 15 combinations. The graphs only show the relationship for a limited range of decreases of GoF[1]. The limited range is chosen because the interest is in the points which only slightly differ in objective value from the optimal value.

The graphs in Figure J.1 show a clear difference between the low and the high density scenarios. The low density scenarios have many points in their objective-spaces which differ only slightly in objective value but are many step-size removed from the optimal parameter set. This is especially the case for the bidirectional low density scenario which has the points with the smallest decrease in GoF and the highest distance values (i.e. points near the top-right corner of the graph). The high density scenarios show a different pattern compared to the low density scenarios whereby for the same decrease in GoF the maximal distance is far smaller compared to the low density scenarios. Compared to each other it is clear that, similar to the low density scenarios, the bidirectional scenarios has the points with the lowest decrease in GoF and the highest distance values. In summary, the graphs show that small changes in the objective space are far more likely to affect the location of the optimal parameter set of the low density scenarios than the high density scenarios and are more likely to affect the bidirectional scenario than other scenarios of a similar density level.

The graphs in Figure J.2 show clear differences between the four metrics. Especially the flow differs from the other three given that it has many points with small decrease in the GoF but a high distance value whilst none of the other three metrics has any points with small decreases in the GoF and a high distance value. This shows that the model can obtain good estimates of the flow using many different parameter sets and hence it also shows the flow is probably not a good metric to be used on its own given that it cannot differentiate well between the different parameter sets and the differences in the underlying behaviour.

The graphs in the first column Figure J.3 show the clear difference between the combinations of low and high density scenarios. As was concluded earlier, theses graphs show that small changes in the objective space are far more likely to affect the location of the optimal parameter set of the low

[1] Figures J.4 to J.6 show the graphs over the whole range of decreases in the GoF and also show which areas are displayed in the graphs of figures J.1 to J.3.

density scenarios than the high density scenarios. The second column shows the differences between the different combinations of metrics of the same aggregation level. The graph of the macroscopic metric clearly shows the dominance of the spatial distribution over the flow given that there are no longer any points with very small decreases in the GoF but high distance values as was the case for the flow. The graph of the mesoscopic metrics shows that combining them affects the relationship negatively in the sense that there are points with a smaller decrease in the GoF and a higher distance value than can be found in the graphs of either of the individual metrics.

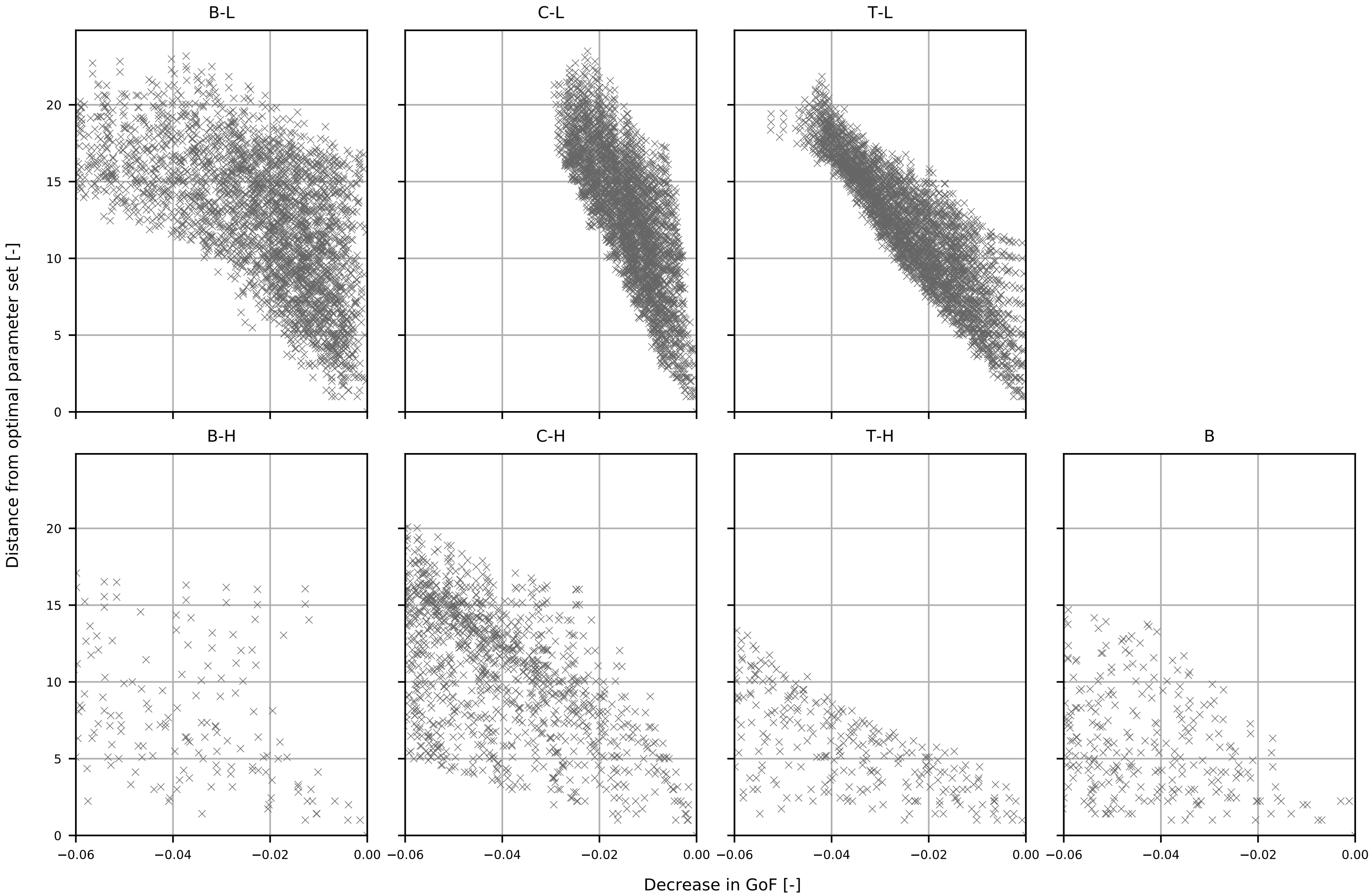


Figure J.1: The decrease in the GoF versus the distance from the optimal parameter set for the individual scenarios - all metrics (combinations 1-7).

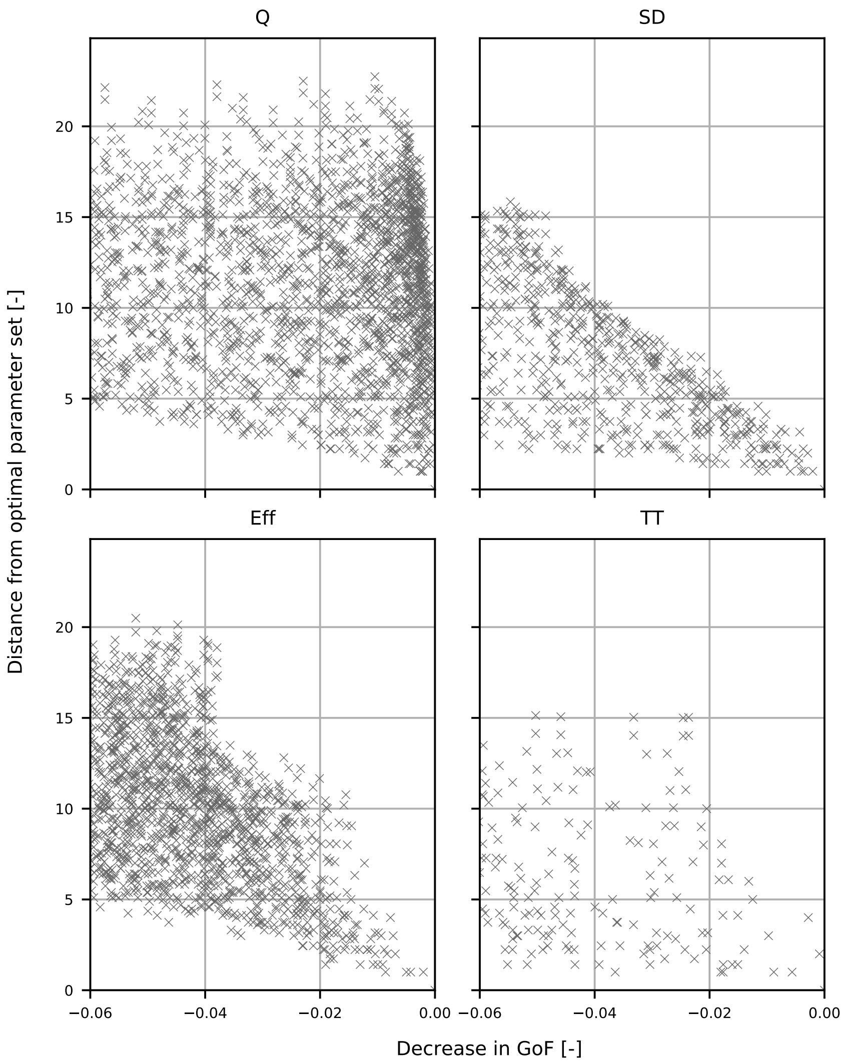


Figure J.2: The decrease in the GoF versus the distance from the optimal parameter set for the individual metrics - all scenarios (combinations 8-11)

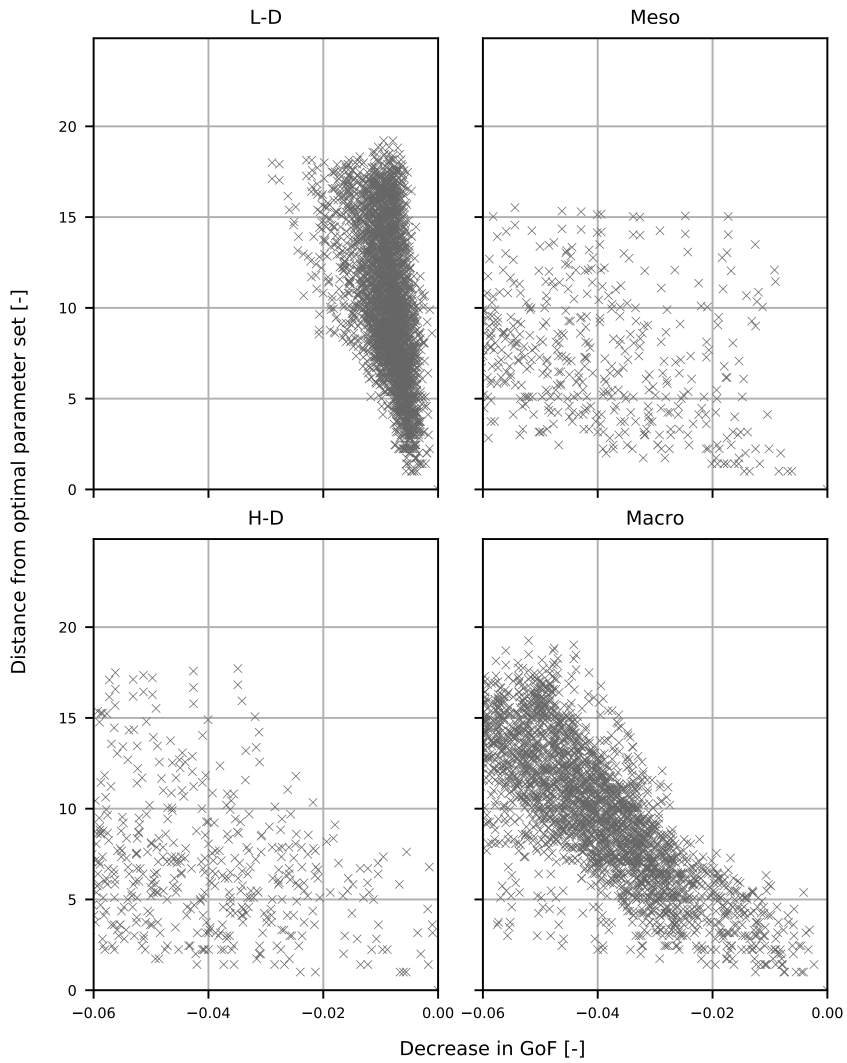


Figure J.3: The decrease in the GoF versus the distance from the optimal parameter set for the different combinations of scenarios and metrics (combinations 12-15)

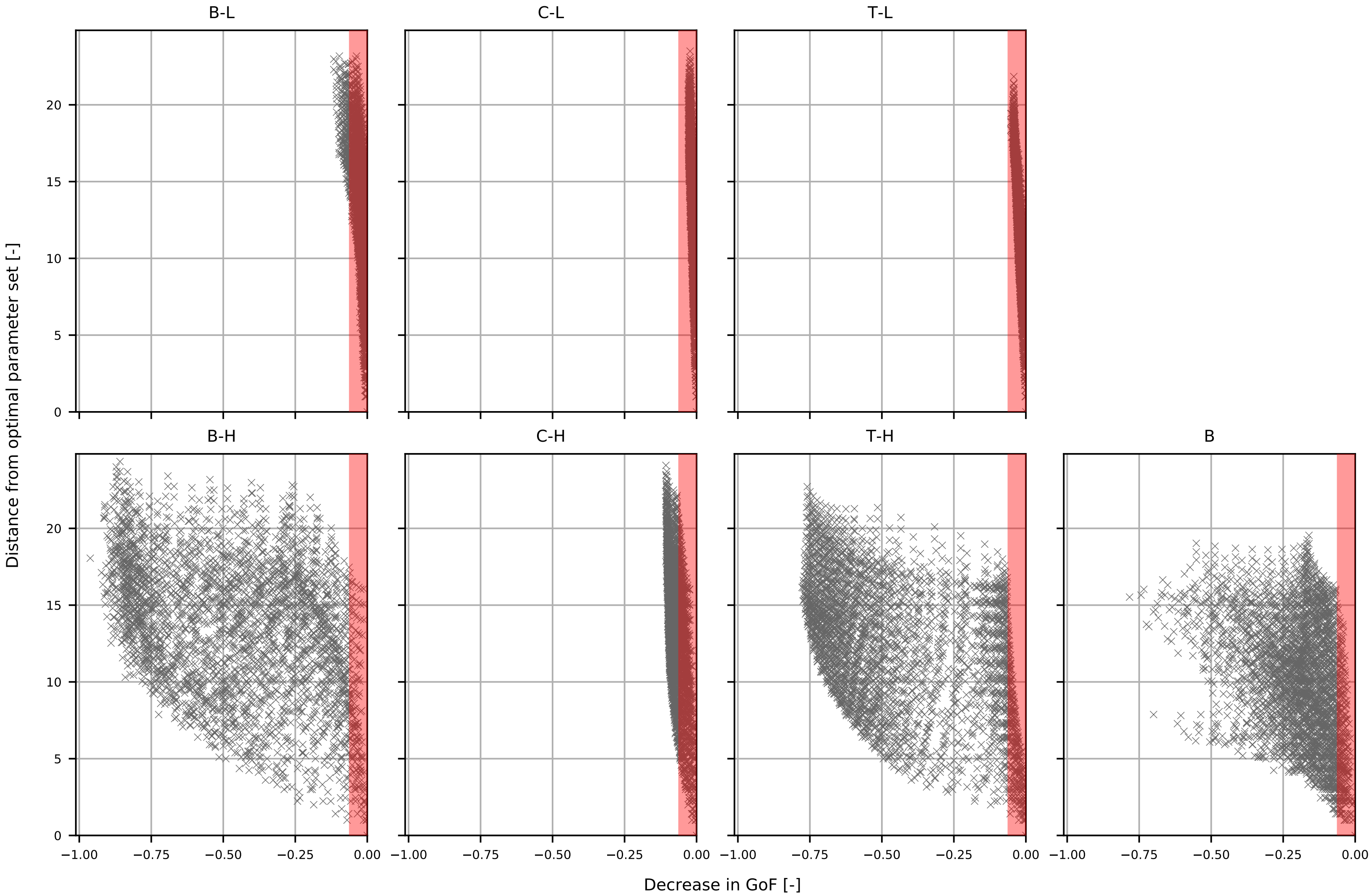


Figure J.4: The decrease in the GoF versus the distance from the optimal parameter set for the whole range of decreases in GoF for the individual scenarios - all metrics (combinations 1-7). The shaded area corresponds with the zoomed-in area displayed in Figure J.1

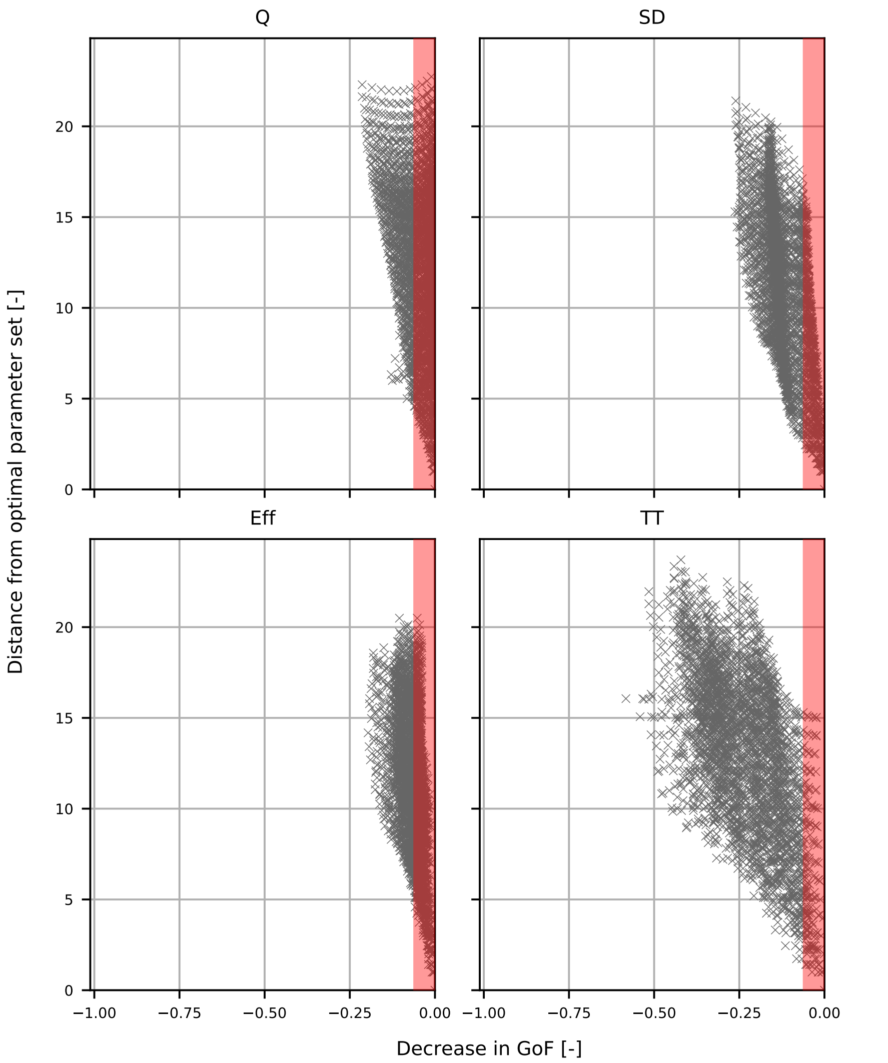


Figure J.5: The decrease in the GoF versus the distance from the optimal parameter set for the whole range of decreases in GoF for the individual metrics - all scenarios (combinations 8-11). The shaded area corresponds with the zoomed-in area displayed in Figure J.2

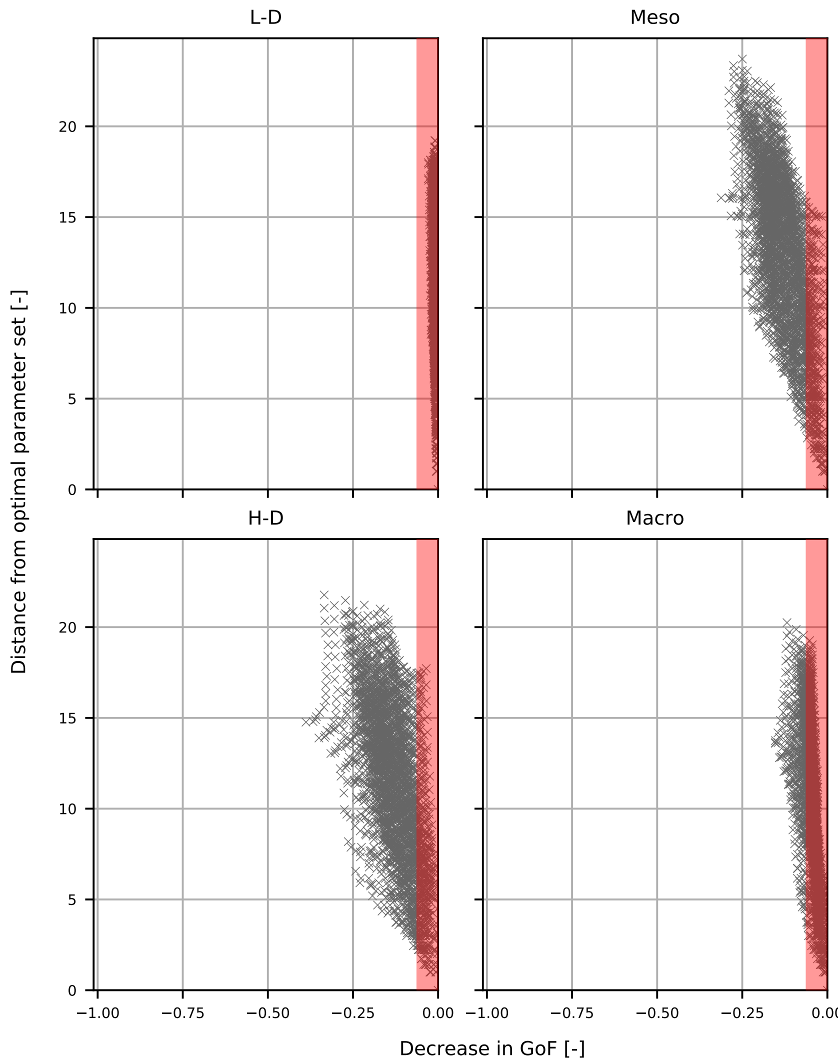


Figure J.6: The decrease in the GoF versus the distance from the optimal parameter set for the whole range of decreases in GoF for the combinations of scenarios and metrics (combinations 12-15). The shaded area corresponds with the zoomed-in area displayed in Figure J.3