

Research Article

Multiobjective Calibration Framework for Pedestrian Simulation Models: A study on the Effect of Movement Base Cases, Metrics, and Density Levels

Martijn Sparnaaij , Dorine C. Duives , Victor L. Knoop , and Serge P. Hoogendoorn 

Department of Transport & Planning, Delft University of Technology, Stevinweg 1, 2628 CN Delft, Netherlands

Correspondence should be addressed to Martijn Sparnaaij; m.sparnaaij@tudelft.nl

Received 31 January 2019; Revised 15 April 2019; Accepted 28 May 2019; Published 3 July 2019

Guest Editor: Milad Haghani

Copyright © 2019 Martijn Sparnaaij et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ideally, a multitude of steps has to be taken before a commercial implementation of a pedestrian model is used in practice. Calibration, the main goal of which is to increase the accuracy of the predictions by determining the set of values for the model parameters that allows for the best replication of reality, has an important role in this process. Yet, up to recently, calibration has received relatively little attention within the field of pedestrian modelling. Most studies focus only on one specific movement base case and/or use a single metric. It is questionable how generally applicable a pedestrian simulation model is that has been calibrated using a limited set of movement base cases and one metric. The objective of this research is twofold, namely, to (1) determine the effect of the choice of movement base cases, metrics, and density levels on the calibration results and (2) to develop a multiple-objective calibration approach to determine the aforementioned effects. In this paper a multiple-objective calibration scheme is presented for pedestrian simulation models, in which multiple normalized metrics (i.e., flow, spatial distribution, effort, and travel time) are combined by means of weighted sum method that accounts for the stochastic nature of the model. Based on the analysis of the calibration results, it can be concluded that (1) it is necessary to use multiple movement base cases when calibrating a model to capture all relevant behaviours, (2) the level of density influences the calibration results, and (3) the choice of metric or combinations of metrics influence the results severely.

1. Introduction

The creation and implementation of a commercial pedestrian simulation will, ideally, consist of multiple steps. One of those steps is calibrating the model whereby the goal is to increase the accuracy of the model predictions by obtaining the parameter set that results in the best replication of reality. As such, calibration is an important step.

Yet, up to recently, calibration has received relatively little attention within the field of pedestrian modelling [1, 2]; this is mainly attributed to the lack of data [1, 3–5] especially at high densities. Despite this issue, there are many studies in which authors calibrate a pedestrian model (e.g., [6–10]) usually by using a fundamental diagram [11] or trajectories. However, as multiple authors mention, the calibration attempts in these studies are limited and mostly focus on only one or a few

aspects [1, 4, 5, 11, 12]. Most studies focus on one specific movement base case (e.g., a bidirectional flow in a straight corridor), use only one metric, or do not look at various population compositions.

It is questionable how generally applicable a pedestrian simulation model is that has been calibrated using a limited set of movement base cases. Campanella, Hoogendoorn, and Daamen [13] and Duives [14] show that using different flow situations leads to different optimal parameter values. That is, both studies identify that for general usage (i.e., using a single model for many different applications) one needs to calibrate a pedestrian simulation model using multiple movement bases to capture all relevant behaviours. The effect of using different metrics during the calibration has been investigated by Duives [14] in relation to pedestrian dynamics and among others [15–17] in relation to vehicular traffic. These studies

illustrate that different combinations of metrics clearly lead to different calibration results. Wolinski et al. [18] also calibrate a number of models using different metrics. Though they do not show the effect of using different metrics on the resulting optimal parameter set, the results show clearly that the model fit to the data depends on the metric used. Furthermore, Hänseler, Bierlaire, Farooq, and Mühlematter [19] also note differences between the optimal parameter sets obtained using different metrics or a combination of them when calibrating their macroscopic model.

To overcome the problem of obtaining different results when using different movement base cases and/or metrics, three multiple-objective calibration frameworks have been proposed in recent years which try to take a more inclusive approach. Wolinski et al. [18] propose a framework which can potentially incorporate multiple objectives. However, during their benchmarking tests they only apply combinations of one movement base case and one metric. Campanella et al. [13] show how a microscopic model can be calibrated using multiple movement base cases and how this compares to calibrating the model with only one movement base case. This study still only uses one metric during the calibration. The work by Duives [14] uses multiple movement base cases with multiple metrics and furthermore includes different combinations of weights in the objective function and is thus the most extensive of the three. However, except for the spatial distribution metric, the study does not show the results for the other individual metrics. Hence, it is not possible to exactly determine the effect the different individual metrics have on the resulting optimal parameter set.

So, even though these works illustrate that the choice of the combinations of movement bases and metrics will influence the optimal parameter set obtained during calibration, all three studies have their limitations. For example, they do not explicitly examine the effect of the level of density which might be a relevant factor given that work by Campanella et al. [13] shows that poorness of data can affect calibration results and work by [20] also shows some differences in the obtained parameter sets for a low and high density case of a bidirectional flow. Furthermore, the effects of using different metrics are also still poorly understood. And, as work by Campanella, Hoogendoorn, and Daamen [21] shows that using multiple objectives during the calibration will lead to a better validation score for general usage, it is clear that increased insights into which objectives to use during calibration can improve model validity.

Given these observations the objective of this research is twofold. Firstly, the objective is to determine the effect of the choice of movement base cases, metrics, and density levels on the calibration results. Secondly, the objective is to develop a multiple-objective calibration method for pedestrian simulation models to determine the aforementioned effects, taking into account the stochastic nature common to many microscopic pedestrian models.

This study aims to add value to the current body of literature by means of a more extensive study of the impact of calibration framework setup on the validity of a pedestrian simulation model. This extension provides, among other things, novel insights into the effect of the level of density of

the movement base case and more detailed insights into the effect of using a range of metrics in the calibration process. Furthermore, this study features a different type of model (i.e., vision-based model [22]) than the previous most extensive studies (i.e., [13, 14]), which both calibrated NOMAD. Thus, this study also illustrates the replicability of their results and the conclusions of those previous studies.

The rest of the paper is organized as follows. Section 2 briefly describes the microscopic pedestrian simulation model. Section 3 shortly introduces the methodology of the sensitivity analysis and presents the results of the analysis. In Section 4 the calibration methodology is elaborated upon. This is followed by the presentation of the results of calibrating the model using a single objective in Section 5. Section 6 presents the results of the multiobjective calibration. Finally, this paper closes off with a discussion of the results, conclusions, and the implications of this work for practice.

2. Brief Introduction to Pedestrian Dynamics

This section introduces pedestrian dynamics (PD), a microscopic pedestrian simulation model developed by INCONTROL Simulation Solutions. It offers a user the ability to model the movement behaviour of pedestrians at all three behavioural levels (strategic, tactical, and operational). Though, in this research pedestrians only have one activity, namely, to walk from their origin to their destination via a single route, and hence there is no need to model the activity choice, the activity scheduling or the route choice. The model featuring the operational walking dynamics is discussed in more detail underneath.

The operational behaviour of the INCONTROL model consists of two parts, i.e., route following and collision avoidance, which together determine the acceleration of a pedestrian at every time step. PD determines the acceleration of a pedestrian by the combination of 'social forces' and a desired velocity component. The pedestrian itself is represented by a circle with a radius r . At every time step the acceleration of a pedestrian is determined as follows:

$$\frac{d\vec{v}_i}{dt} = \frac{\vec{v}_{des,i} - \vec{v}_i}{\tau} + \sum_j \vec{f}_{i,j} + \sum_W \vec{f}_{i,W} \quad [\text{m/s}^2] \quad (1)$$

where $\vec{v}_{des,i}$ and \vec{v}_i are, respectively, the desired and current velocity of pedestrian i . $\vec{f}_{i,j}$ and $\vec{f}_{i,W}$ are the physical forces that occur on contact with another pedestrian or a static obstacle. And lastly, τ is the relaxation time. Furthermore, in case the speed of the pedestrian drops below a certain threshold (i.e., the minimal desired speed parameter) the pedestrian does not move until the next time step when the resulting speed is higher than the threshold.

The desired velocity is determined according to the method proposed by Moussaïd et al. [22]. The method uses a vision-based approach to avoid collisions. This approach combines the collision avoidance with the preferred speed and the desired destination to determine the desired velocity. The desired velocity is determined by means of two heuristics, namely,

- (1) A pedestrian chooses the direction that results in the most direct path to its desired destination given the presence of both static and dynamic obstacles.
- (2) A pedestrian chooses the speed that, in case there is an obstacle in the preferred direction, results in the lowest time-to-collision whereby this time is always larger than τ .

The pedestrian only takes into account obstacles that are within its field of vision ($[-\phi, \phi]$), which is determined by the current orientation of the pedestrian, the viewing angle ϕ , and the viewing distance d_{\max} . The desired direction is determined by minimizing (2).

$$d_h = \begin{cases} d_{i;\text{exp:col}} - d_{i;\text{pers}} & \text{If the expected collision is with another} \\ & \text{pedestrian travelling in the same direction [m]} \\ d_{i;\text{exp:col}} & \text{Otherwise} \end{cases} \quad (3)$$

where $d_{i;\text{exp:col}}$ is the distance to the first expected collision in the desired direction α of pedestrian i and $d_{i;\text{pers}}$ the personal distance of pedestrian i . The personal distance is the distance a pedestrian wants to keep between itself and another pedestrian.

There are two important notes regarding the implementation of this method in PD, namely: (1) Regardless of the settings for the viewing angle and viewing distance, the model will only take into account the four closest pedestrians (who are within the field of vision) when determining the desired velocity. (2) Not all parameters can be adapted by the user; the parameters governing the physical forces (i.e., $\vec{f}_{i;j}$ and $\vec{f}_{i;W}$) are, namely, not user-adaptable.

The desired destination of each pedestrian is determined using the Indicative Route Method proposed by Karamouzas, Geraerts, and Overmars [23]. This desired destination is influenced by two (user-adaptable) parameters which are the “Preferred clearance”, which influences the minimal distance a pedestrian wants to keep between its desired destination and a static obstacle, and the “Side preference update factor”, which influences the strength of the desired destination location changes given the current position of the pedestrian and the current deviation from the originally planned path.

As is the case for many pedestrian models, PD is stochastic by nature (i.e., two simulations with exactly the same parameters and input but with different seeds result in different outcomes). In this study there are three main causes for this stochasticity, namely, the preferred speed, the initial destination point, and the exact point of origin. The first contributes to the stochasticity due to the fact that every pedestrian is randomly assigned a preferred speed from a given distribution. The latter two causes of stochasticity are points whose location influences the desired destination and whose exact position is a randomly determined location within a respective origin or destination area. The fact that the model is stochastic by nature has to be taken into account

$$d(\alpha) = d_{\max}^2 + f(\alpha)^2 - 2d_{\max}f(\alpha)\cos(\alpha_0 - \alpha), \quad (2)$$

for $\alpha \in [-\phi, \phi]$

where $f(\alpha)$ is the distance to the closest expected obstacle in the direction of α , which is equal to d_{\max} if there are no obstacles within the viewing range. α_0 is the angle towards the desired destination. 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 pedestrian and d_h is the distance between the pedestrian and the first expected collision in the desired direction. d_h is determined as follows:

during the calibration and is discussed in more detail in the next section.

3. Sensitivity Analysis

A sensitivity analysis is performed to determine to which particular parameters the model is sensitive. This section describes the methodology of the sensitivity analysis and presents the results of this analysis. The results of the sensitivity analysis are used to determine which parameters should be incorporated in the calibration process, as recalibrating all model parameters is not feasible within the time frame of this study. How the results are used to determine the calibration search space and why it is not feasible to include all parameters is explained in more detail in Section 4.5, Search Space Definition.

3.1. Methodology of the Sensitivity Analysis. The goal of the sensitivity analysis is to determine which of the 7 model parameters of the INCONTROL model (see Table 1) that influence the operational behaviour most affect the model's results. The authors expect that the sensitivity depends on the scenario. Thus, the analysis is performed for all seven scenarios used in the calibration. These scenarios are as follows: A high density bidirectional flow, a high density corner flow, a high density t-junction flow, a bottleneck flow and low density variants of the bidirectional, corner, and t-junction flows. For a detailed description of the scenarios, see Section 4.1.

The distribution of the instantaneous speeds of all pedestrians is used as the sole metric. This distribution contains all instantaneous speeds of all pedestrians and all replications. This metric is chosen because 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

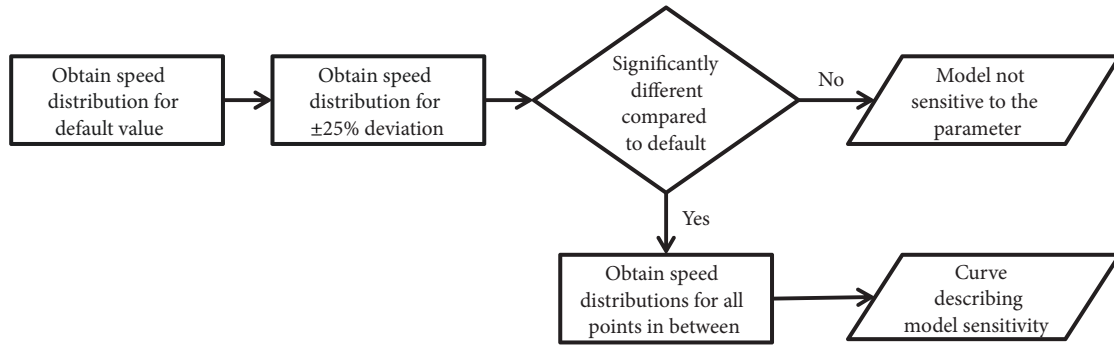


FIGURE 1: Overview of the sensitivity analysis methodology. The process depicted in the figure is performed for all combinations of the 7 scenarios and 7 parameters.

TABLE 1: The seven parameters included in the sensitivity analysis.

Route following:	Preferred clearance
	Side pref. update factor
Collision avoidance:	Relaxation time
	Viewing angle
	Viewing distance
	Min. desired speed
	Personal distance

that interactions are not solved efficiently causing pedestrians to change their speed a lot).

The Anderson-Darling test [24] is used to determine if enough replications have been performed to allow for the comparison of two distributions of the instantaneous speed. This statistical test determines whether a distribution has converged (for more details see Section 4.6).

To limit the amount of simulations only first-order effects are investigated. Note that the number of replications is already extensive due to the incorporation of seven scenarios and a vast number of replications to account for model stochasticity. Figure 1 depicts the process which is applied to every combination of a scenario and a parameter. The remaining text in this subsection describes the process and its steps in more detail. For every scenario, the first step is to obtain the speed distribution using the defaults values. This distribution serves as the base line. Accordingly, for all combinations of the 7 parameters and 7 scenarios, the value of the parameter in question is increased by 25%. After which the distribution of the speeds is again obtained. The same procedure is followed for a decrease of the parameter values by 25%. The limit of 25% is chosen as the INCONTROL model already has undergone some basic calibration and hence it is assumed that the optimal values will not deviate much from the current default values.

For all these new distributions of speeds accordingly the following two checks are identified: (a) is the new distribution significantly different from the default distribution according to the Anderson-Darling test and (b) are the differences between the means and standard deviations of the distributions larger than one would expect based on the influence

of the stochasticity (for more details see [25]). The second check provides insight into the magnitude of the difference and, thus, the degree of the sensitivity of the model to changes in this parameter. If both these conditions hold (i.e., significant differences are found between the new and old distributions) all distributions between the 25% boundary and the default value, using a precision of 1% point, are obtained to gain insight into how the model's sensitivity changes as the deviation from the default value increases.

For a more detailed description of the methodology the reader is referred to [25].

3.2. Results of the Sensitivity Analysis. Based on the methodology described above the following results are obtained. Firstly, the model is not sensitive to changes in the “Preferred clearance” parameter and changes in the “Viewing distance” parameter. Even in high density cases the model is not sensitive to changes in the viewing distance parameter. This is most likely the result of the fact that PD only takes into account the four closest pedestrians within the viewing field, which might all reside well within this radius.

In general, the model is not sensitive to changes in the “Personal distance”, the “Side pref. update factor”, and the “Min. desired speed”. In the case of the “Personal distance”, the model is slightly sensitive to changes in this parameter in the case of the bottleneck scenario. In the case of the “Side pref. update factor”, the model is slightly sensitive to the t-junction high density scenario. The model is also slightly sensitive to changes in the “Personal distance” for both scenarios. However, as the maximum differences between the means and the standard deviations are at most 2%, we conclude that within the $\pm 25\%$ range changes in these parameters will not affect the model results much.

The only parameter to which the model is sensitive in all seven scenarios is the relaxation time. The model is, furthermore, sensitive to the viewing angle in all four high density scenarios. Figure 2 presents that, for both parameters and the seven scenarios, the differences between the mean and the standard deviation of the distributions are obtained using an increased or decreased parameter value. From this figure a number of observations can be made. Firstly, the figure clearly shows that the model is more sensitive to changes in the relaxation time than to the changes in

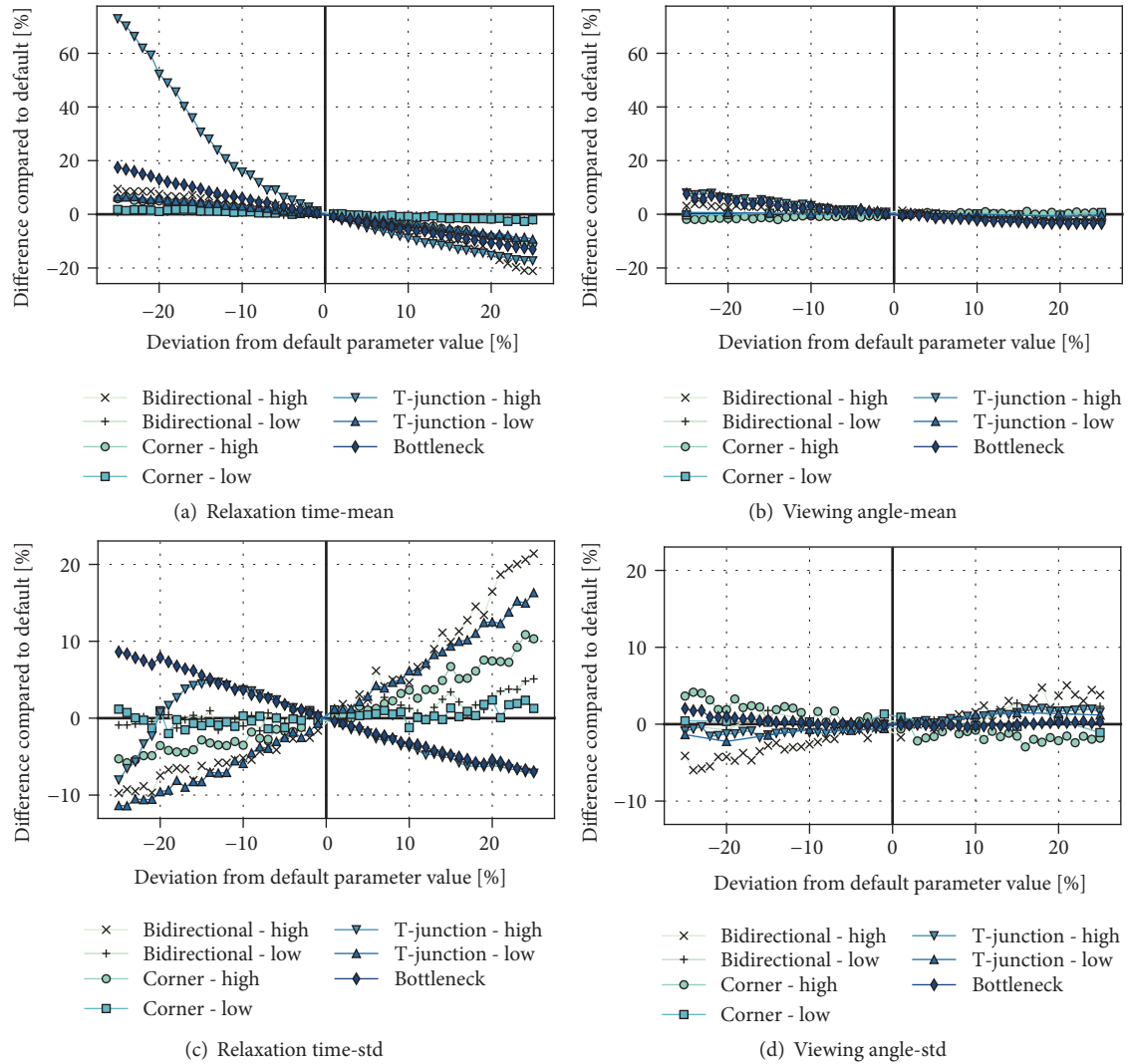


FIGURE 2: Results of the sensitivity analysis.

the viewing angle. Secondly, clear differences between the scenarios can be identified. For example, the model is more sensitive to changes in the parameters when simulating high density scenarios. And thirdly, in many cases, an asymmetry can be observed between the slope and shape of the curve left from the default value (i.e., the decreased parameter values) and the one on the right. Overall, we conclude that the effect of changing a parameter's value differs a lot between the two parameters and the seven scenarios.

In conclusion, overall the INCONTROL model is not very sensitive to changes in many of the parameters given the $\pm 25\%$ boundaries. The model is primarily sensitive to changes in the relaxation time, whose value influences the outcome of the simulation in all seven scenarios, and the viewing angle, in the case of the high density scenarios.

4. Methodology

This section presents the reasoning behind the newly developed calibration methodology. Figure 3 depicts the multiple-objective calibration methodology and its parts. In line with

previous research, the methodology uses multiple scenarios and multiple metrics. For every combination of a scenario and a metric an objective value is obtained which represents the difference between the simulation and the reference data for the given parameter set. These objective values are then combined into a single objective value which in turn is used by the optimization method to determine if the current parameter set is optimal. In case the current parameter set is deemed optimal, the calibration stops and the next step would be to validate the model (which is not part of this study). Alternatively, in case the parameter set is not deemed optimal a new parameter set is created and the calibration process continues.

All of these parts are discussed in more detail in this section. First the scenarios are identified. Accordingly, the metrics and the objective function are presented. This section furthermore elaborates on the optimization method and the manner that the stochasticities of the pedestrian simulation model are handled.

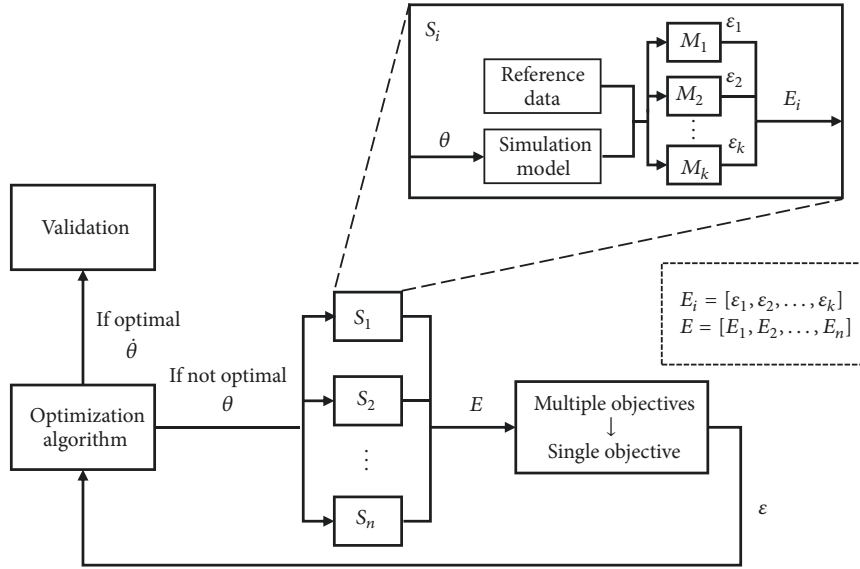


FIGURE 3: Overview of the multiple-objective methodology where S_1 to S_n are the used scenarios, M_1 to M_k the used metrics, ϵ_1 to ϵ_k the individual errors, and E_i the combined error for scenario i , ϵ the combined error and θ the parameter set.

4.1. Scenarios. Contemporary, several datasets are available that feature the movement of pedestrians in multiple movement base cases and a similar population of pedestrians, among others [26–28]. Since the experiments within the HERMES project represent the most comprehensive set of movement base cases featuring a similar population and different levels of density, this dataset will be used in this calibration procedure. Based on this dataset seven scenarios are constructed whereby every scenario contains a single movement base case and a single density level. Four movement base cases are studied; these are a bidirectional flow, a unidirectional corner flow, a merging flow at a t-junction, and a bottleneck flow. All base cases have both a low and high density variant except for the bottleneck which only has a high density variant. Figure 4 shows the layout of the four simulated movement base cases. For a more detailed overview of the experimental setup within the HERMES project the reader is referred to [27]. Care is taken to ensure a similar flow pattern over time, speed distribution, and route choice; details on the exact simulation setup of the seven scenarios in PD are mentioned in [25].

4.2. Metrics. In this multiple-objective framework four different metrics are used to identify how different metrics impact the calibration results. 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 travel time distribution and the effort distribution. These metrics are chosen because, on both levels, they describe different aspects of the walking behaviour. Microscopic metrics, i.e., trajectories, are not used for three reasons. Firstly, calibration based on trajectories requires a different approach than calibrating based on macroscopic and mesoscopic metrics. It would require many more simulations to use both microscopic and mesoscopic and/or macroscopic metrics and due

to time limits this was not considered to be viable within this study. Secondly, the current approaches for calibrating based on trajectories do not deal with the stochastic nature of the model. Lastly, since pedestrian simulation models are mostly used to approximate the macroscopic properties of the infrastructure (e.g., capacity, density distribution) [21] and given that calibrating based on microscopic metrics does not necessarily result in a macroscopically valid model [11], macroscopic and mesoscopic metrics take priority over microscopic metrics. The four metrics adopted in this research to calibrate PD are discussed in more detail below.

Flow. The flow is chosen as a macroscopic metric to check to what extent the model can reproduce the throughput in different situations. In all seven scenarios the average flow is measured along a certain cross-section (see Figure 4) during a certain measurement period. The average flow is calculated as follows:

$$\bar{q}_i = \frac{N_i}{\Delta t \times l} \quad [\text{ped/s/m}] \quad (4)$$

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 measurement period (Δt). The flow is normalized to a flow per meter of measurement line, whereby l is the length of the measurement line, in order to allow for comparisons between scenarios.

Distribution over Space. Reference [14] showed that microscopic models might not always be able to accurately reproduce the spatial distribution patterns. Hence, it is essential to check whether a model performs well with respect to this property. The distribution over space measures how the pedestrians are distributed over the measurement area. A grid of 0.4×0.4 m, which is approximately the size of one pedestrian during a high density situation, overlays the

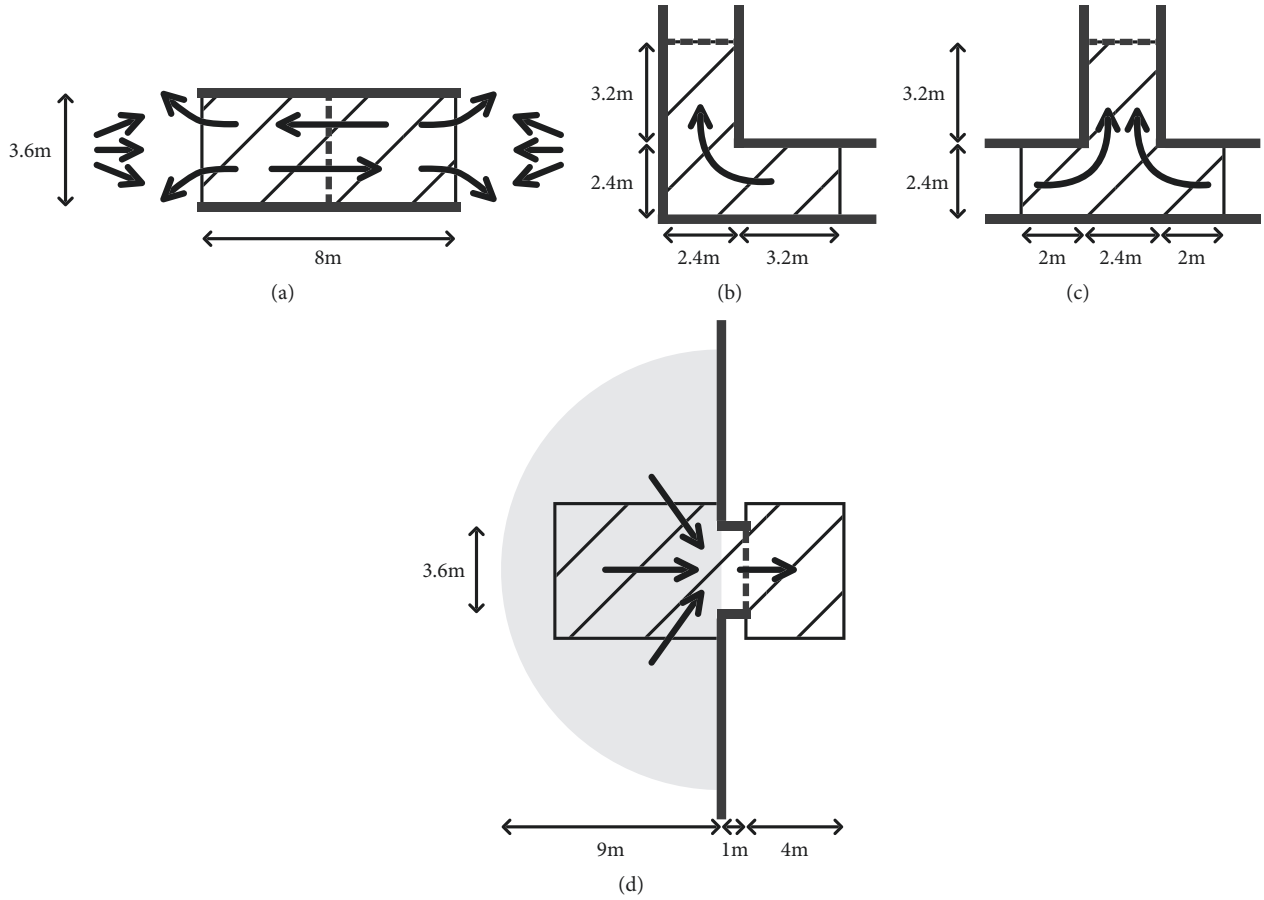


FIGURE 4: Overview of the layout of the four movement base cases used during the calibration. (a) Bidirectional flow. (b) Unidirectional corner flow. (c) Merging t-junction flow. (d) Bottleneck. In the figure the hatched areas indicate the measurement areas; the dashed lines the location where the flow is measured and in the case of the bottleneck the grey area indicates the waiting area at the start of the simulation.

measurement area and for every cell the percentage of the time it is occupied is determined by (5).

$$F_j = \frac{N_{occ;j}}{N_{steps}} \quad [-] \quad (5)$$

where $N_{occ;j}$ is the number of time steps cell j 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.

Travel Time. The travel time is the time it takes a pedestrian to traverse the measurement area as determined by (5).

$$TT_k = \frac{t_{end} - t_{start}}{l_{ref}} \quad [s/m] \quad (6)$$

where t_{start} and t_{end} are, respectively, the time pedestrian k first which entered the measurement area and time pedestrian k which left the area. l_{ref} is the average length of the path in the measurement area, as obtained from the reference data. The travel time is normalized in order to simplify the comparison of the goodness-of-fit of different scenarios with different average path lengths. Note that this metric approximates the realized pace of each individual. That is, if

an individual makes a detour at a very high speed it will not affect its travel time, but it will influence the effort required to get to its destination.

Only the travel times of those pedestrians who successfully traversed the whole measurement area during the measurement period are included in the distribution of the travel times.

Effort. Several studies have identified the difficulty of smooth interactions between simulated pedestrians in bidirectional flows. In order to ensure realistic interaction behaviour the effort metric is introduced, which captures how much effort it takes a pedestrian to traverse the measurement area. The effort for pedestrian k is defined as the average change in velocity per time step (see (7)).

$$e_k = \frac{\sum_{n=1}^{n-1} (|v_x(t) - v_x(t-1)| + |v_y(t) - v_y(t-1)|)}{n-1}, \quad [m/s] \quad (7)$$

$$\begin{aligned} v_x(t) &= \frac{x(t) - x(t-1)}{\Delta t}, \quad [m/s] \\ v_y(t) &= \frac{y(t) - y(t-1)}{\Delta t}, \quad [m/s] \end{aligned} \quad (8)$$

TABLE 2: Normalization values.

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

where $v_x(t)$ and $v_y(t)$ are, respectively, the speed in the x and y-direction at time step t and n the number of time steps. The speeds are obtained by differentiating the current and previous positions of pedestrian k (see (8)), where $x(t)$ and $y(t)$ are, respectively, the x and y-position at time step t and Δt is the duration of the time step. The effort measurements of all pedestrians are combined into one distribution.

4.3. Objectives. In this research multiple objectives are combined into a single objective using the weighted sum method [29]. This is in line with research by Duives [14]; the only example in literature using both multiple metrics and scenarios to calibrate a pedestrian model.

In order to make a fair comparison between objectives, normalization is necessary, as the metrics have different units and different orders of magnitude. Table 2 shows the normalization values used during calibration. The adopted normalization method is based on two main assumptions. Firstly, it is assumed that for a single metric a deviation of 1 unit (for example, 1 ped/m/s is case of the flow) in one scenario is equally wrong as a deviation of 1 unit in another scenario (i.e., an absolute error is used instead of a relative error). Secondly, it is assumed that, for every metric, a deviation of 1 ped/m/s in the flow is equal to a deviation equal to the average ratio between the values of flow and the respective metric in the reference data (i.e., $(1/S_n) \sum_{s \in S} (y_{m;s}/\bar{q}_s)$, where S is the set of scenarios, S_n is the number of scenarios, \bar{q}_s is the average flow of scenario s , and $y_{m;s}$ is the value for metric m for scenario s). This method is chosen because it does not explicitly assume a bias towards any of the metrics or scenarios which is deemed appropriate for this study given its goal. However, this might not necessarily be the case if one intends to calibrate a model for a specific intended use. For a more detailed explanation of this method and an underpinning of the choice to specifically use this method, the reader is referred to [25].

The objective function for a given metric and scenario is given by the normalized Squared Error (SE) which for the macroscopic metrics is determined by (9) and for the mesoscopic metrics is determined by (10).

$$SE_{\text{norm}}(\theta) = \frac{1}{m} \sum_j \left(\frac{\sum_i M_{\text{sim};j}(\theta)/n - M_{\text{ref};j}}{M_{\text{norm}}} \right)^2 \quad (9)$$

$$SE_{\text{norm;meso}}(\theta) = \frac{1}{2} \left(\frac{M_{\text{sim};\mu}(\theta) - M_{\text{ref};\mu}}{M_{\text{norm};\mu}} \right)^2 + \frac{1}{2} \left(\frac{M_{\text{sim};\sigma}(\theta) - M_{\text{ref};\sigma}}{M_{\text{norm};\sigma}} \right)^2 \quad (10)$$

where M_{sim} is the metric's value according to the simulation, M_{ref} the reference value according to the data, M_{norm} the value used for the normalization, and θ the vector of model parameters. In the case of (9) n is the number of replications and m is the number of travel directions in case of the flow and the number of cells in case of the spatial distribution. In the case of the mesoscopic metrics (10) shows that the difference between the distributions is approximated by taking both the error in the mean (μ) and the standard deviation (σ). These distributions contain the measurements of all replications.

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

$$O(\theta) = \frac{1}{N_s * N_m} \sum_s \sum_m SE_{\text{norm};s;m}(\theta) \quad (11)$$

where $SE_{\text{norm};s;m}(\theta)$ is the value of the objective function of scenario 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. A likelihood method, which multiplies probabilities, might not work in this case, as this method will always attempt to fix the worst parameter first. In an additive scheme weights can be applied to prioritize certain combinations of scenarios and metrics over other combinations. However, as this research studies the effects of the different choices of scenarios and metrics, in this research all combinations are considered to be of equal importance and hence equal weights are used.

4.4. Optimization Method. In this research a grid search is used to obtain the optimal parameter set, as it provides the researcher with more insight into the shape of the surface of objective space. The disadvantage of using a grid search is that other optimization methods, e.g., Greedy and Genetic algorithms, can potentially be faster. However, these methods do run the risk of getting stuck in a local minimum and do not necessarily give a good insight into the shape of surface of the objective space. As can be derived from (11), smaller values of the objective function represent a better goodness-of-fit (GoF) and hence the goal of the optimization is to minimize the objective value.

4.5. Search Space Definition. Rudimentary calibration of PD has already been performed by the company. Thus, instead of calibrating all model parameters, the presented calibration method will be used in this research to identify the correctness of the variables with respect to which this model is most sensitive, namely the relaxation time and the viewing angle. Even though the model is less sensitive with respect to the radius, this parameter will also be included in the calibration as initial tests of the implementation of the scenarios illustrated that in the case of the bidirectional high density scenario the default radius of this model produced problematic results. The search space is defined as follows:

- (i) The upper and lower boundaries of the relaxation time and viewing angle are determined by a deviation of $-0.24 * \theta_{\text{def}} < \theta < 0.24 * \theta_{\text{def}}$ with respect to the

default parameters. The step size is 3% of the default value. The resulting ranges are $[0.380\text{s}^{-1}-0.620\text{s}^{-1}]$ for the relaxation time and $[57.00^\circ-92.00^\circ]$ for the viewing angle.

- (ii) For the radius the upper boundary is equal to the default value, the lower bound has a deviation of $-0.40 * \theta_{\text{def}}$ and the step size is 4% of the default value. This results in a range of $[0.14340\text{m}-0.23900\text{m}]$.

As this research focuses on the effect of density levels, the metrics that are part of the objective function and movement base cases, the search space is not continuous and has been restricted in order to create reasonable computation times and a reasonably good insight into the shape of the objective function.

4.6. Dealing with Stochasticity in Pedestrian Simulation Models. Similar to most pedestrian simulation models, PD is stochastic in nature. Therefore, it is essential to determine the minimum amount of replications one would need in order to assure that statistical differences are due to differences in model parameters instead of stochasticity in the model realization.

In this research the required number of replications is determined using a convergence method similar to [30] whereby the distribution of speeds is used as the sole metric. To determine if two subsequent distributions can be considered to be samples drawn from the same distribution the Anderson-Darling test is used [24]. Equation (12) shows that if b subsequent distributions are considered to be similar according to the Anderson-Darling test (i.e., the test return a p-value greater than $p_{\text{threshold}}$ indicating that the null hypothesis that the two samples come from the same distribution can be rejected at the $p_{\text{threshold}}$ significance level) the distribution has converged.

$$AD(S_n, S_{n-1}) > p_{\text{threshold}} \quad (12)$$

$$\forall n \in [m - b + 1, m - b + 2, \dots, m]$$

whereby S_n is the speed distribution containing all instantaneous speed measurements of all pedestrians for all time steps they spent within the infrastructure for all n subsequent replications.

Tests showed that regardless of the chosen values for b and $p_{\text{threshold}}$ the required number of replications depends on the exact seeds that are used and their order. Therefore, a predefined seed set was used during the calibration of PD to ensure that any differences between simulations using different parameter sets were not caused by the stochastic nature of the model. Using this predefined set, a value of 10 for b and a value of 0.25 for $p_{\text{threshold}}$, it was determined that the required number of replications was a 100 for the bidirectional scenarios, 50 for the corner high density scenarios, 40 for the t-junctions and corner low density scenarios, and 30 for the bottleneck scenario. For a more detailed discussion of the method and the choice for the values of b and $p_{\text{threshold}}$, the reader is referred to [25].

5. Calibration Results Based on Single Objectives

In this section the results of the individual objectives (a combination of a single scenario and a single metric) are discussed. Figures 5(a)–5(d) show boxplots of the objective values, determined by (11), per individual objective containing the objective values of all 3179 points of the search space. These plots provide insight into the distribution of the objective values. More importantly, these figures show the order of magnitude of the minimal objective value if the model would be calibrated using a single metric/objective. Here, smaller objective values represent a better fit of the model results to the reference data.

These plots provide insight into how the objective values are distributed but primarily show the order of magnitude of the minimal objective value if the model would be calibrated using only a single objective. The smaller the objective values the better the model results fit the reference data. By using a logarithmic scale the figure clearly illustrates how well the model fits to the data (when the optimal parameter set is used) and how this differs between the different scenarios and metrics.

Figures 5(e)–5(j) show boxplots of the nonnormalized, nonsquared errors and these provide insight into the size of the errors and how they are distributed. Here it is the case, the closer to zero the values are, the smaller the error between the simulation and the reference data is. Due to the linear scale these plot provide a better insight into how the value of the parameters influences the error and therefore the objective value.

Figures 5(a) and 5(e) show that for all scenarios the model can reproduce the flows well given both the small errors and the low minimal objective values. Furthermore, one can observe that on average the low density scenarios have a lower objective value but that the minimal objective value is smaller for the high density scenarios.

Figures 5(b) and 5(f) show that the model cannot reproduce the spatial patterns very well given the high errors and the large minimal objective values. The low density scenarios both have a lower objective value on average and a lower minimal objective value. Furthermore, especially for the high density case of the t-junction scenario the performance is relatively bad compared to the other scenarios.

Figures 5(d), 5(h), and 5(j) show that, except for the bidirectional high and t-junction high density scenarios, the travel time distribution can be reproduced well by the model. In the case of the bidirectional high and t-junction high density scenarios the figures show that the model can reproduce the mean and the standard deviation of the travel time distribution well individually but apparently not when they are combined. Also, similar to the spatial distribution, the low density scenarios have a lower objective value on average as well as a lower minimum objective value.

In the case of the effort metric Figures 5(c), 5(g), and 5(i) show that for most scenario the model cannot reproduce the effort distribution very well. The two exceptions are the bottleneck and t-junction high density scenarios. In these cases the model can reproduce the effort distributions well.

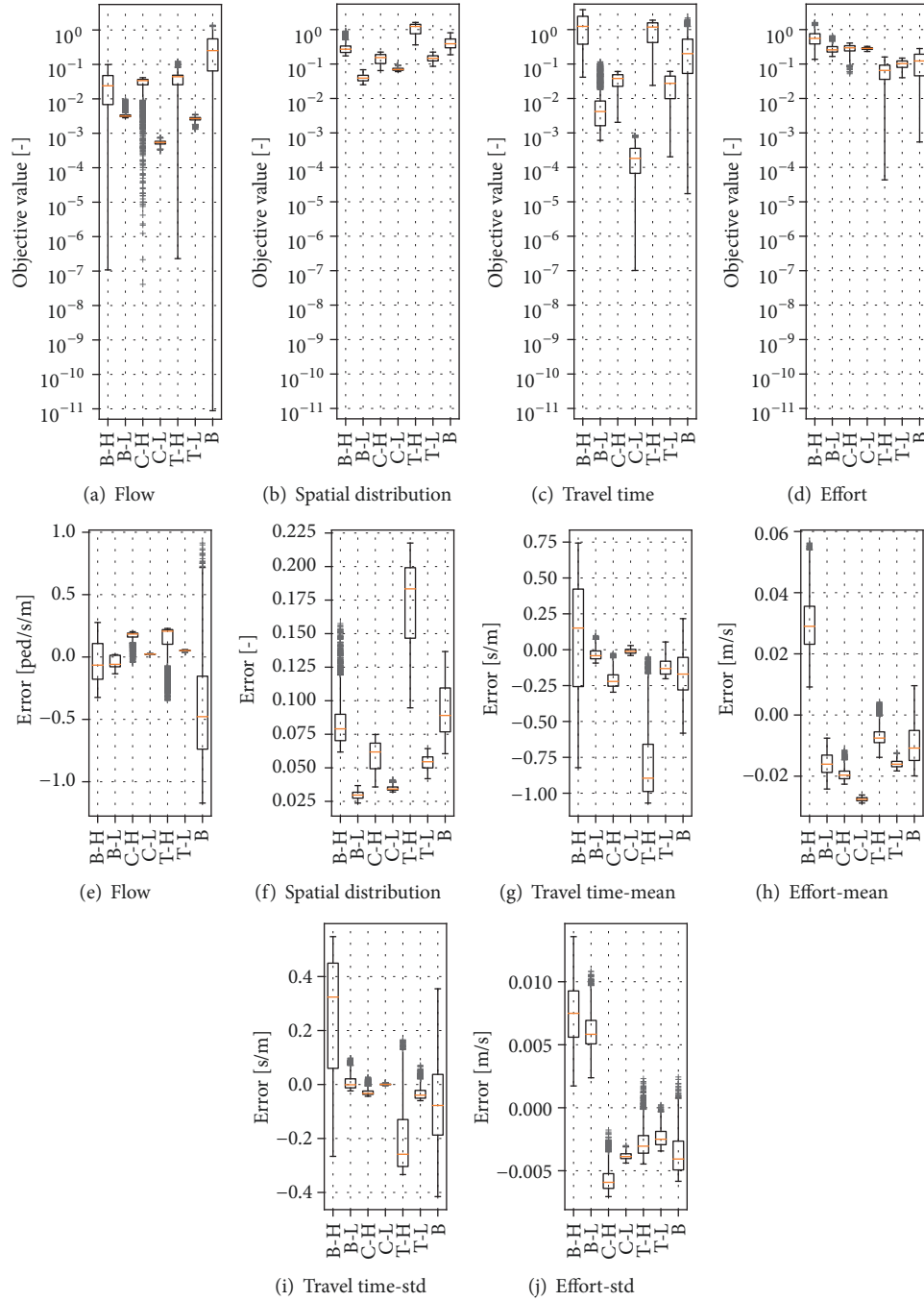


FIGURE 5: Results of calibrating the model using a single objective. Graphs (a)-(d) show, per combination of metric and scenario, how the objective values (calculated according to (11)) are distributed. Graphs (e)-(j) show the nonnormalized, nonsquared errors (i.e., $M_{\text{sim}}(\theta) - M_{\text{ref}}$) which give insight into size and the distribution of the errors. The flow scenarios are identified by their acronyms (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, and T-L = T-junction low).

Generally, the same pattern can be observed as the flow regarding the difference between the high and low density scenarios. That is, the high density scenarios generally have a lower minimum objective value but the low density scenarios generally have a lower objective value on average.

All figures show that both the size of the minimal objective value and the distribution of the errors depend on the

particular combination of scenario and metric. Furthermore, the figure illustrates that the model can generally reproduce the metrics related to the performance of the infrastructure (the flow and travel time) better than those more related to the underlying microscopic and macroscopic pedestrian dynamics (spatial distribution and the effort). However, regarding the difference between the performances of the

model on the different metrics, three things have to be noted.

Firstly, as Liao, Zhang, Zheng, and Zhao [31] show, even though the flow of their calibrated model is similar to the flow in the data, the underlying fundamental diagrams differ slightly. Regarding the results of this study, a quick review of the average velocities and average densities illustrates that differences exist between the reference data and the data obtained during the calibration. This finding suggests that the model and the data have slightly different underlying fundamental diagrams. Furthermore, it also suggests that, within the given search space, there does not seem to exist a parameter set that aligns the model's fundamental diagram to the data. This may in part explain why especially for the spatial distribution and effort metrics the model fit is worse than compared to the flow.

Secondly, as pointed out by Benner, Kretz, Lohmiller, and Sukennik [32], also a lack of detailed information about the boundary conditions (e.g., the exact distribution of the desired speeds and the order in which pedestrians enter the infrastructure) might negatively influence a model's capability to fit the data. Again, this lack of detailed information is likely to have the smallest impact on the flow as this is the most aggregated metrics of the four.

Lastly, some metrics might be more sensitive to changes in parameters of the pedestrian simulation model that are not taken into account in this study than the parameters included in the search space of the calibration. The sensitivity to these 'other' parameters is due to two things. First, the sensitivity analysis was performed before the choice of metrics for this calibration. Second, the speed distribution was used in the sensitivity analysis to determine significant differences in model result. Thus, this suggests that it is not only important to include multiple scenarios in the sensitivity analysis but also multiple metrics. This would ensure enhanced insight into the model's sensitivity and hence also better insight into which are the most important parameters to include in the search space during calibration.

This research aims to determine how different choices regarding scenarios and metrics influence the calibration and not calibration of the InControl model. Therefore, differences between the simulation model results and the data are not investigated in more detail in this paper.

6. Differences in Performance between Calibration Strategies

In this section the results of different calibration strategies are discussed. First, a general analysis of the results is performed based on the obtained optimal parameter sets for all of the 16 combinations. Afterwards, the results of different strategies are compared to determine the influence of movement base cases, density levels and metrics. Table 3 identifies 16 different calibration strategies whereby the table indicates which scenarios and metrics are included during the calibration according to a certain strategy.

Table 4 presents the optimal parameter sets for all 16 strategies. 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, the optimal objective values in Table 4 are notably higher than those found in Figure 5 which indicates that a combination of objectives decreases the fit of the model with respect to the data. Next to that, for all 16 strategies, 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, the results of the calibration indicate that it is more important to take those pedestrians into account who are in front rather than those who are more to the side. Furthermore, in the case of the relaxation time the parameter value also frequently lies on the boundary of the search space. Due to time constraints it was not possible to extend the search space. However, an analysis is performed to ascertain the likely effects changing the search space, this includes both extending it and increasing the precision, would have on the location of the optimal parameter sets.

6.1. Precision of the Grid Search and the Values on the Search Space Boundaries. The search space is limited both by its boundaries and by its precision. To obtain insight into the likelihood that the results would change significantly (i.e., the location of the optimal parameter set changes significantly) an analysis is performed. The analysis is based on visual inspection of two types of graphs.

Figure 6 provides insight into how likely it is that the location of the optimal parameter changes significantly if the precision of the search space is increased. It is considered likely that the location of the optimal parameter set can change significantly if the search space contains points with a very similar objective value, compared to the minimal objective value, which are located in a significant different part of the search space. The graphs show the relation between the decrease in GoF and the distance from the optimal parameter set. The decrease in GoF is calculated by (14), whereby a small decrease in the GoF equals a small difference between the objective value of a given parameter set and the minimal objective value. The distance is calculated by

$$D(\theta, \theta^*) = \sqrt{N_\tau(\theta, \theta^*)^2 + N_\phi(\theta, \theta^*)^2 + N_r(\theta, \theta^*)^2} \quad (13)$$

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.

Figure 6(a) shows that for the bidirectional low case there are many points that have a fairly similar GoF and that many of them are in significantly different parts of the search space (i.e., are at a large distance from the optimal parameter set). Hence, it could be likely that a change in the precision of the search space would lead to a significant different location of the optimal parameter set.

TABLE 3: Tested combination of scenarios and metrics, where the acronyms identify the metrics (i.e., Q = flow, SD = spatial distribution, Eff = effort, and 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, and T-L = T-junction low).

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

TABLE 4: Calibration results, where $O(\theta)$ represents the optimal value of the objective function. The cells in italic font 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]
1. Bidirectional high	0.1329	<i>0.620</i>	<i>57.00</i>	0.15296
2. Bidirectional low	0.0588	<i>0.620</i>	<i>57.00</i>	0.19120
3. Corner high	0.0561	0.395	<i>57.00</i>	0.23900
4. Corner low	0.0742	<i>0.380</i>	61.50	0.23900
5. T-junction high	0.1190	0.590	<i>57.00</i>	0.21998
6. T-junction low	0.0468	<i>0.380</i>	68.25	0.23900
7. Bottleneck	0.1093	0.395	68.25	0.20076
8. Flow	0.0146	<i>0.380</i>	59.25	0.20076
9. Spatial distribution	0.2015	0.575	59.25	0.21988
10. Travel time	0.1814	<i>0.620</i>	59.25	0.15296
11. Effort	0.1798	0.500	<i>57.00</i>	0.23900
12. High density scenarios	0.2647	0.575	<i>57.00</i>	0.21032
13. Low density scenarios	0.0722	0.500	<i>57.00</i>	0.23900
14. All scenarios - Macro	0.1444	0.545	59.25	0.21988
15. All scenarios - Meso	0.2012	<i>0.620</i>	59.25	0.15296
16. All combined	0.1814	0.575	<i>57.00</i>	0.21032

Figure 6(b) shows that this is not the case for the bottleneck case. Figures 6(c) and 6(d) show a few points whose GoF is similar to that of the optimal point but whose distance to the optimal point is rather large (see the red rectangles). For these cases it is determined which parameter(s) is/are the main contributor to this large distance. In both examples here, it is found that primarily the viewing angle changes. So, in these cases it is likely that, even if the precision of the search space is increased, the location of the optimal parameter set

will not change significantly in relation to both the relaxation time and the radius. The location regarding the viewing angle is less certain though.

The main conclusion of the visual inspection of these graphs for all 16 combinations is that for most of the low density cases, the exception is the T-junction low case, and the flow case it is likely that the location of the optimal parameter set could change significantly if the precision of the search space is increased. For all other cases this is not the case.

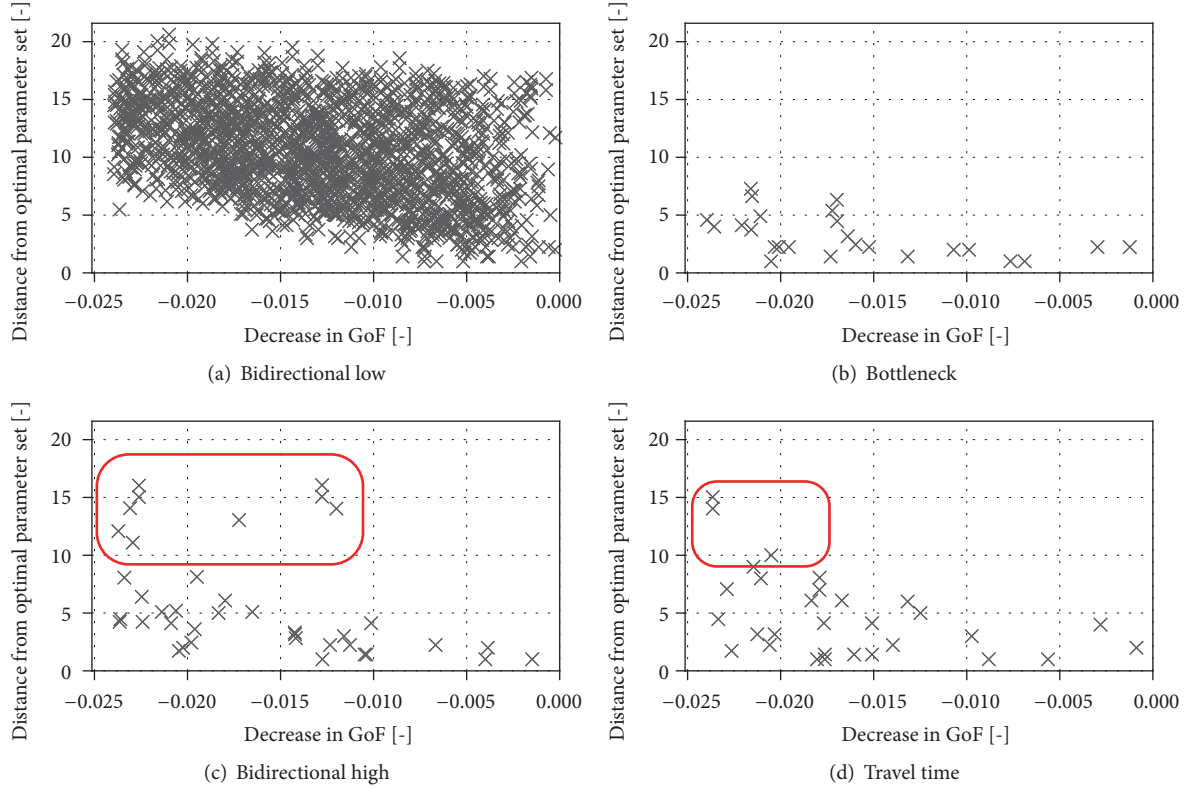


FIGURE 6: The deviation from optimal GoF versus the distance to the optimal parameter set.

Figure 7 shows how the objective value changes over the search space for two examples. These graphs provide insight into the question if it is likely that an extension of the search space would cause the optimal parameter to be located in a significantly different part of the search space in relation to the parameter whose value is on the search space boundary.

For example, Figures 7(a)–7(c) show no discernible pattern. Hence one cannot state with a high degree of certainty that if the search space is extended the relaxation time will be equal to or larger than the current value and that the viewing angle will be equal to or smaller than its current value.

Figures 7(d)–7(f), on the other hand, do show a clear pattern. In this case it is likely that an extension of the search space will result in an optimal parameter set whose viewing angle is equal to or smaller than the current optimal viewing angle. The patterns shown in the graphs also make it likely that the optimal radius and relaxation time will be fairly similar to the current optimal values.

Performing this analysis for all 16 combinations yields the conclusion that, for all cases where the previous analysis did conclude that an increase in the precision is unlikely to change the results, an extension of the search space is likely to increase the differences between the cases instead of decreasing it.

Overall, the analysis shows that for some cases the location of the optimal parameter set could change significantly if the search space is changed. However, it also shows that the large differences between the optimal parameter sets, currently found, are also likely to be found if the search space

is adapted. Hence, the authors expect that an extension of the search space results in even larger differences in the optimal parameter sets than the differences identified in this paper.

6.2. Identification of Differences in Performance between Calibration Procedures. In order to illustrate the differences between the optimal parameter sets a cross-comparison of the goodness-of-fit is performed. These comparisons are based on the difference between the optimal GoF of combination A and the GoF of combination A when the optimal parameter set of combination B is used (see (14)).

$$\Delta GoF_{A:B} = -(O_A(\theta_B^*) - O_A(\theta_A^*)) \quad (14)$$

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 combination B is used. As stated in the methodology section, an increase in the value of the objective function equals a decrease in the GoF, hence the minus value equation (14). Thus, the larger the decrease in GoF, the worse the fit of the model to the data is if the given parameter set is used instead of the optimal parameter set. In the remainder of this section the effects the choice of movement base case, the density level and metrics on the GoF are discussed in more detail.

6.3. Effect of Movement Base Case on Multiple-Objective Calibration Results. Figure 8 presents the results of a comparison

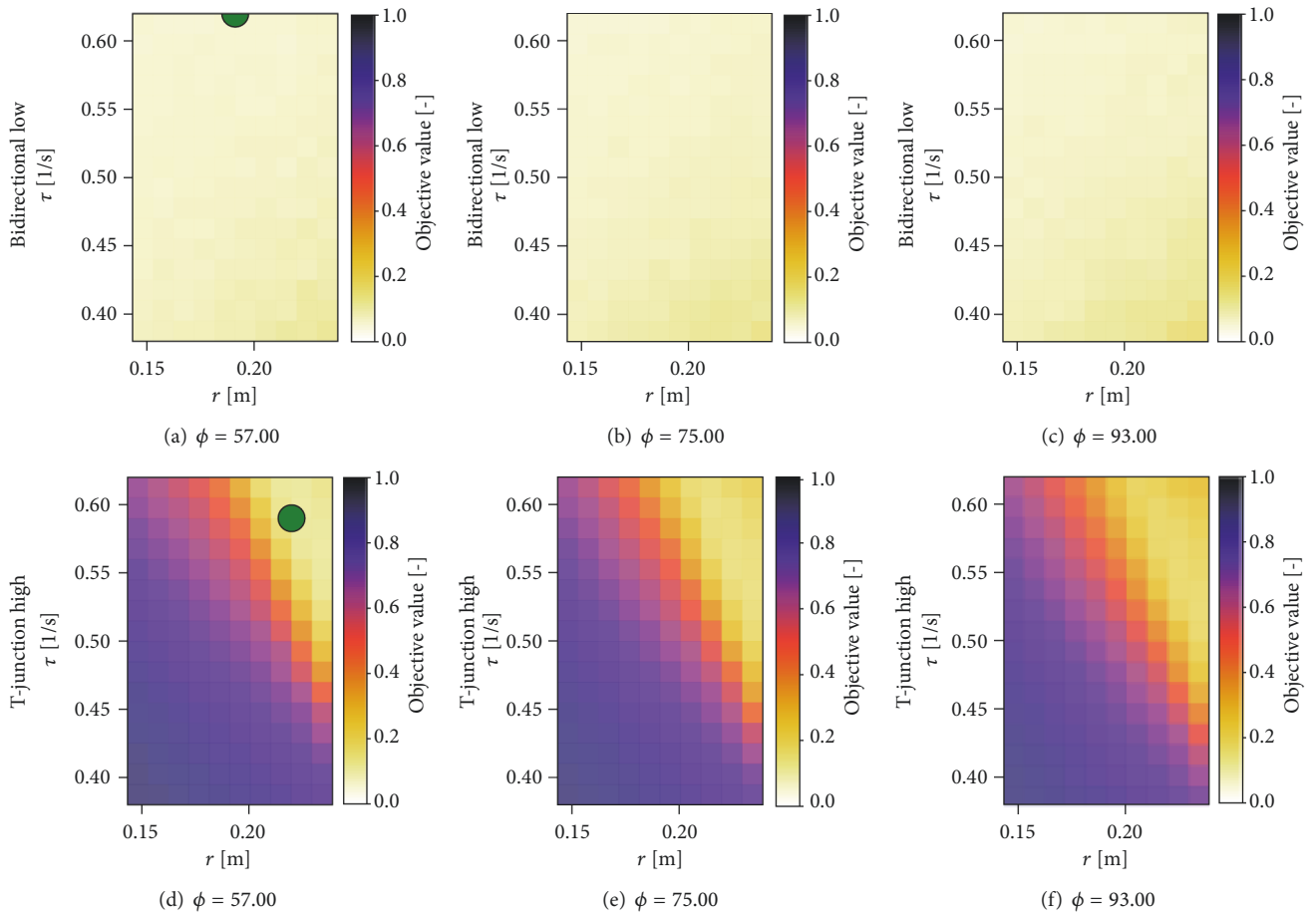


FIGURE 7: Examples of surface plots showing how the objective value changes over the search space. The green dot identifies the location of the optimal parameter set.

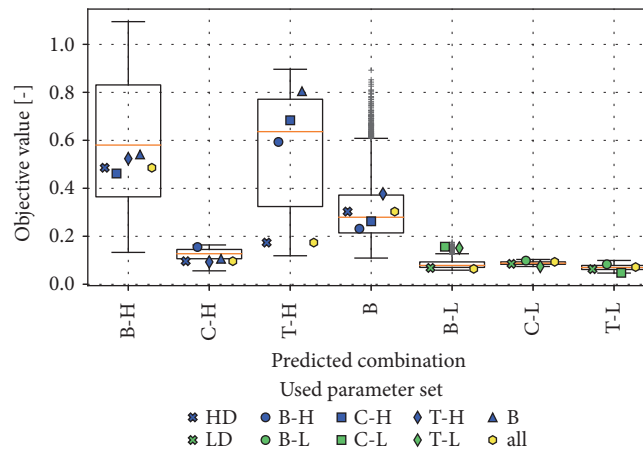


FIGURE 8: The distribution of the objective values per movement base case combination is displayed by the boxplots. The markers indicate, per used parameter set, how much the objective value would deviate from its optimum if the optimal parameter set, obtained using another (combination of) movement base case(s), would be used. The combinations are identified by their acronyms as found in Table 3.

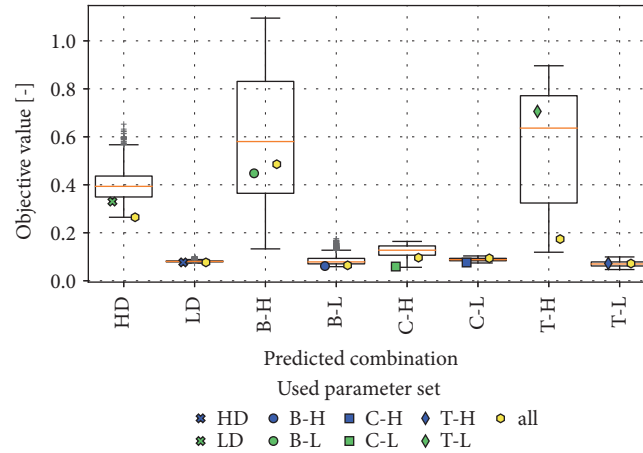


FIGURE 9: The distribution of the objective values per density level combination is displayed by the boxplots. The markers indicate, per used parameter set, how much the objective value would deviate from its optimum if the optimal parameter set, obtained using the same (combination of) movement base case(s) but a different density level, would be used. The combinations are identified by their acronyms as found in Table 3.

between different calibration strategies, in which the difference in goodness-of-fit is depicted. All comparisons are made between (combinations of) scenarios of the same density level, in order to exclude the possibility that differences are caused by a difference in the level of density and not by a difference in movement base case.

The figure shows the distribution of the objective values for the different movement base cases. The markers depict the objective value if the optimal parameter set obtained using another movement base case is used. The difference between the location of the marker and the minimal objective value, as indicated by the boxplot, indicates the difference in GoF.

The boxplots show that, generally, the low density cases have a higher GoF and that they are less sensitive to changes in the parameter set regarding their fit to the data. Furthermore, for both the low and high density cases the corner scenarios seem least sensitive to changes in the parameter set. The data, moreover, illustrates that in all cases the GoF of the individual movement base cases decreases when the parameter set based on another movement base case or a set of movement base cases is used. However, the size of the decrease clearly depends on the scenario as well as which scenario's optimal parameter set is used. A few notable observations can be made regarding the decreases in GoF.

Firstly, in the case of the high density bidirectional scenario (B-H) all other optimal parameter sets from the other high density combinations lead to a similar decrease in GoF. This is also the case for the low density bidirectional scenario (B-L). However, in this case the optimal parameter set obtained when combining all three low density scenarios results in a far smaller decrease in GoF indicating that the bidirectional scenario has a strong influence on the objective function of this combination. Overall, both observations indicate that the bidirectional movement base case contains behaviours which are not well captured by other movement base cases.

Secondly, the high density t-junction scenario (T-H) also seems to contain behaviours which are not captured well by other high density movement base cases. Furthermore, as the decrease of GoF is clearly smallest for the optimal parameter set obtained using a combination of all four high density scenarios, it is clear that the objective function of this combination is strongly influenced by the t-junction scenario.

Thirdly, in the cases of the low density corner (C-L) and t-junction scenarios (T-H, T-L) the decrease in GoF is very small when the optimal parameter set of the other scenario is used. Table 4 furthermore identifies that the optimal parameter sets for these two scenarios are very similar. This indicates that at low densities the t-junction movement base case effectively reduces to a corner movement base case as the densities are so low that there is probably little merging behaviour.

Overall, based on this analysis we conclude that different movement base cases contain different behaviours which are not necessarily captured when the model is calibrated using other movement base cases. The exceptions are the corner and t-junction base cases at low densities (C-L, T-L). Using an optimal parameter set obtained by combining multiple movement base cases mitigates this problem somewhat. However, this still results in clear decreases in the GoF for all movement base cases.

6.4. Effect of Density Level on Multiple-Objective Calibration Results. In Figure 9 the results of the comparison between the GoF's for different density levels are presented. The data shows that in all three cases the decrease in the GoF is smaller when the optimal parameter set of the high density case is used in the low density case than vice versa. However, there are clear differences to be observed between the movement base cases. In the case of the corner movement base case the level of density does not have a clear effect as using the optimal parameter set of the other density level only leads to

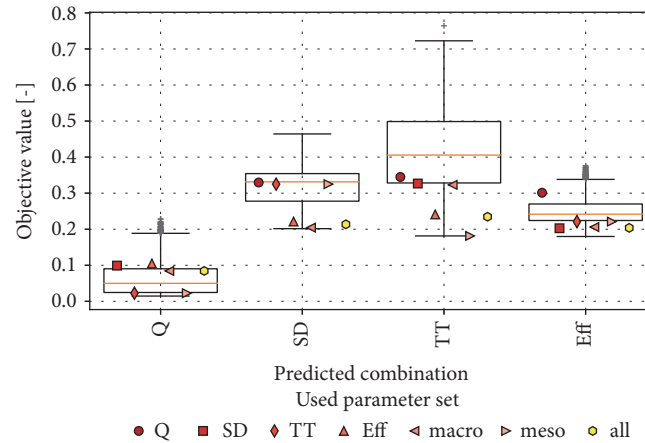


FIGURE 10: The distribution of the objective values per metric combination is displayed by the boxplots. The markers indicate, per used parameter set, how much the objective value would deviate from its optimum if the optimal parameter set, obtained using another (combination of) metrics(s), would be used. The combinations are identified by their acronyms as found in Table 3.

a very small decrease in the GoF. This is also the case when the optimal parameter set of the high density bidirectional scenario is used for the low density bidirectional scenario. However, vice versa it is not the case. For the t-junction scenarios, the figure clearly shows that in both cases using the optimal parameter set of the other density level results in a large decrease in the GoF.

The data also illustrates that the decrease in GoF of the combination of high density scenarios is larger when the optimal parameter set of the combination of low density is used than vice versa. This remains the case even if the bottleneck scenario is omitted from the high density set, such that the high density set contains exactly the same movement base case as the low density set. In this case the decrease in GoF for the high density set becomes even larger.

Overall, it can be concluded that the level of density of the scenario does influence the calibration results. Therefore, it is concluded that it is more important to include the high density scenarios than the low density scenarios. Furthermore, depending on the movement base case, it can even be the case that the low density variant can be omitted as it will not add any value.

6.5. Effect of the Metrics on the Multiple-Objective Calibration Results. In Figure 10 a comparison is visualised between the influences of the different parameter sets on the performance of the metrics. There seems to be a correlation between the distribution of the effort and the spatial distribution (i.e., SD, Eff, Meso, and all). When the model is calibrated using only one of them, the decrease in the GoF of the other is small. Besides that, both the use of the spatial distribution and the use of the distribution of the effort result in a far worse prediction of the flow compared to the distribution of the travel times. That is, the decrease in GoF of the flow is far larger in case the optimal parameter set of the spatial distribution or the use of the distribution of the effort is used. Lastly, the optimal parameter sets obtained using combinations of metrics are more heavily influenced

by certain metrics. When only the macroscopic metrics are applied, the spatial distribution clearly has a larger impact on the location of the optimal parameter set given the lower decrease in GoF. When solely using the mesoscopic metrics, the distribution of the travel time has a larger impact compared to the distribution of the effort.

These results show that the choice of metrics does influence 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.

7. Conclusions, Discussion, and Implications for Practice

The findings of this research regarding the influence of the movement base cases are found to be consistent with both [13, 14]. Similar to those studies, this research finds that (1) it is necessary to use multiple movement base cases, when calibrating a model, to capture all relevant behaviours and (2) the GoF of the individual movement base cases decreases when the parameter set based on multiple movement base cases is used.

Hence, this research confirms that 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 (i.e., it does not need to accurately replicate all movement base cases and/or metrics), it might be preferred to use a limited set of movement base cases during the calibration, in particular, given the fact that the GoF of the individual movement base case decreases when multiple movement base cases are used during the calibration.

The level of density also influences the calibration results. Thus, depending on the intended use of the model different density levels should be taken into account during the calibration. Furthermore, this study concludes that it is more

important to incorporate high density scenarios. As a result, one can omit some of low density scenarios, in particular the bidirectional and corner low density scenarios.

This study also finds that the calibration results depend on the choice of metric or combinations of metrics. Depending on the combination of metrics, also the choice of objective function and normalization method influences the results. Consequently, depending on the usage of the model, one should decide which metric or metrics are most important and how to reflect the difference in importance of these metrics when combining multiple objectives into one.

The results also show that the relaxation time is the only parameter to which the model is sensitive to in all scenarios. Its exact value thus has a large impact on how well the simulations fit the data. Reference [33] found similar results for a different model, though they only studied a bottleneck. Furthermore, it is also the only parameter where the optimal value can lie on both sides of the search space depending on the used combination of scenarios and metrics. This can possibly be explained by the fact that the relaxation parameter has multiple roles as is described in [34]. As Johansson et al. [34] conclude, this indicates that the models may be too simplistic.

A number of things have to be noted when reflecting on the method of handling the multiple objectives. In light of the goal of this study, the method presented in this study was chosen to assure as little as possible bias towards any of the metrics or scenarios. However, if one calibrates a model given a certain type of model usage, one might use different weights or even different optimization method. For example, one can search for the Pareto optimal solution [29] or combine multiple objectives using the ϵ -constraint method [29]. Both the choice of the optimization method and its influence on the calibration results are relevant topics for future research.

Though this study is more extensive than previous studies, it still has a number of limitations. Firstly, due to limitations of the available dataset, this study did not include a crossing movement base case. So, it is unclear whether and to what extent the crossing movement base case contains behaviours which are not captured by other movement base cases. More research, in which also effects of the intersecting movement base cases are included, is needed to create a comprehensive overview of the effect the choice of movement base cases has on the calibration. Secondly, only one microscopic model was used in this study. Therefore, as such, it is unclear to what extent the current findings can be generalized and whether the conclusions of this study also hold for other microscopic models. The fact that the results are consistent with [13, 14] and the fact that these studies used another microscopic model does indicate that this is the case. Lastly, as already reflected upon, this study did not look into the effects of the chosen calibration methodology. Insights into these effects could be relevant when deciding on the calibration methodology for a microscopic model with a certain intended use.

All in all, the results show two important things. Firstly, the optimal parameter set obtained using a single or limited amount of objectives does not always provide an accurate fit of the model to the data for other combinations of scenarios

and metrics. Besides that, using multiple objectives (e.g., using multiple high density scenarios) to calibrate the model decreases the GoF of the model to the data for all the individual objectives. These two conclusions imply 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. These results also raise an important question. Is the implicit assumption that the behaviour of the pedestrians is independent of the flow situation, which is at the foundation of most pedestrian simulation models, valid? This research cannot answer this question because of, among other, its limitations on the number of parameters included in the search space. Hence it is an interesting topic for future research.

Lastly, these results show that a model only provides accurate results for scenarios and metrics that the model has been calibrated and validated on. Using the model for predictions of other scenarios and metrics is likely to result in large inaccuracies. Hence, it is essential that calibration or validation attempts include multiple scenarios and multiple metrics.

Data Availability

The empirical data used in this study can be found at <http://ped.fz-juelich.de/da/>.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The research presented in this paper is part of the research program “Allegro: Unravelling slow mode traveling and traffic: with innovative data to a new transportation and traffic theory for pedestrians and bicycles” (ERC Grant Agreement no. 669792), a Horizon 2020 project which is funded by the European Research Council. Additionally, the research presented in this paper has been presented at the Transportation Research Board 97th Annual Meeting (2018, Washington DC, United States).

References

- [1] C. Rudloff, T. Matyus, S. Seer, and D. Bauer, “Can walking behavior be predicted? Analysis of calibration and fit of pedestrian models,” *Transportation Research Record*, no. 2264, pp. 101–109, 2011.
- [2] S. Seer, N. Brändle, and C. Ratti, “Kinects and human kinetics: A new approach for studying pedestrian behavior,” *Transportation Research Part C: Emerging Technologies*, vol. 48, pp. 212–228, 2014.
- [3] A. Abdelghany, K. Abdelghany, and H. Mahmassani, “A hybrid simulation-assignment modeling framework for crowd dynamics in large-scale pedestrian facilities,” *Transportation Research Part A: Policy and Practice*, vol. 86, pp. 159–176, 2016.

- [4] J. L. Berrou, J. Beecham, P. Quaglia, M. A. Kagarlis, and A. Gerodimos, "Calibration and validation of the legion simulation model using empirical data," in *Pedestrian and evacuation dynamics 2005*, N. Waldau, P. Gattermann, H. Knoflacher, and M. Schreckenberg, Eds., pp. 167–181, 2007.
- [5] M. Davidich and G. Köster, "Predicting pedestrian flow: A methodology and a proof of concept based on real-life data," *PLoS ONE*, vol. 8, no. 12, Article ID e83355, 2013.
- [6] W. Klein, G. Köster, and A. Meister, "Towards the calibration of pedestrian stream models," in *Proceedings of the International conference on parallel processing and applied mathematics*, R. Wyrzykowski, J. Dongarra, K. Karczewski, and J. Wasniewski, Eds., pp. 521–528, 2010.
- [7] R. Lovreglio, E. Ronchi, and D. Nilsson, "Calibrating floor field cellular automaton models for pedestrian dynamics by using likelihood function optimization," *Physica A: Statistical Mechanics and its Applications*, vol. 438, pp. 308–320, 2015.
- [8] T. Robin, G. Antonini, M. Bierlaire, and J. Cruz, "Specification, estimation and validation of a pedestrian walking behavior model," *Transportation Research Part B: Methodological*, vol. 43, no. 1, pp. 36–56, 2009.
- [9] A. Schadschneider, C. Eilhardt, S. Nowak, and R. Will, "Towards a calibration of the floor field cellular automaton," in *Pedestrian and evacuation dynamics*, pp. 557–566, 2011.
- [10] M. Davidich and G. Köster, "Towards automatic and robust adjustment of human behavioral parameters in a pedestrian stream model to measured data," *Safety Science*, vol. 50, no. 5, pp. 1253–1260, 2012.
- [11] M. C. Campanella, S. P. Hoogendoorn, and W. Daamen, "Improving the nomad microscopic walker model," *IFAC Proceedings Volumes*, vol. 42, no. 15, pp. 12–18, 2009.
- [12] S. P. Hoogendoorn and W. Daamen, "Microscopic calibration and validation of pedestrian models: Cross-comparison of models using experimental data," in *Traffic and granular flow'05*, A. Schadschneider, T. Pöschel, R. Kühne, M. Schreckenberg, and D. E. Wolf, Eds., pp. 329–340, Springer, Berlin, Germany, 2007.
- [13] M. C. Campanella, S. P. Hoogendoorn, and W. Daamen, "A methodology to calibrate pedestrian walker models using multiple-objectives," in *Pedestrian and evacuation dynamics*, pp. 755–759, 2011.
- [14] D. C. Duives, *Analysis and modelling of pedestrian movement dynamics at large-scale events [Doctoral, thesis]*, 2016.
- [15] A. Duret, C. Buisson, and N. Chiabaut, "Estimating individual speed-spacing relationship and assessing ability of newell's car-following model to reproduce trajectories," *Transportation Research Record*, no. 2088, pp. 188–197, 2008.
- [16] S. Ossen and S. P. Hoogendoorn, "Validity of trajectory-based calibration approach of car-following models in presence of measurement errors," *Transportation Research Record*, no. 2088, pp. 117–125, 2008.
- [17] V. Punzo, B. Ciuffo, and M. Montanino, "Can results of car-following model calibration based on trajectory data be trusted?" *Journal of the Transportation Research Board*, vol. 2315, pp. 11–24, 2012.
- [18] D. Wolinski, S. J. Guy, A.-H. Olivier, M. Lin, D. Manocha, and J. Pettré, "Parameter estimation and comparative evaluation of crowd simulations," *Computer Graphics Forum*, vol. 33, no. 2, pp. 303–312, 2014.
- [19] F. S. Hänseler, M. Bierlaire, B. Farooq, and T. Mühlematter, "A macroscopic loading model for time-varying pedestrian flows in public walking areas," *Transportation Research Part B: Methodological*, vol. 69, pp. 60–80, 2014.
- [20] M. Ko, T. Kim, and K. Sohn, "Calibrating a social-force-based pedestrian walking model based on maximum likelihood estimation," *Transportation*, vol. 40, no. 1, pp. 91–107, 2013.
- [21] M. C. Campanella, S. Hoogendoorn, and W. Daamen, "Quantitative and qualitative validation procedure for general use of pedestrian models," in *Pedestrian and evacuation dynamics 2012*, U. Weidmann, U. Kirsch, and M. Schreckenberg, Eds., pp. 891–905, 2014.
- [22] M. Moussaid, D. Helbing, and G. Theraulaz, "How simple rules determine pedestrian behavior and crowd disasters," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 17, pp. 6884–6888, 2011.
- [23] I. Karamouzas, R. Geraerts, and M. Overmars, "Indicative routes for path planning and crowd simulation," in *Proceedings of the 4th International Conference on the Foundations of Digital Games, ICFDG 2009*, pp. 113–120, USA, April 2009.
- [24] T. W. Anderson and D. A. Darling, "Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes," *Annals of Mathematical Statistics*, vol. 23, no. 2, pp. 193–212, 1952.
- [25] M. Sparnaaij, *How to calibrate a pedestrian simulation model [Master's Thesis]*, 2017, <https://repository.tudelft.nl/islandora/search/?collection=education>.
- [26] W. Daamen, *Modelling Passenger Flows in Public Transport Facilities*, 2004, <http://resolver.tudelft.nl/uuid:e65fb66c-1e55-4e63-8c49-5199d40f60e1>.
- [27] C. Keip and K. Ries, *Dokumentation Von Versuchen Zur Personenstromdynamik - Projekt "Hermes"*, Bergische Universität Wuppertal, 2009, http://ped.fz-juelich.de/experiments/2009.05.12_Duesseldorf_Messe_Hermes/docu/VersuchsdokumentationHERMES.pdf.
- [28] S. C. Wong, W. L. Leung, S. H. Chan et al., "Bidirectional pedestrian stream model with oblique intersecting angle," *Journal of Transportation Engineering*, vol. 136, no. 3, pp. 234–242, 2010.
- [29] S. H. Zak and E. K. P. Chong, "Multiobjective optimization," in *An Introduction to Optimization*, vol. 24, pp. 577–598, 2013.
- [30] E. Ronchi, E. D. Kuligowski, P. A. Reneke, R. D. Peacock, and D. Nilsson, "The process of verification and validation of building fire evacuation models," Technical Note (NIST TN) - 1822, 2013.
- [31] W. Liao, J. Zhang, X. Zheng, and Y. Zhao, "A generalized validation procedure for pedestrian models," *Simulation Modelling Practice and Theory*, vol. 77, pp. 20–31, 2017.
- [32] H. Benner, T. Kretz, J. Lohmiller, and P. Sukennik, "Is calibration a straight-forward task if detailed trajectory data is available?" in *Transportation Research Board 96th Annual Meeting*, 2017.
- [33] M. Haghighi and M. Sarvi, "Simulating pedestrian flow through narrow exits," *Physics Letters A*, vol. 383, no. 2, pp. 110–120, 2019.
- [34] F. Johansson, D. Duives, W. Daamen, and S. Hoogendoorn, "The many roles of the relaxation time parameter in force based models of pedestrian dynamics," *Transportation Research Procedia*, pp. 300–308, 2014.

