MULTI-OBJECTIVE CALIBRATION FRAMEWORK FOR PEDESTRIAN SIMULATION MODELS

- Study on the Effect of Movement Base Cases, Metrics and Density Levels -

Martijn Sparnaaij
Stevinweg 1, 2628 CN, Delft, The Netherlands
m.sparnaaij@tudelft.nl

Dr. Dorine C. Duives (Corresponding author)
Stevinweg 1, 2628 CN, Delft, The Netherlands
d.c.duives@tudelft.nl
+31 (15) 278 6304

Dr. Victor Knoop
Stevinweg 1, 2628 CN, Delft, The Netherlands
v.knoop@tudelft.nl
+31 (15) 278 8413

Prof. Serge P. Hoogendoorn
Stevinweg 1, 2628 CN, Delft, The Netherlands
s.p.hoogendoorn@tudelft.nl
+31 (15) 278 5475

1 Augustus 2017

Keywords:
Pedestrian simulation models, Multi-objective calibration, Effect study,
Movement base case, Metrics, Density levels

Submitted at TRB 2018:
6048 words + 1 figures * 250 words + 5 tables * 250 words
= 7548 words

ABSTRACT

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 on one specific movement base case only 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 two-fold, namely to 1) determine the effect of the choice of movement base cases, metrics and density levels on the calibration results and 2) 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 normalised 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.

INTRODUCTION

Ideally, a multitude of steps has to be taken before a commercial implementation of a pedestrian model is used in practice. Within this process, 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.

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. 6-8) usually by using the fundamental diagram (9). 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, 9, 10). 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.

It is questionable how generally applicable a pedestrian simulation model is that has been calibrated using a limited set of movement base cases. Research by (11) and (12) shows that using different flow situations and different metrics lead to different optimal parameter values. That is, both (11) and (12) identify that for general usage (i.e. using a single model for many different applications) one needs to calibrate using multiple movement bases to capture all relevant behaviours. The effect of using different metrics during the calibration has been investigated by (12) in relation to pedestrian dynamics and among others (13-15) in relation to vehicular traffic. The study shows that different combinations of metrics clearly lead to different calibration results.

To overcome this problem, three multiple-objective frameworks have been proposed which try to take a more inclusive approach. (16) Uses multiple metrics to compare the model results to the reference data, while (11) uses multiple movement base cases with multiple metrics. However, during the calibration procedure both studies used only one metric. The work by (12) 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.

Even though these works illustrate that using a parameter set obtained by calibrating the model using multiple objectives results in a better validation score, these works still simplified the calibration procedure. These preliminary studies into the calibration of microscopic simulation models did not account for all movement base cases and used fairly ad-hoc methods to balance the objectives in the objective function.

The objective of this research is two-fold, namely to determine the effect of the choice of movement base cases, metrics and density levels on the calibration results and to develop a multiple-objective calibration approach for pedestrian simulation models to determine the aforementioned effects. The following section first elaborates on the calibration methodology. The main features of the pedestrian simulation model are briefly described next. Afterwards, the results of the multiple-objective calibration procedure are presented. This paper closes of with a discussion of the results, conclusions and the implications of this work for practice.

METHODOLOGY

This section presents the reasoning behind the newly developed calibration methodology. First the scenarios are identified. Accordingly the metrics and the objective function are presented. This section furthermore elaborates on the stopping criteria and the manner that the stochasticities of the pedestrian simulation model are handled.

Scenarios

Research by both (11) and (12) 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 (17) which show that models perform better in a general setting when they are calibrated using multiple base cases.

Contemporary, several data sets are available that feature the movement of pedestrians in multiple movement base cases and a similar population of pedestrians, among others (18-20). 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 data set 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 unidirectional corner flow, a merging flow, a bidirectional flow 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. For a more detailed overview of the experimental setup within the HERMES project the reader is referred to (20). Care is taken to ensure a similar flow pattern over time, speed distribution and route choice, details on the exact simulation of the seven scenarios are mentioned in (21).

Metrics

In this multiple-objective framework four different metrics are used to identify how different metrics impacts 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 effort distribution and the travel time distribution. Microscopic metrics, i.e. trajectories, are not used for three reasons. Firstly, calibration based on trajectories requires a different approach than calibrating based on macro and mesoscopic metrics. 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) (17) and given that calibrating based on microscopic metrics does not necessarily result in a macroscopically valid model (9) macroscopic and mesoscopic metrics take priority over microscopic metrics.

Flow

The flow is chosen as a macroscopic metric to check how well the model is capable of reproducing the throughput in different situations. In all seven scenarios the average flow is measured along a certain cross-section during a certain measurement period according to Eq. 1. The average flow is calculated as follows:

$$\bar{q}_i = \frac{N_i}{\Delta t * l} \tag{1}$$

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 measurement period (Δt). The flow in normalised 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

(12) Showed that microscopic models might not always be able to accurately reproduce the spatial distribution patterns. Hence, it is essential to check whether this 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 x 0.4 m, which is approximately the size of one pedestrian during a high density situation, overlays the measurement area and for every cell the percentage of the time it is occupied is determined in Eq. 2.

$$F_j = \frac{N_{occ;j}}{N_{steps}} \tag{2}$$

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.

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 Eq. 3).

$$e_k = \frac{\sum_{n-1}(|v_{l;x} - v_{l-1;x}| + |v_{l;y} - v_{l-1;y}|)}{n-1} \tag{3}$$

$$v_{l;x} = \frac{x_l - x_{l-1}}{\Delta t} \tag{4}$$

where $v_{l;x}$ and $v_{l;y}$ are respectively the speed in the x and y-direction at time step l and n the number of time steps. The speeds are obtained by differentiating the positions (Eq. 4), where x_l is the x-position at time step l and Δt is the duration of the time step. The effort measurements of all pedestrians are combined into a distribution.

Travel time

The travel time is the time it takes a pedestrian to traverse the measurement area (Eq. 5)

$$TT_k = \frac{t_{end} - t_{start}}{l_{ref}} \quad (5)$$

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. The travel time is normalised in order to simplify the comparison between different scenario with different average path lengths. Note that this metric approximates the realized pace of each individual. That is, an individual who makes a detour at a very high speed does not influence the travel time, but will influence the effort metric.

Only the travel time of those pedestrians who successfully traversed the whole measurement area during the measurement period are used to form the distribution of the travel times.

Objectives

In this research multiple objectives are combined into a single objective using the weighted sum method (22). This is in line with research by (12), 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, normalisation is necessary, as the metrics have different units and different orders of magnitude. The adopted normalisation method uses a single normalization value per metric whereby this normalization value is determined based on the ratios between the metric values obtained from the reference data. For a detailed explanation of this method and an underpinning of the choice to specifically use this method, the reader is referred to (21). The objective function for a given metric and scenario is given by the normalised Squared Error (SE) for the macroscopic metrics (Eq. 6) and the mesoscopic metrics (Eq. 7).

$$SE_{norm}(\theta) = \frac{1}{m} \sum_j \left(\frac{\frac{\sum_i M_{sim;i;j}(\theta)}{n} - M_{ref;j}}{M_{norm}} \right)^2 \quad (6)$$

$$SE_{norm;meso}(\theta) = \frac{1}{2} \left(\frac{M_{sim;\mu}(\theta) - M_{ref;\mu}}{M_{norm;\mu}} \right)^2 + \frac{1}{2} \left(\frac{M_{sim;\sigma}(\theta) - M_{ref;\sigma}}{M_{norm;\sigma}} \right)^2 \quad (7)$$

where M_{sim} is the metrics value according to the simulation, M_{ref} the reference value according to the data, M_{norm} the value used for the normalisation and θ the vector of model parameters. In the case Eq. 6 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 Eq. 7 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:

$$O(\theta) = \frac{1}{N_s * N_m} \sum_s \sum_m SE_{norm;s;m}(\theta) \quad (8)$$

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. A likelihood method, which multiplies probabilities, might not work in this case, as the method will always attempt to fix the worst parameter first. In an additive scheme weights can be applied in order to limit the effect of certain variables on the end result. Here, smaller values of the objective function represent a better Goodness-of-Fit (GoF).

Optimization method

In this research a grid search will be used to obtain the optimal parameter set, as it provides the researcher with more insight into the shape of the GOF surface. The disadvantage of using a grid-search that other optimization methods, e.g. Greedy, Genetic algorithms etc., can potentially be faster and find the exact global optimum. However, these methods do run the risk of getting stuck in a local minimum.

Search space definition

A basic calibration of the adopted pedestrian simulation model has already been performed. Therefore, 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 as initial tests of the implementations of the scenarios illustrated that in the case of the bidirectional high density scenario the default radius of this model produced problematic results.

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

- 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.
- 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.

As this research focusses 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.

Dealing with stochasticity in pedestrian simulation models

Similar to most pedestrian simulation models, the used simulation model 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 realisations.

In this research the required number of replications is determined using a convergence method similar to (23) 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). Eq. 9 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 smaller than or equal to $p_{threshold}$) the distribution has converged.

$$AD(S_n, S_{n-1}) \leq p_{threshold} \;\; \forall \, n \in [m - b + 1, m - b + 2, \dots, m] \tag{9}$$

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

Tests showed that regardless of the chosen values for b and $p_{threshold}$ the required number of replications depends on the exact seeds that are used and their order. Due to this finding a pre-defined seed set was used during the calibration to ensure that any differences between simulations using different parameter sets were not caused by the stochastic nature of the model. Using this pre-defined set, a value of 10 for b and a value of 0.25 for $p_{threshold}$, it was determined that the required number of replications was between 30 and 100 depending on the scenario.

BRIEF INTRODUCTION TO PEDESTRIAN DYNAMICS

This section introduces Pedestrian Dynamics (PD), a microscopic pedestrian simulation model by INCONTROL Simulation Solutions. PD offers a user the ability to model the behaviour of pedestrians at all three behavioural levels (strategic, tactical and operational). 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 modelling of the operational walking dynamics will be discussed underneath in more detail.

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 by the combination of 'social forces' with a desired velocity component. The desired velocity is determined according to the method proposed by (25). 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 destination is determined by the location of the attraction point which in turn is determined by the Indicative Route Method (26).

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 1a-d show boxplots, containing the objective values of all 3179 points of the search space, of the objective values per individual objective. These plots provide insight into how the objective values are distributed and the order of magnitude of the minimal objective value if the model would be calibrated using only a single objective. Figures 1e-j show boxplots of the non-normalised, non-squared errors and these provide insight into the size of the errors and how they are distributed.

Figure 1a and 1e show that for all scenarios the model can reproduce the flows well given both the small errors and the low minimal objective values. Figure 1b and f show that the model cannot reproduce the spatial patterns very well compared to the flows given both the higher errors and the larger minimal objective values. In the case of the effort metric figures 1c, 1g and 1i 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. Figures 1d, 1h and 1j show that, with the exception of 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.

All figures show that both the size of the minimal objective value and how the errors are distributed depend on the particular combination of scenario and metric. Furthermore, the figure shows 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).

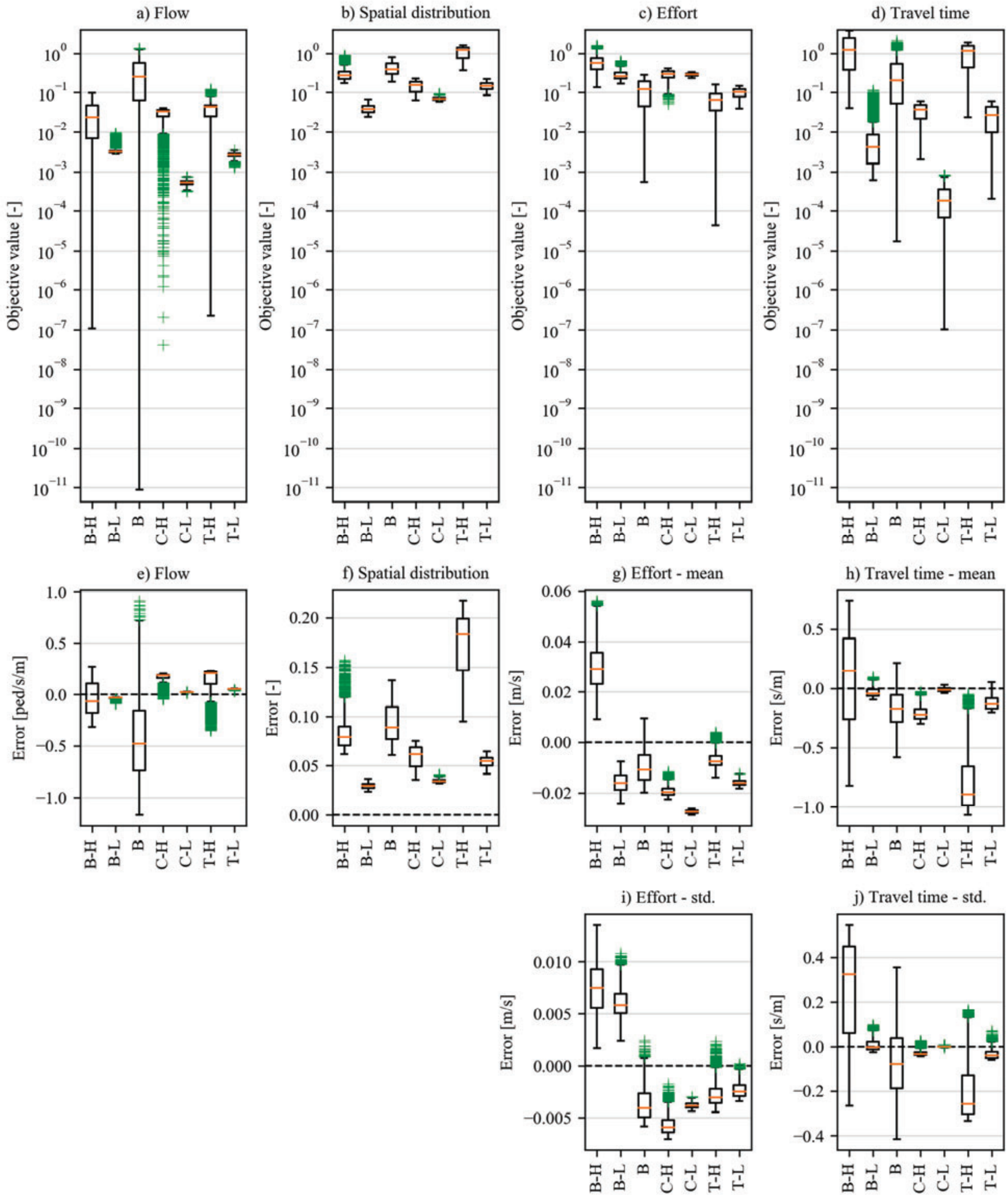

FIGURE 1 Results of calibrating the model using a single objective. Graphs a-d show, per combination of metric and scenario, how the objective values (calculated according to Eq. 8) are distributed. Graphs e-j show the non-normalized, non-squared errors (i.e. $M_{sim}(\theta) - M_{ref}$) which give insight 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, T-L = T-junction low).

DIFFERENCES IN PERFORMANCE BETWEEN CALIBRATION STRATEGIES

In this section the results of different calibration strategies will be 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 will be compared to get insight into the influence of movement base cases, density levels and metrics. Table 1 shows the 16 different strategies whereby the table indicates which scenarios and metrics are included during the calibration.

Table 2 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 2 are notably higher than those found in figure 1 already indicating that combining objective decreases the fit of the model 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. Due to time constraints it was not possible to extent the search space to obtain insight into the question whether the parameter values on the boundaries are the optimal values or that these lie beyond the current search space.

TABLE 1 Tested combination of scenarios and metrics, where 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
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
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
12. High density scenarios	x	x	x	x	x		x	x		x	
13. Low density scenarios	x	x	x	x		x			x		x
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
16. All combined	x	x	x	x	x	x	x	x	x	x	x

TABLE 2 Calibration results, where O(θ) represents the optimal value of the objective function

Calibration strategy	Combination	O(θ) [-]	Relaxation time [1/s]*	Viewing angle [degree]*	Radius [m]*
Individual scenarios – all metrics	1. Bidirectional high	0.1329	0.620	57.00	0.15296
	2. Bidirectional low	0.0588	0.620	57.00	0.19120
	3. Bottleneck	0.1093	0.395	68.25	0.20076
	4.Corner high	0.0561	0.395	57.00	0.23900
	5.Corner low	0.0742	0.380	61.50	0.23900
	6.T-junction high	0.1190	0.590	57.00	0.21988
	7. T-junction low	0.0468	0.380	68.25	0.23900
Individual scenarios – all scenarios	8. Flow	0.0146	0.380	59.25	0.20076
	9. Spatial distribution	0.2015	0.575	59.25	0.21988
	10. Effort	0.1798	0.500	57.00	0.23900
	11. Travel time	0.1814	0.620	59.25	0.15296
Combination of scenarios – all metrics	12. High density scenarios	0.2647	0.575	57.00	0.21032
	13. Low density scenarios	0.0722	0.500	57.00	0.21032
Combination of metrics – all scenarios	14. All scenarios – macro	0.1444	0.545	59.25	0.21988
	15. All scenarios – meso	0.2012	0.620	59.25	0.15296
Combination of all metrics and all scenarios	16. All combined	0.1841	0.575	57.00	0.21032

** range tested relaxation time [0.380 - 0.620], viewing angle [57 – 92] and radius [0.14340 – 0.23900]*

Identification of differences in performance between calibration procedures

In order to illustrate the differences between the optimal parameter sets, that were derived by the 16 distinct combinations, a cross-comparison of the goodness-of-fit is performed (see Tables 3, 4 and 5). 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 (Eq. 10).

$$\Delta GoF_{A;B} = -\left(O_A(\theta_B^*) - O_A(\theta_A^*)\right) \tag{10}$$

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.

Effect of movement base case on multiple-objective calibration results

Table 3 presents the results of a comparison 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 data shows 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. On average this decrease is smallest when the optimal parameter set is used that has been obtained using the combination of movement base cases. Moreover, the level of density influences the size of the decrease and the difference between the movement base cases regarding the size of the decrease in GoF. From this result it can be concluded that, in the case of low densities, the deviation of the GoF for a parameter set of one movement base case with respect to another movement base case is limited when using a single parameter set. However, this is not the case for high levels of density. The large decreases in GoF and the large differences between the movement base cases show that, in the case of a high density level, the model has difficulties predicting flows in different movement base cases well when using a single parameter set.

TABLE 3 Comparison difference in Goodness-of-fit with respect movement base case scenarios given a parameter set that is calibrated using a certain set of movement base cases, where the scenarios are identified by their acronyms (i.e. H-D = All high density scenarios, B=H = bidirectional high, B = bottleneck, C-H = corner high, T-H = T-junction high, L-D = all low density scenarios, B-L = bidirectional low, C-L = corner low, T-L = T-junction low) .

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

Effect of density level on multiple-objective calibration results

In Table 4 the results of the comparison between parameter sets that are found using 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, especially in the bidirectional and t-junction movement base cases. Moreover, in the case of the t-junction movement base case the decrease in GoF for both the low and high density levels are clearly larger than the other two movement base cases.

The data also shows 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.

TABLE 4 Comparison difference in Goodness-of-fit with respect to the density level given a parameter set that is calibrated using another density level, where the scenarios are identified by their acronyms (see Table 3).

Used parameter set	Predicted combination: B-H	C-H	T-H	B-L	C-L	T-L	H-D	L-D
H-D								-0.0044
L-D							-0.0655	
B-H				-0.0026				
B-L	-0.3149							
C-H					-0.0019			
C-L		-0.0032						
T-H						-0.0258		
T-L			-0.5869					

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

Effect of the metrics on the multiple-objective calibration results

In Table 5 a comparison is visualised between the influence of the used metrics on the performance of the resulting parameter set. 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. Besides that, both the use of the spatial distribution and the use of the distribution of the effort results 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 compared to the other metrics. When only the two macroscopic metrics are used, 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 to the individual metrics.

TABLE 5 Comparison absolute errors with respect to each metric given a parameter set that is calibrated using a certain set of metrics, where Q = flow, SD = spatial distribution, Eff = effort, TT = travel time, Macro = combination of flow and spatial distribution, Meso = combination of effort and travel time, All = combination of all four metrics

		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

DISCUSSION, CONCLUSIONS AND IMPLICATIONS FOR PRACTICE

The findings of this research regarding the influence of the movement base cases are found to be consistent with both (11) and (12). 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, it might be preferred to use a more limited set of movement base cases during the calibration due to 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 does influence the calibration results. From this it can be concluded that, again, depending in 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.

The choice of metric or combinations of metrics influence the results. Depending on the combination of metrics, also the choice of objective function and normalisation 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 these metrics when combining multiple objectives into one. Different approaches could be used to combine multiple objectives, among others normalisation methods in combination with the weighted sum method.

All in all, 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.

ACKNOWLEDGEMENTS

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.

REFERENCES

1. Rudloff, C., T. Matyus, S. Seer, and D. Bauer. Can walking behavior be predicted? Analysis of calibration and fit of pedestrian models. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2264, 2011, pp. 101-109. https://doi.org/10.3141/2264-12.
2. Seer, S., N. Brändle, and C. Ratti. Kinects and human kinetics: A new approach for studying pedestrian behavior. *Transportation research part C: emerging technologies*, Vol. 48, 2014, pp. 212-228. https://doi.org/10.1016/j.trc.2014.08.012.
3. Abdelghany, A., 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, 2016, pp. 159-176. http://dx.doi.org/10.1016/j.tra.2016.02.011.
4. Berrou, J.L., 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, ed.), Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. pp. 167-181.
5. Davidich, M. and G. Köster. Predicting Pedestrian Flow: A Methodology and a Proof of Concept Based on Real-Life Data. *PLoS ONE*, Vol. 8, No. 12, 2013, p. e83355. https://doi.org/10.1371/journal.pone.0083355.
6. Davidich, M. 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, 2012, pp. 1253-1260. https://doi.org/10.1016/j.ssci.2011.12.024.
7. Robin, T., 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, 2009, pp. 36-56. https://doi.org/10.1016/j.trb.2008.06.010.
8. Klein, W., G. Köster, and A. Meister. Towards the Calibration of Pedestrian Stream Models. In *International Conference on Parallel Processing and Applied Mathematics* (R. Wyrzykowski, J. Dongarra, K. Karczewski, and J. Wasniewski, ed.), Springer Berlin Heidelberg, Berlin, Heidelberg, 2010. pp. 521-528.
9. Campanella, M., S. Hoogendoorn, and W. Daamen. Improving the Nomad microscopic walker model. *IFAC Proceedings Volumes*, Vol. 42, No. 15, 2009, pp. 12-18. https://doi.org/10.3182/20090902-3-US-2007.0076.
10. Hoogendoorn, S.P. 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, ed.), Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. pp. 329-340.
11. Campanella, M.C., S.P. Hoogendoorn, and W. Daamen. A methodology to calibrate pedestrian walker models using multiple-objectives. In *Pedestrian and Evacuation Dynamics* (R.D. Peacock, E.D. Kuligowski, and J.D. Averill, ed.), Springer US, Boston, MA, 2011. pp. 755-759.

12. Duives, D.C. Analysis and Modelling of Pedestrian Movement Dynamics at Large-scale Events. PhD Thesis. 2016. Delft University of Technology. http://dx.doi.org/10.4233/uuid:08831f69-9b8e-44cf-8afe-f4a3e7bc9a9c
13. Duret, A., 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: Journal of the Transportation Research Board*, No. 2088, 2008, pp. 188-197. https://doi.org/10.3141/2088-20.
14. Ossen, S. and S. Hoogendoorn. Validity of Trajectory-Based Calibration Approach of Car-Following Models in Presence of Measurement Errors. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2088, 2008, pp. 117-125. https://doi.org/10.3141/2088-13.
15. Punzo, V., B. Ciuffo, and M. Montanino. Can Results of Car-Following Model Calibration Based on Trajectory Data Be Trusted? *Transportation Research Record: Journal of the Transportation Research Board*, No. 2315, 2012, pp. 11-24. https://doi.org/10.3141/2315-02.
16. Wolinski, D., 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, 2014, pp. 303-312. http://dx.doi.org/10.1111/cgf.12328.
17. Campanella, M., 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, ed.), Springer International Publishing, Cham, 2014. pp. 891-905.
18. Daamen, W. Modelling passenger flows in public transport facilities. PhD Thesis. 2004. Delft University of Technology. Available from: https://repository.tudelft.nl/islandora/search/?collection=research
19. Wong, S.C., W.L. Leung, S.H. Chan, W.H.K. Lam, N.H.C. Yung, C.Y. Liu, and P. Zhang. Bidirectional Pedestrian Stream Model with Oblique Intersecting Angle. *Journal of Transportation Engineering*, Vol. 136, No. 3, 2010, pp. 234-242. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000086.
20. Keip, C. and K. Ries. Dokumentation von Versuchen zur Personenstromdynamik - Projekt "HERMES". 2009. Bergische Universität Wuppertal. http://ped.fz-juelich.de/experiments/2009.05.12_Duesseldorf_Messe_Hermes/docu/VersuchsdokumentationHERMES.pdf. Accessed January 17, 2016.
21. Sparnaaij, M. How to calibrate a pedestrian simulation model. MSc. Thesis. 2017. Delft University of Technology. Available from: https://repository.tudelft.nl/islandora/search/?collection=education
22. Zak, S.H. and E.K.P. Chong. Multiobjective Optimization. In *An Introduction to Optimization* ed.), Wiley, Somerset, United States, 2013. pp. 577-598.
23. Ronchi, E., 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. NIST, U.S. Department of Commerce, 2013
24. Anderson, T.W. and D.A. Darling. Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes. *Ann. Math. Statist.*, Vol. 23, No. 2, 1952, pp. 193-212. https://doi.org/10.1214/aoms/1177729437.
25. Moussaïd, M., D. Helbing, and G. Theraulaz. How simple rules determine pedestrian behavior and crowd disasters. *Proceedings of the National Academy of Sciences*, Vol. 108, No. 17, 2011, pp. 6884-6888. https://doi.org/10.1073/pnas.1016507108.
26. Karamouzas, I., R. Geraerts, and M. Overmars. Indicative routes for path planning and crowd simulation.In *Proceedings of the 4th International Conference on Foundations of Digital Games*, ACM, Orlando, Florida, 2009. pp. 113-120.