

**Multi-objective calibration framework for pedestrian simulation models
study on the effect of movement base cases, metrics and density levels**

Sparnaaij, Martijn; Duives, Dorine; Knoop, Victor; Hoogendoorn, Serge

Publication date

2018

Document Version

Accepted author manuscript

Published in

97th Annual meeting of the Transportation Research board

Citation (APA)

Sparnaaij, M., Duives, D., Knoop, V., & Hoogendoorn, S. (2018). Multi-objective calibration framework for pedestrian simulation models: study on the effect of movement base cases, metrics and density levels. In *97th Annual meeting of the Transportation Research board* Article 18-02495.

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

1
2
3
4
5
6
7 **MULTI-OBJECTIVE CALIBRATION FRAMEWORK FOR PEDESTRIAN**
8 **SIMULATION MODELS**
9 **- Study on the Effect of Movement Base Cases, Metrics and Density Levels -**
10

11
12
13
14
15
16 Martijn Sparnaaij
17 Stevinweg 1, 2628 CN, Delft, The Netherlands
18 m.sparnaaij@tudelft.nl
19

20 Dr. Dorine C. Duives (Corresponding author)
21 Stevinweg 1, 2628 CN, Delft, The Netherlands
22 d.c.duives@tudelft.nl
23 +31 (15) 278 6304
24

25 Dr. Victor Knoop
26 Stevinweg 1, 2628 CN, Delft, The Netherlands
27 v.knoop@tudelft.nl
28 +31 (15) 278 8413
29

30 Prof. Serge P. Hoogendoorn
31 Stevinweg 1, 2628 CN, Delft, The Netherlands
32 s.p.hoogendoorn@tudelft.nl
33 +31 (15) 278 5475
34

35
36
37 *1 Augustus 2017*
38

39 Keywords:
40 Pedestrian simulation models, Multi-objective calibration, Effect study,
41 Movement base case, Metrics, Density levels
42

43
44 Submitted at TRB 2018:
45 6048 words + 1 figures * 250 words + 5 tables * 250 words
46 = 7548 words

1 **ABSTRACT**

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

19

20 **INTRODUCTION**

21 Ideally, a multitude of steps has to be taken before a commercial implementation of a pedestrian
22 model is used in practice. Within this process, calibration, the main goal of which is to increase
23 the accuracy of the predictions by determining the set of values for the model parameters that
24 allows for the best replication of reality, has an important role.

25 Yet, up to recently, calibration has received relatively little attention within the field of
26 pedestrian modelling (1, 2). This is mainly attributed to the lack of data (1, 3-5) especially at
27 high densities. Despite this issue, there are many studies in which authors calibrate a pedestrian
28 model (e.g. 6-8) usually by using the fundamental diagram (9). However, as multiple authors
29 mention, the calibration attempts in these studies are limited and mostly focus on only one or a
30 few aspects (1, 4, 5, 9, 10). Most studies focus on one specific movement base case (e.g. a
31 bidirectional flow in a straight corridor), only use a single metric or do not look at various
32 compositions of the population.

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

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

1 uses multiple movement base cases with multiple metrics and furthermore includes different
2 combinations of weights in the objective function and is thus the most extensive of the three.

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

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

15 16 **METHODOLOGY**

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

21 22 **Scenarios**

23 Research by both (11) and (12) has shown that using different base cases during calibration
24 results in different optimal parameter sets and hence both studies show the importance of using
25 multiple bases cases when calibrating a pedestrian model for general usage. This is further
26 supported by the findings in (17) which show that models perform better in a general setting
27 when they are calibrated using multiple base cases.

28 Contemporary, several data sets are available that feature the movement of pedestrians in
29 multiple movement base cases and a similar population of pedestrians, among others (18-20).
30 Since the experiments within the HERMES project represent the most comprehensive set of
31 movement base cases featuring a similar population and different levels of density, this dataset
32 will be used in this calibration procedure. Based on this data set seven scenarios are constructed
33 whereby every scenario contains a single movement base case and a single density level. Four
34 movement base cases are studied, these are a unidirectional corner flow, a merging flow, a
35 bidirectional flow and a bottleneck flow. All base cases have both a low and high density variant
36 except for the bottleneck which only has a high density variant. For a more detailed overview of
37 the experimental setup within the HERMES project the reader is referred to (20). Care is taken to
38 ensure a similar flow pattern over time, speed distribution and route choice, details on the exact
39 simulation of the seven scenarios are mentioned in (21).

40 41 **Metrics**

42 In this multiple-objective framework four different metrics are used to identify how different
43 metrics impacts the calibration results. In this research the choice is made to use two metrics at
44 the macroscopic level, the flow and the spatial distribution, and two at the mesoscopic level, the
45 effort distribution and the travel time distribution. Microscopic metrics, i.e. trajectories, are not
46 used for three reasons. Firstly, calibration based on trajectories requires a different approach than

1 calibrating based on macro and mesoscopic metrics. Secondly, the current approaches for
 2 calibrating based on trajectories do not deal with the stochastic nature of the model. Lastly, since
 3 pedestrian simulation models are mostly used to approximate the macroscopic properties of the
 4 infrastructure (e.g. capacity, density distribution) (17) and given that calibrating based on
 5 microscopic metrics does not necessarily result in a macroscopically valid model (9)
 6 macroscopic and mesoscopic metrics take priority over microscopic metrics.

8 *Flow*

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

$$13 \quad \bar{q}_i = \frac{N_i}{\Delta t * l} \quad (1)$$

14 where N_i is the number of unique pedestrians with main travel direction i that passed the line
 15 in the direction equal to the main travel direction and during the measurement period (Δt). The
 16 flow is normalised to a flow per meter of measurement line whereby l is the length of the
 17 measurement line in order to allow for comparisons between scenarios.

18 *Distribution over space*

19 (12) Showed that microscopic models might not always be able to accurately reproduce the
 20 spatial distribution patterns. Hence, it is essential to check whether this model performs well with
 21 respect to this property. The distribution over space measures how the pedestrians are distributed
 22 over the measurement area. A grid of 0.4 x 0.4 m, which is approximately the size of one
 23 pedestrian during a high density situation, overlays the measurement area and for every cell the
 24 percentage of the time it is occupied is determined in Eq. 2.

$$25 \quad F_j = \frac{N_{occ;j}}{N_{steps}} \quad (2)$$

26 where $N_{occ;j}$ is the number of time steps cell j is occupied by one or more pedestrians (based on
 27 the centre point of the pedestrians) and N_{steps} is the number of time steps taken into account.

28 *Effort*

29 Several studies have identified the difficulty of smooth interactions between simulated
 30 pedestrians in bidirectional flows. In order to ensure realistic interaction behaviour the effort
 31 metric is introduced, which captures how much effort it takes a pedestrian to traverse the
 32 measurement area. The effort for pedestrian k is defined as the average change in velocity per
 33 time step (see Eq. 3).
 34
 35

$$36 \quad e_k = \frac{\sum_{n=1}^{n-1} (|v_{l;x} - v_{l-1;x}| + |v_{l;y} - v_{l-1;y}|)}{n-1} \quad (3)$$

$$37 \quad v_{l;x} = \frac{x_l - x_{l-1}}{\Delta t} \quad (4)$$

38 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
 39 number of time steps. The speeds are obtained by differentiating the positions (Eq. 4), where x_l
 40 is the x-position at time step l and Δt is the duration of the time step. The effort measurements of
 41 all pedestrians are combined into a distribution.
 42

1 *Travel time*

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

3
$$TT_k = \frac{t_{end} - t_{start}}{l_{ref}} \quad (5)$$

4 where t_{start} and t_{end} are respectively the time the pedestrian first entered the measurement area
 5 and time the pedestrian left the area. l_{ref} is the average length of the path in the measurement
 6 area, as obtained from the reference data. The travel time is normalised in order to simplify the
 7 comparison between different scenario with different average path lengths. Note that this metric
 8 approximates the realized pace of each individual. That is, an individual who makes a detour at a
 9 very high speed does not influence the travel time, but will influence the effort metric.

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

12
 13 **Objectives**

14 In this research multiple objectives are combined into a single objective using the weighted sum
 15 method (22). This is in line with research by (12), the only example in literature using both
 16 multiple metrics and scenarios to calibrate a pedestrian model.

17 In order to make a fair comparison between objectives, normalisation is necessary, as the
 18 metrics have different units and different orders of magnitude. The adopted normalisation
 19 method uses a single normalization value per metric whereby this normalization value is
 20 determined based on the ratios between the metric values obtained from the reference data. For a
 21 detailed explanation of this method and an underpinning of the choice to specifically use this
 22 method, the reader is referred to (21). The objective function for a given metric and scenario is
 23 given by the normalised Squared Error (SE) for the macroscopic metrics (Eq. 6) and the
 24 mesoscopic metrics (Eq. 7).

25
$$SE_{norm}(\theta) = \frac{1}{m} \sum_j \left(\frac{\frac{\sum_i M_{sim;i,j}(\theta) - M_{ref;j}}{n}}{M_{norm}} \right)^2 \quad (6)$$

26
$$SE_{norm;meso}(\theta) = \frac{1}{2} \left(\frac{M_{sim;\mu}(\theta) - M_{ref;\mu}}{M_{norm;\mu}} \right)^2 + \frac{1}{2} \left(\frac{M_{sim;\sigma}(\theta) - M_{ref;\sigma}}{M_{norm;\sigma}} \right)^2 \quad (7)$$

27
 28 where M_{sim} is the metrics value according to the simulation, M_{ref} the reference value according
 29 to the data, M_{norm} the value used for the normalisation and θ the vector of model parameters. In
 30 the case Eq. 6 n is the number of replications and m is the number of travel directions in case of
 31 the flow and the number of cells in case of the spatial distribution. In the case of the mesoscopic
 32 metrics Eq. 7 shows that the difference between the distributions is approximated by taking both
 33 the error in the mean (μ) and the standard deviation (σ). These distributions contain the
 34 measurements of all replications.

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

37
$$O(\theta) = \frac{1}{N_s * N_m} \sum_s \sum_m SE_{norm;s;m}(\theta) \quad (8)$$

38 where $SE_{norm;s;m}(\theta)$ is the value of the objective function of scenarios s and metric m for the
 39 parameter set θ and N_s and N_m are, respectively, the number of scenarios and metrics in the set.
 40 A likelihood method, which multiplies probabilities, might not work in this case, as the method
 41 will always attempt to fix the worst parameter first. In an additive scheme weights can be applied

1 in order to limit the effect of certain variables on the end result. Here, smaller values of the
2 objective function represent a better Goodness-of-Fit (GoF).

3 4 **Optimization method**

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

10 11 **Search space definition**

12 A basic calibration of the adopted pedestrian simulation model has already been performed.
13 Therefore, the presented calibration method will be used in this research to identify the
14 correctness of the variables with respect to which this model is most sensitive, namely the
15 relaxation time and the viewing angle. Even though the model is less sensitive with respect to the
16 radius, this parameter will also be included as initial tests of the implementations of the scenarios
17 illustrated that in the case of the bidirectional high density scenario the default radius of this
18 model produced problematic results.

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

- 20 • The upper and lower boundaries of the relaxation time and viewing angle are determined
21 by a deviation of $-0.24 * \theta_{\text{def}} < \theta < 0.24 * \theta_{\text{def}}$ with respect to the default parameters.
22 The step size is 3% of the default value.
- 23 • For the radius the upper boundary is equal to the default value, the lower bound has a
24 deviation of $-0.40 * \theta_{\text{def}}$ and the step size is 4% of the default value.

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

29 30 **Dealing with stochasticity in pedestrian simulation models**

31 Similar to most pedestrian simulation models, the used simulation model is stochastic in nature.
32 Therefore, it is essential to determine the minimum amount of replications one would need in
33 order to assure that statistical differences are due to differences in model parameters instead of
34 stochasticity in the model realisations.

35 In this research the required number of replications is determined using a convergence method
36 similar to (23) whereby the distribution of speeds is used as the sole metric. To determine if two
37 subsequent distributions can be considered to be samples drawn from the same distribution the
38 Anderson-Darling test is used (24). Eq. 9 shows that if b subsequent distributions are considered
39 to be similar according to the Anderson-Darling test (i.e. the test return a p-value smaller than or
40 equal to $p_{\text{threshold}}$) the distribution has converged.

$$41 \quad AD(S_n, S_{n-1}) \leq p_{\text{threshold}} \quad \forall n \in [m - b + 1, m - b + 2, \dots, m] \quad (9)$$

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

44 Tests showed that regardless of the chosen values for b and $p_{\text{threshold}}$ the required number of
45 replications depends on the exact seeds that are used and their order. Due to this finding a pre-
46 defined seed set was used during the calibration to ensure that any differences between

1 simulations using different parameter sets were not caused by the stochastic nature of the model.
2 Using this pre-defined set, a value of 10 for b and a value of 0.25 for $p_{threshold}$, it was
3 determined that the required number of replications was between 30 and 100 depending on the
4 scenario.

5 6 **BRIEF INTRODUCTION TO PEDESTRIAN DYNAMICS**

7 This section introduces Pedestrian Dynamics (PD), a microscopic pedestrian simulation model by
8 INCONTROL Simulation Solutions. PD offers a user the ability to model the behaviour of
9 pedestrians at all three behavioural levels (strategic, tactical and operational). In the case of this
10 research the pedestrians only have one activity, walking from their origin to their destination, and
11 hence there is no need to model the activity choice or the scheduling. The modelling of the
12 operational walking dynamics will be discussed underneath in more detail.

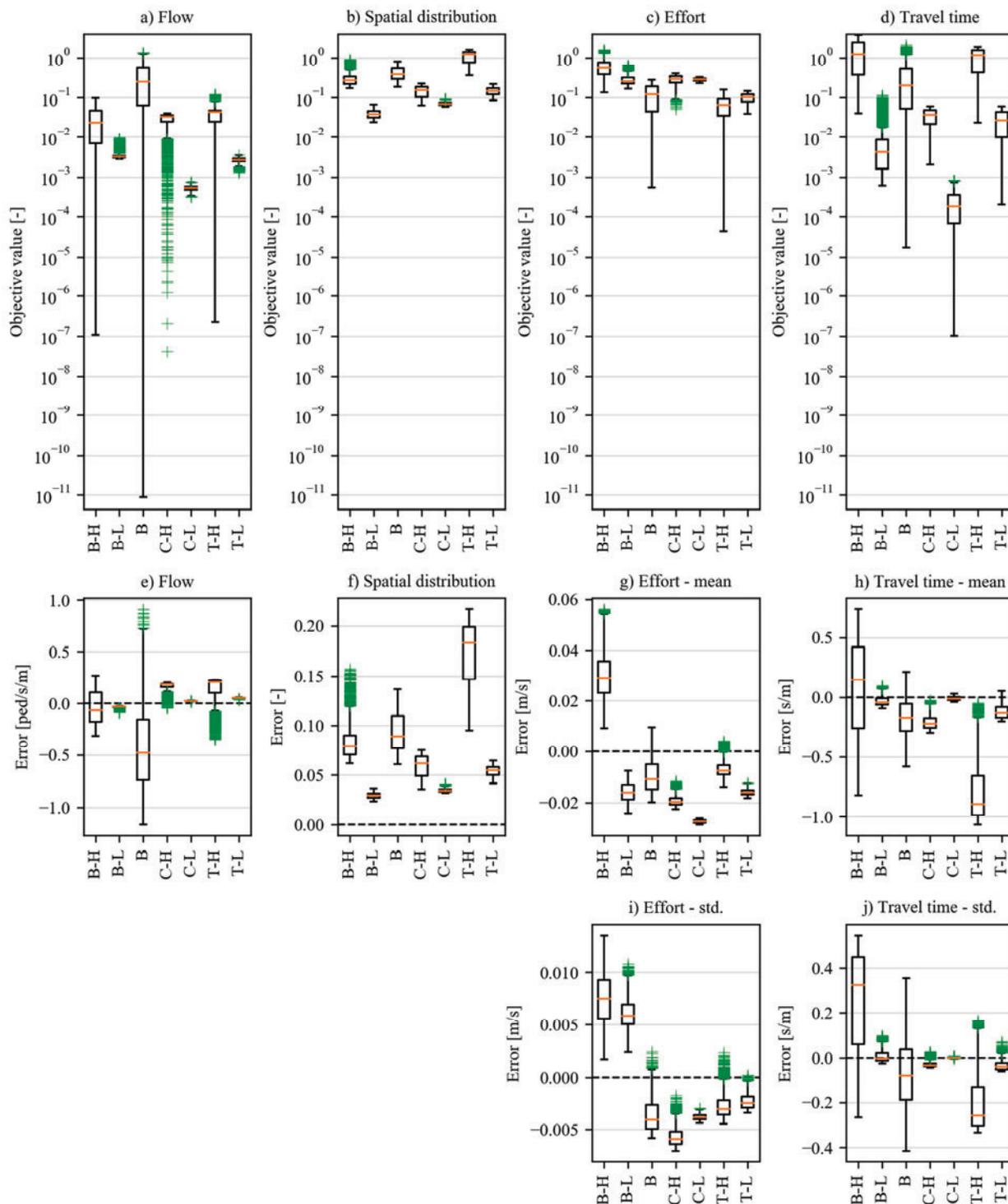
13 The operational behaviour consists of two parts, route following and collision avoidance,
14 which together determine the acceleration of a pedestrian at every time step. In PD the
15 acceleration of an agent is determined by the combination of ‘social forces’ with a desired
16 velocity component. The desired velocity is determined according to the method proposed by
17 (25). The method uses a vision based approach to avoid collisions and combines the collision
18 avoidance with the preferred speed and the desired destination to determine the desired velocity.
19 The desired velocity is a combination of the desired speed and the desired direction. The desired
20 destination is determined by the location of the attraction point which in turn is determined by
21 the Indicative Route Method (26).

22 23 **CALIBRATION RESULTS BASED ON SINGLE OBJECTIVES**

24 In this section the results of the individual objectives (a combination of a single scenario and a
25 single metric) are discussed. Figures 1a-d show boxplots, containing the objective values of all
26 3179 points of the search space, of the objective values per individual objective. These plots
27 provide insight into how the objective values are distributed and the order of magnitude of the
28 minimal objective value if the model would be calibrated using only a single objective. Figures
29 1e-j show boxplots of the non-normalised, non-squared errors and these provide insight into the
30 size of the errors and how they are distributed.

31 Figure 1a and 1e show that for all scenarios the model can reproduce the flows well given both
32 the small errors and the low minimal objective values. Figure 1b and f show that the model
33 cannot reproduce the spatial patterns very well compared to the flows given both the higher
34 errors and the larger minimal objective values. In the case of the effort metric figures 1c, 1g and
35 1i show that for most scenario the model cannot reproduce the effort distribution very well. The
36 two exceptions are the bottleneck and t-junction high density scenarios. In these cases the model
37 can reproduce the effort distributions well. Figures 1d, 1h and 1j show that, with the exception of
38 the bidirectional high and t-junction high density scenarios the travel time distribution can be
39 reproduced well by the model. In the case of the bidirectional high and t-junction high density
40 scenarios the figures show that the model can reproduce the mean and the standard deviation of
41 the travel time distribution well individually but apparently not when they are combined.

42 All figures show that both the size of the minimal objective value and how the errors are
43 distributed depend on the particular combination of scenario and metric. Furthermore, the figure
44 shows that the model can generally reproduce the metrics related to the performance of the
45 infrastructure (the flow and travel time) better than those more related to the underlying
46 microscopic and macroscopic pedestrian dynamics (spatial distribution and the effort).



1

2 **FIGURE 1** Results of calibrating the model using a single objective. Graphs a-d show, per combination of
 3 metric and scenario, how the objective values (calculated according to Eq. 8) are distributed. Graphs e-j show
 4 the non-normalized, non-squared errors (i.e. $M_{sim}(\theta) - M_{ref}$) which give insight into size and the
 5 distribution of the errors. The flow scenarios are identified by their acronyms (i.e. B-H = bidirectional high,
 6 B-L = bidirectional low, B = bottleneck, C-H = corner high, C-L = corner low, T-H = T-junction high, T-L =
 7 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

24
25

1 **TABLE 2 Calibration results, where $O(\theta)$ represents the optimal value of the objective function**

Calibration strategy	Combination	$O(\theta)$ [-]	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]

2
3
4 **Identification of differences in performance between calibration procedures**

5 In order to illustrate the differences between the optimal parameter sets, that were derived by the
6 16 distinct combinations, a cross-comparison of the goodness-of-fit is performed (see Tables 3, 4
7 and 5). These comparisons are based on the difference between the optimal GoF of combination
8 A and the GoF of combination A when the optimal parameter set of combination B is used (Eq.
9 10).

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

11 where $O_A(\theta_A^*)$ is the value of the objective function of combination A when its optimal parameter
12 set θ_A^* is used. $O_A(\theta_B^*)$ is the value of the objective function of combination A if the optimal
13 parameter set of the combination B is used.

14
15 **Effect of movement base case on multiple-objective calibration results**

16 Table 3 presents the results of a comparison between different calibration strategies, in which the
17 difference in goodness-of-fit is depicted. All comparisons are made between (combinations of)
18 scenarios of the same density level, in order to exclude the possibility that differences are caused
19 by a difference in the level of density and not by a difference in movement base case.

20 The data shows that in all cases the GoF of the individual movement base cases decreases
21 when the parameter set based on another movement base case or a set of movement base cases is
22 used. On average this decrease is smallest when the optimal parameter set is used that has been
23 obtained using the combination of movement base cases. Moreover, the level of density
24 influences the size of the decrease and the difference between the movement base cases
25 regarding the size of the decrease in GoF. From this result it can be concluded that, in the case of
26 low densities, the deviation of the GoF for a parameter set of one movement base case with
27 respect to another movement base case is limited when using a single parameter set. However,
28 this is not the case for high levels of density. The large decreases in GoF and the large
29 differences between the movement base cases show that, in the case of a high density level, the

1 model has difficulties predicting flows in different movement base cases well when using a
 2 single parameter set.

3

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

		Predicted combination						
		B-H	B	C-H	T-H	B-L	C-L	T-L
Used parameter set	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

9

10 **Effect of density level on multiple-objective calibration results**

11 In Table 4 the results of the comparison between parameter sets that are found using different
 12 density levels are presented. The data shows that in all three cases the decrease in the GoF is
 13 smaller when the optimal parameter set of the high density case is used in the low density case
 14 than vice versa, especially in the bidirectional and t-junction movement base cases. Moreover, in
 15 the case of the t-junction movement base case the decrease in GoF for both the low and high
 16 density levels are clearly larger than the other two movement base cases.

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

22

23 **TABLE 4 Comparison difference in Goodness-of-fit with respect to the density level given a parameter set**
 24 **that is calibrated using another density level, where the scenarios are identified by their acronyms (see Table**
 25 **3).**

		Predicted combination							
		B-H	C-H	T-H	B-L	C-L	T-L	H-D	L-D
Used parameter set	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					

26

27 Overall, it can be concluded that the level of density of the scenario does influence the
 28 calibration results. Therefore, it is concluded that it is more important to include the high density
 29 scenarios than the low density scenarios

30

1 **Effect of the metrics on the multiple-objective calibration results**

2 In Table 5 a comparison is visualised between the influence of the used metrics on the
 3 performance of the resulting parameter set. There seems to be a correlation between the
 4 distribution of the effort and the spatial distribution. When the model is calibrated using only one
 5 of them, the decrease in the GoF of the other is small. Besides that, both the use of the spatial
 6 distribution and the use of the distribution of the effort results in a far worse prediction of the
 7 flow compared to the distribution of the travel times. That is, the decrease in GoF of the flow is
 8 far larger in case the optimal parameter set of the spatial distribution or the use of the distribution
 9 of the effort is used. Lastly, the optimal parameter sets obtained using combinations of metrics
 10 are more heavily influenced by certain metrics compared to the other metrics. When only the two
 11 macroscopic metrics are used, the spatial distribution clearly has a larger impact on the location
 12 of the optimal parameter set given the lower decrease in GoF. When solely using the mesoscopic
 13 metrics, the distribution of the travel time has a larger impact compared to the distribution of the
 14 effort.

15 These results show that the choice of metrics does influence the results of the calibration.
 16 Depending on the choice of metric or combination of metrics, different optimal parameter sets
 17 are found which in turn lead to different results regarding the GoF to the individual metrics.

18
 19 **TABLE 5 Comparison absolute errors with respect to each metric given a parameter set that is calibrated**
 20 **using a certain set of metrics, where Q = flow, SD = spatial distribution, Eff = effort, TT = travel time, Macro**
 21 **= combination of flow and spatial distribution, Meso = combination of effort and travel time, All =**
 22 **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

23 **DISCUSSION, CONCLUSIONS AND IMPLICATIONS FOR PRACTICE**

24 The findings of this research regarding the influence of the movement base cases are found to be
 25 consistent with both (11) and (12). Similar to those studies, this research finds that 1) it is
 26 necessary to use multiple movement base cases, when calibrating a model, to capture all relevant
 27 behaviours and 2) the GoF of the individual movement base cases decreases when the parameter
 28 set based on multiple movement base cases is used.

29 Hence, this research confirms that one needs to use multiple movement base cases when
 30 calibrating a model intended for general usage. However, when the intended use of the model is
 31 more limited, it might be preferred to use a more limited set of movement base cases during the
 32 calibration due to the fact that the GoF of the individual movement base case decreases when
 33 multiple movement base cases are used during the calibration.

34 The level of density does influence the calibration results. From this it can be concluded that,
 35 again, depending in the intended use of the model different density levels should be taken into
 36 account during the calibration. Furthermore, as the results show, it is far more important to take
 37 the higher levels of density into account.
 38

1 The choice of metric or combinations of metrics influence the results. Depending on the
2 combination of metrics, also the choice of objective function and normalisation method
3 influences the results. Consequently, depending on the usage of the model, one should decide
4 which metric or metrics are most important, and how to reflect these metrics when combining
5 multiple objectives into one. Different approaches could be used to combine multiple objectives,
6 among others normalisation methods in combination with the weighted sum method.

7 All in all, the main implication of the results for practice is that the intended use of the model
8 should be taken into account when deciding which scenarios, metrics, objective functions and
9 method for combining multiple objectives one should use.

10 11 **ACKNOWLEDGEMENTS**

12 The research presented in this paper is part of the research program “Allegro: Unravelling slow
13 mode traveling and traffic: with innovative data to a new transportation and traffic theory for
14 pedestrians and bicycles” (ERC Grant Agreement no. 669792), a Horizon 2020 project which is
15 funded by the European Research Council.

16 17 **REFERENCES**

- 18 1. Rudloff, C., T. Matyus, S. Seer, and D. Bauer. Can walking behavior be predicted? Analysis of calibration
19 and fit of pedestrian models. *Transportation Research Record: Journal of the Transportation Research*
20 *Board*, No. 2264, 2011, pp. 101-109. <https://doi.org/10.3141/2264-12>.
- 21 2. Seer, S., N. Brändle, and C. Ratti. Kinects and human kinetics: A new approach for studying pedestrian
22 behavior. *Transportation research part C: emerging technologies*, Vol. 48, 2014, pp. 212-228.
23 <https://doi.org/10.1016/j.trc.2014.08.012>.
- 24 3. Abdelghany, A., K. Abdelghany, and H. Mahmassani. A hybrid simulation-assignment modeling
25 framework for crowd dynamics in large-scale pedestrian facilities. *Transportation Research Part A: Policy*
26 *and Practice*, Vol. 86, 2016, pp. 159-176. <http://dx.doi.org/10.1016/j.tra.2016.02.011>.
- 27 4. Berrou, J.L., J. Beecham, P. Quaglia, M.A. Kagarlis, and A. Gerodimos. Calibration and validation of the
28 Legion simulation model using empirical data. In *Pedestrian and Evacuation Dynamics 2005* (N. Waldau,
29 P. Gattermann, H. Knoflacher, and M. Schreckenberg, ed.), Springer Berlin Heidelberg, Berlin, Heidelberg,
30 2007. pp. 167-181.
- 31 5. Davidich, M. and G. Köster. Predicting Pedestrian Flow: A Methodology and a Proof of Concept Based on
32 Real-Life Data. *PLoS ONE*, Vol. 8, No. 12, 2013, p. e83355. <https://doi.org/10.1371/journal.pone.0083355>.
- 33 6. Davidich, M. and G. Köster. Towards automatic and robust adjustment of human behavioral parameters in
34 a pedestrian stream model to measured data. *Safety science*, Vol. 50, No. 5, 2012, pp. 1253-1260.
35 <https://doi.org/10.1016/j.ssci.2011.12.024>.
- 36 7. Robin, T., G. Antonini, M. Bierlaire, and J. Cruz. Specification, estimation and validation of a pedestrian
37 walking behavior model. *Transportation Research Part B: Methodological*, Vol. 43, No. 1, 2009, pp. 36-
38 56. <https://doi.org/10.1016/j.trb.2008.06.010>.
- 39 8. Klein, W., G. Köster, and A. Meister. Towards the Calibration of Pedestrian Stream Models. In
40 *International Conference on Parallel Processing and Applied Mathematics* (R. Wyrzykowski, J. Dongarra,
41 K. Karczewski, and J. Wasniewski, ed.), Springer Berlin Heidelberg, Berlin, Heidelberg, 2010. pp. 521-
42 528.
- 43 9. Campanella, M., S. Hoogendoorn, and W. Daamen. Improving the Nomad microscopic walker model.
44 *IFAC Proceedings Volumes*, Vol. 42, No. 15, 2009, pp. 12-18. [https://doi.org/10.3182/20090902-3-US-](https://doi.org/10.3182/20090902-3-US-2007.0076)
45 [2007.0076](https://doi.org/10.3182/20090902-3-US-2007.0076).
- 46 10. Hoogendoorn, S.P. and W. Daamen. Microscopic calibration and validation of pedestrian models: Cross-
47 comparison of models using experimental data. In *Traffic and Granular Flow'05* (A. Schadschneider, T.
48 Pöschel, R. Kühne, M. Schreckenberg, and D.E. Wolf, ed.), Springer Berlin Heidelberg, Berlin,
49 Heidelberg, 2007. pp. 329-340.
- 50 11. Campanella, M.C., S.P. Hoogendoorn, and W. Daamen. A methodology to calibrate pedestrian walker
51 models using multiple-objectives. In *Pedestrian and Evacuation Dynamics* (R.D. Peacock, E.D.
52 Kuliowski, and J.D. Averill, ed.), Springer US, Boston, MA, 2011. pp. 755-759.

- 1 12. Duives, D.C. Analysis and Modelling of Pedestrian Movement Dynamics at Large-scale Events. PhD
2 Thesis. 2016. Delft University of Technology. <http://dx.doi.org/10.4233/uuid:08831f69-9b8e-44cf-8afe->
3 [f4a3e7bc9a9c](http://dx.doi.org/10.4233/uuid:08831f69-9b8e-44cf-8afe-f4a3e7bc9a9c)
- 4 13. Duret, A., C. Buisson, and N. Chiabaut. Estimating Individual Speed-Spacing Relationship and Assessing
5 Ability of Newell's Car-Following Model to Reproduce Trajectories. *Transportation Research Record:*
6 *Journal of the Transportation Research Board*, No. 2088, 2008, pp. 188-197. <https://doi.org/10.3141/2088->
7 20.
- 8 14. Ossen, S. and S. Hoogendoorn. Validity of Trajectory-Based Calibration Approach of Car-Following
9 Models in Presence of Measurement Errors. *Transportation Research Record: Journal of the*
10 *Transportation Research Board*, No. 2088, 2008, pp. 117-125. <https://doi.org/10.3141/2088-13>.
- 11 15. Punzo, V., B. Ciuffo, and M. Montanino. Can Results of Car-Following Model Calibration Based on
12 Trajectory Data Be Trusted? *Transportation Research Record: Journal of the Transportation Research*
13 *Board*, No. 2315, 2012, pp. 11-24. <https://doi.org/10.3141/2315-02>.
- 14 16. Wolinski, D., S. J Guy, A.H. Olivier, M. Lin, D. Manocha, and J. Pettré. Parameter estimation and
15 comparative evaluation of crowd simulations. *Computer Graphics Forum*, Vol. 33, No. 2, 2014, pp. 303-
16 312. <http://dx.doi.org/10.1111/cgf.12328>.
- 17 17. Campanella, M., S. Hoogendoorn, and W. Daamen. Quantitative and qualitative validation procedure for
18 general use of pedestrian models. In *Pedestrian and Evacuation Dynamics 2012* (U. Weidmann, U. Kirsch,
19 and M. Schreckenberg, ed.), Springer International Publishing, Cham, 2014. pp. 891-905.
- 20 18. Daamen, W. Modelling passenger flows in public transport facilities. PhD Thesis. 2004. Delft University of
21 Technology. Available from: <https://repository.tudelft.nl/islandora/search/?collection=research>
- 22 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
23 Pedestrian Stream Model with Oblique Intersecting Angle. *Journal of Transportation Engineering*, Vol.
24 136, No. 3, 2010, pp. 234-242. [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000086](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000086).
- 25 20. Keip, C. and K. Ries. Dokumentation von Versuchen zur Personenstromdynamik - Projekt "HERMES".
26 2009. Bergische Universität Wuppertal. <http://ped.fz->
27 [juelich.de/experiments/2009.05.12_Duesseldorf_Messe_Hermes/docu/VersuchsdokumentationHERMES.p](http://ped.fz-juelich.de/experiments/2009.05.12_Duesseldorf_Messe_Hermes/docu/VersuchsdokumentationHERMES.p)
28 [df](http://ped.fz-juelich.de/experiments/2009.05.12_Duesseldorf_Messe_Hermes/docu/VersuchsdokumentationHERMES.p). Accessed January 17, 2016.
- 29 21. Sparnaaij, M. How to calibrate a pedestrian simulation model. MSc. Thesis. 2017. Delft University of
30 Technology. Available from: <https://repository.tudelft.nl/islandora/search/?collection=education>
- 31 22. Zak, S.H. and E.K.P. Chong. Multiobjective Optimization. In *An Introduction to Optimization* ed.), Wiley,
32 Somerset, United States, 2013. pp. 577-598.
- 33 23. Ronchi, E., E.D. Kuligowski, P.A. Reneke, R.D. Peacock, and D. Nilsson. *The Process of Verification and*
34 *Validation of Building Fire Evacuation Models*. Technical Note (NIST TN) - 1822. NIST, U.S. Department
35 of Commerce, 2013
- 36 24. Anderson, T.W. and D.A. Darling. Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on
37 Stochastic Processes. *Ann. Math. Statist.*, Vol. 23, No. 2, 1952, pp. 193-212.
38 <https://doi.org/10.1214/aoms/1177729437>.
- 39 25. Moussaïd, M., D. Helbing, and G. Theraulaz. How simple rules determine pedestrian behavior and crowd
40 disasters. *Proceedings of the National Academy of Sciences*, Vol. 108, No. 17, 2011, pp. 6884-6888.
41 <https://doi.org/10.1073/pnas.1016507108>.
- 42 26. Karamouzas, I., R. Geraerts, and M. Overmars. Indicative routes for path planning and crowd simulation. In
43 *Proceedings of the 4th International Conference on Foundations of Digital Games*, ACM, Orlando,
44 Florida, 2009. pp. 113-120.

45