

CALIBRATION OF PEDESTRIAN SIZES IN DECISION-BASED MODELLING

Jana Vacková, Marek Bukáček

FNSPE, Czech Technical University in Prague
Trojanova 13, Prague, Czech Republic

e-mail: janca.vackova@fjfi.cvut.cz

ABSTRACT

Calibration of any model is the crucial part of pedestrian movement prediction. Hence a suitable calibration approach is needed. This paper deals with an author's microscopic decision-based model and a general calibration concept which is focused on the phase of congestion. Dense crowd behaviour is modelled by two parameters describing the agent size, namely social and physical pedestrian sizes. The physical (minimum) pedestrian size is considered to be known – it can be estimated from the real data using distance of human shoulders. However, the social size can vary during the time with respect to the surroundings of the pedestrian. Furthermore, the size estimated by a general metric may significantly differ to the individual pedestrian experience. In other words, pedestrians decide about their own (social) compression in a crowd. Therefore, the calibration process has to be based on sophisticated methods considering individual pedestrian behaviour rather than simple macroscopic quantities. Thus, this calibration study shows how pedestrian trajectories and derived microscopic quantities (especially local density) can be applied. An experimental data of passing through a room with one exit are used.

INTRODUCTION

As the population grows, the pedestrian dynamics research is not only more trendy, but also more needed now than in the last century [11]. To predict the pedestrian movement during evacuations or any other kind of crowd situations, the pedestrian models are used. Individual data analysis from egress experiments stands beside the modelling, but it is necessary as well.

Having a model with defined rules and principles, its parameters can be set up by physical laws. However, this setting does not have to be the one which produces a real system. The aim of the calibration process is to find the eligible parameter values and make the model replicate the reality accurately [14]. Calibration of models is not important only in pedestrian models, it is a general topic [3].

Nevertheless, there is not a single correct way to calibrate the model - there is no universally right method [3]. Different methods are used according to the type of the model, its following application and, of course, the author's preferences. We can apply heterogeneity (especially pedestrian type) in calibration [5], maximum likelihood estimation under assumption that an error (between model and data) is normally distributed [9], macroscopic quantities (fundamental diagram) [13] or individual pedestrian trajectory data [7]. The choice of the quantity, which represents the data properties, and the objective function or method, which describes the error between model and real data, is very difficult. To cover more than one metric, multiobjective calibration was studied [4].

In the last decades, statistical methods used for calibration became more and more popular [1, 16]. However, the most used method is still maximum likelihood estimation [10] which brings only the point estimate of a model parameter. Few publications taking into account a distribution of a parameter in question arose recently, e.g. [6], and showed that using statistical methods is very promising approach in model calibration.

Having designed and developed decision-based model continuous in space [18], the calibration process can be started. Our calibration concept consists of separate calibration episodes. We want to avoid choosing just a few metrics and quantities to describe the whole, complex system and

rising problems with finding the global optimum. Thus, we design the calibration episodes which are separate and every of them covers one type of pedestrian behaviour captured by (one or several) model parameters. Firstly, we need to capture the basic properties of the movement process. From this reason we start¹ with agent sizes which were shown as one of the crucial features of the crowd behaviour in our model [19].

This paper provides an insight into one of many calibration episodes from definition of calibration quantities and estimates of simulation time to hypothesis testing used to find an optimum value (or optimum set) of parameters. As was mentioned above, the calibration process is not universalised. Therefore one of the goals of this contribution is to document the whole calibration episode precisely.

DECISION-BASED MODEL AND PEDESTRIAN SIZES

Rules of used decision-based model are defined properly in [18], all model parameters are denoted in Table 1.

Presented calibration episode solves parameters of pedestrian size. We assume that all pedestrians have the same initial size $s \in \mathbf{R}^+$ when they come into the system and, as they move towards the exit in time $t \in \mathbf{R}^+$, we allow to change their size in order to cover the crowd behaviour when the exit capacity is almost reached – we call it pedestrian social size $s_\alpha(t)$. Their social sizes are reduced only to themselves - they see each other still at the initial size s and, only from their perspective, they may decide about their own pressing, i.e. social compression, see Figure 1. This behaviour reflects the real ability of pedestrian to pass through narrow space using body rotation or pressing. The minimum possible pedestrian (social) size is called pedestrian physical size $\tau_s \in \mathbf{R}^+$. Initial and physical sizes are model parameters (see Table 1) and it is fulfilled that $0 < \tau_s \leq s_\alpha(t) \leq s$. The physical size can be estimated from the real data using distance of human shoulders. However, it is possible to design the physical size smaller than the real one in the model if it is required for the modelling purpose. The pedestrian social size is non-increasing function, i.e. the pedestrian is not allowed to expand again if their path is free again – the function trend considers our experimental data used for model calibration with evacuation of one room through one exit. For this kind of experiment, it is sufficient this specific resizing; for different experimental settings, the model can be easily upgradable for an arbitrary trend in the function of pedestrian social size.

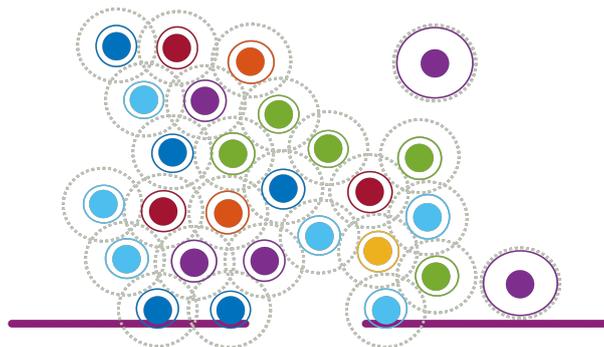


Figure 1: *Pedestrians are drawn as dots in their physical size τ_s . Solid circles around them represent the size $s_\alpha(t)$ at fixed time t , dotted circles depict the initial size s .*

When the reducing of pedestrian (social) sizes is not sufficient for non-zero bottleneck flow, i.e. the exit area is stuck and the pedestrian is the part of the arose arch, a 'crisis rule' is then applied – the pedestrian looks in a specific angle and if there is a free space, the pedestrian leaves the arch and goes right to the exit from the room. Otherwise, the pedestrian has no place to go due to the crowd and their speed becomes zero. Calibration of these 'crisis' parameters will be topic of further research.

¹Calibration episode about maximum speed of agents precedes this one. However, it was really straightforward. Calibration episode about agent sizes is more interesting.

Table 1: List of model parameters.

Parameter	Mark	Units	Meaning	Restrictions
tStep	δt	s	time step	$\delta t \rightarrow 0^+$
xStep	δx	m	spatial step for shortening	$\delta x \rightarrow 0^+$
aStep	$\delta \xi$	rad	angular step for rotation	$\delta \xi \rightarrow 0^+$
wallDist	w	m	min. distance from wall etc.	$w \in \mathbf{R}_0^+$
sizePed	s	m	size of pedestrian	$s \in \mathbf{R}^+$
reducePedStep	δs	m	step to reduce pedestrian size	$\delta s \rightarrow 0^+$
thresholdSizePed	τ_s	m	minimum size of pedestrian	$0 < \tau_s \leq s$
vOpt	v_{opt}	m/s	pedestrian optimum speed	$v_{\text{opt}} \in \mathbf{R}^+$
vIni	v_{ini}	m/s	pedestrian initial velocity	$\ v_{\text{ini}}\ \leq v_{\text{opt}}$
nutAngle	ν	rad	ped. field of vision	$0 < \nu \leq 2\pi$
maxCourseChange	φ	rad	max. change of ped. course	$0 < \varphi \leq 2\pi$
acce	a	m/s ²	pedestrian acceleration	$a \in \mathbf{R}^+$
acceCrisis	a_{crisis}	m/s ²	accel. if an arch occurs	$a_{\text{crisis}} \geq a$
viewAngle	θ	rad	field of vision if arch occurs	$0 < \theta < \pi$
nCP	n_{CP}	-	current checkpoint benefit	$n_{\text{CP}} \geq 0$
m	m_{CP}	-	power of distance to checkpoint	$m_{\text{CP}} \geq 0$

EXPERIMENTAL DATA

The used scenario during model calibration is fixed in accordance with the egress experiment from [2], including size of the room (observed area is 6 m \times 2.24 m with centred exit with a width of 0.6 m) and random inflow (using exponentially distributed time headways at the input) set-up as 1.5 ped/s. Pedestrians (undergraduate students) randomly entered the room by one of three entrances, walked to the opposite wall and left the room by one exit, as seen in Figure 2. By controlling input flow, different conditions from free flow to congestion in the exit area were achieved. In total, our sample is made up of 2000 paths through 10 experimental runs. In this calibration episode, we exclude pure free flow experimental runs.

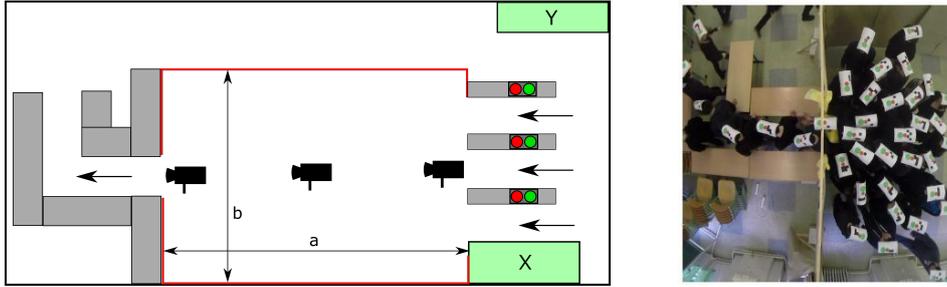


Figure 2: Schema of the experiment and a snapshot from a camera, from [2].

METHODOLOGY

In our calibration episode, we will use for finding an optimum value (or set) following techniques. Firstly, *Chebyshev's inequality*

$$\mathbb{P}\left[|\bar{\xi} - \mu| < \Delta(n, \varepsilon)\right] \geq 1 - \varepsilon, \quad (1)$$

where μ is mean value of a random quantity ξ and $\xi_i, i \in \mathbf{N}, i \leq n$, are random realisations of random quantity ξ . If $\text{Var}(\xi) = \sigma^2 < +\infty$, $\Delta(n, \varepsilon) = \frac{\sigma}{\sqrt{n\varepsilon}}$, where $\varepsilon \in (0, 1)$. In popular terms, we know that we are able to compute the mean value of random quantity ξ using n iterations with an error (from real mean value μ) less than $\Delta(n, \varepsilon)$ with probability greater or equal to $1 - \varepsilon$. This method can be used for estimates of number of iterations needed to obtain a result with specific error for calibration itself. Besides, we will use tight and wide version for finding an *optimum set*, i.e. a set of parametric sets ensuring similarity to experimental data. We can say that a parametric

set belongs to a *tight* (accepted) optimum set if this point including a neighbourhood delimited by its Chebyshev's error belongs to neighbourhood of the experimental value with radius of experimental Chebyshev's error. We can say that a parametric set belongs to a *wide* (cannot be rejected) optimum set if intersection of error neighbourhoods of this point and experimental value is not an empty set. In this version, it cannot be easily said the probability under which we cannot reject any parametric value. This issue is treated in hypothesis testing.

Hypothesis testing is a statistical approach that can be used here in the following way. We test for each parametric set $j \in \mathbf{N}$ null hypothesis H_0 versus its alternative H_1 as

$$H_0 : \boldsymbol{\mu}_j = \boldsymbol{\mu}_E \quad \text{vs.} \quad H_1 : \boldsymbol{\mu}_j \neq \boldsymbol{\mu}_E, \quad (2)$$

where $\boldsymbol{\mu}_E$ and $\boldsymbol{\mu}_j$ represent experimental and j th parametric set respectively. We have to use multivariate test, specifically we choose James' test, see [8, 12]. This test is derived for unknown covariance matrices and unequal number of observations used for obtaining $\boldsymbol{\mu}_j, \boldsymbol{\mu}_E$ and its test statistic, evaluated for every parametric set $j \in \mathbf{N}$,

$$T := (\boldsymbol{\mu}_j - \boldsymbol{\mu}_E)' \left(\frac{1}{n_1} \mathbf{S}_j + \frac{1}{n_2} \mathbf{S}_E \right)^{-1} (\boldsymbol{\mu}_j - \boldsymbol{\mu}_E), \quad (3)$$

where $\mathbf{S}_j, \mathbf{S}_E$ are estimates of covariance matrices, is approximately distributed as $T \sim \chi_d^2$ when H_0 is true. Although there is an assumption which we should examine (model and experimental data should be normally distributed), James' test is robust to non-normality. To check this property, we use test based on skewness and kurtosis of multivariate distribution from [12]. In the both tests, we use fixed significance level $\alpha = 0.05$. Parametric sets with p-value greater than α then belong to the optimum set.

Minimization of Euclidean distance is mentioned here just to be compared with the methods described above, it is the simplest way how to find an optimum value, i.e. to find a minimum of an objective (error) function (chosen as Euclidean distance here). Then the error can be computed for each parametric set $j \in \mathbf{N}$ as follows

$$\text{error}(j) = \sqrt{(\bar{N}_{1,j} - N_1^E)^2 + (\bar{N}_{2,j} - N_2^E)^2}. \quad (4)$$

Major weakness of this method is that it does not take into account a variance of data. We understand top 10% parametric sets with the smallest deviations as the optimum points constituting the optimum set.

CALIBRATION EPISODE

Calibration Quantities

Calibration quantities need to be defined properly to capture pedestrian behaviour for which they were designed in the model. The initial pedestrian size s represents pedestrian size in free flow. On the contrary, the physical pedestrian size τ_s (the minimum reachable social pedestrian size) represents pedestrian size reached in a crowd. In order to capture these two different modes, we evaluate a number of pedestrians in a detector under specific conditions (regulated using another detector). At first, we need to define these detectors and a method how the conditions will be regulated, secondly the number of pedestrians (mean and variance) needs to be evaluated, and finally, the definition of two calibration quantities has to be established.

We define first detector S (small rectangle) for measuring the number of pedestrians and second (helping) detector L (large rectangle) which will be used for regulation of conditions to obtain two regimes discussed above:

$$S := \{x = (x, y) \in \mathbf{R}^2 | x \in \langle 2, 4 \rangle \wedge y \in \langle 0.5, 1.5 \rangle\}, L := \{x \in \mathbf{R}^2 | x \in \langle 1, 5 \rangle \wedge y \in \langle 0, 2.24 \rangle\}. \quad (5)$$

Motivation for placing the small rectangle 0.5 m far from the exit is that we expect higher number of pedestrians in this area than in the exit area. The position and size of the small rectangle is chosen

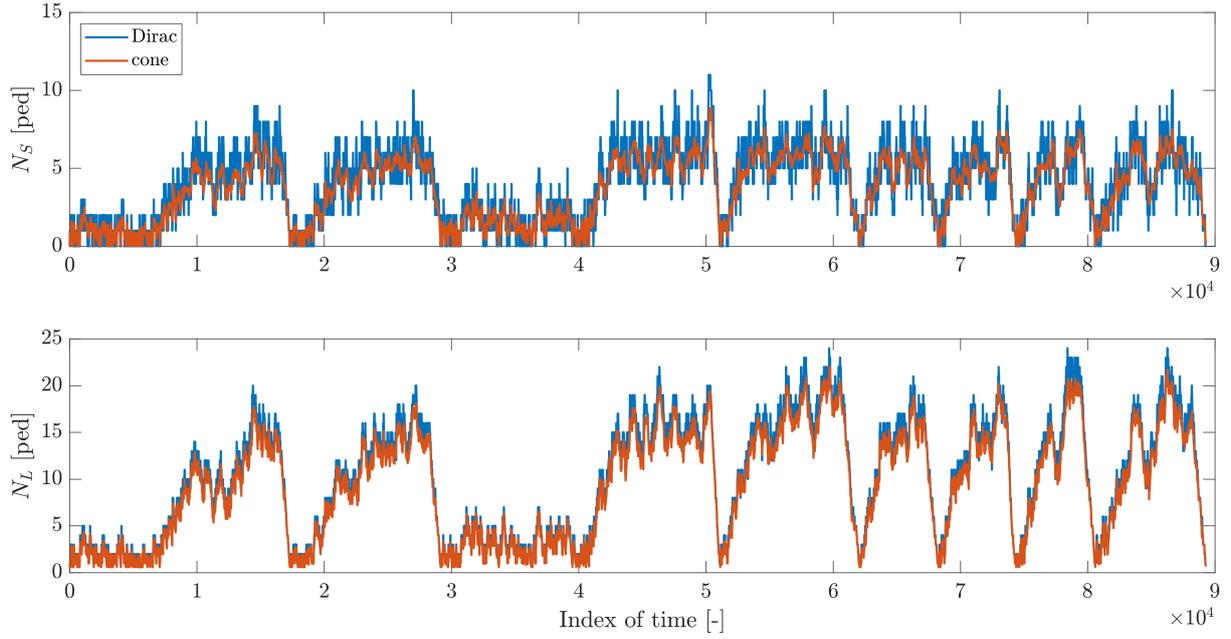


Figure 3: Comparison of standard (Dirac) and conic counting of pedestrians in detector S and L for experimental dataset (all experimental rounds are depicted one by one).

to capture the main corridor. The position and size of the large rectangle is chosen to cover the most of an observed area, excluding its underused peripheries.

To evaluate the number of pedestrians in these detectors, the kernel method [15, 17] will be used, especially

$$N_S := \int_S p(\mathbf{x}, R) = \int_S \sum_{\alpha} p_{\alpha}(\mathbf{x}, R) \quad (6)$$

with conic kernel $p_{\alpha}(\mathbf{x}, R)$ and blur of size $R = 0.9$ m according to parametric study from [20]. The number of pedestrians in large rectangle N_L is defined analogously. We choose this kernel method, see Figure 3, because it produces continuous number of pedestrians in a comparison to the standard (Dirac) counting of pedestrians. Therefore we do not need to decide between 10 or 11 pedestrians during the calibration process - we could choose an arbitrary value between them, for instance 10.759 pedestrians.

Having the detectors and evaluation method already defined, the number of pedestrians can be measured in the small rectangular detector under two regimes

$$N_1 := \max_{t \in \mathbb{R}^+} \{N_S(t) | N_L(t) \leq \tau_L^{(1)}\}, \quad N_2 := \max_{t \in \mathbb{R}^+} \{N_S(t) | N_L(t) > \tau_L^{(2)}\}. \quad (7)$$

We need to find an appropriate value of the thresholds $\tau_L^{(1,2)}$ representing the separation lines between two regimes needed for this calibration episodes. Therefore we will analyse it in our experimental data to set it correctly.

We need to find the value which allows pedestrians to walk in a free flow, i.e. minimum and maximum speed measured in detector S may help

$$\min v_S(t) := \min_{\alpha \in \mathbb{N}} \{\|v_{\alpha}(t)\| : \mathbf{x}_{\alpha}(t) \in S\},$$

where \mathbf{x}_{α} denotes the position of pedestrian α and v_{α} their velocity. This definition works analogously for $\max v_S$. With help of these two quantities, we can distinguish these two regimes in Figure 4. Threshold $\tau_L^{(1)}$ may be chosen as 8 ped because $\min v_S$ and $\max v_S$ clearly decreased around this point. Besides, 8 pedestrians occupying detector L produce average density around 0.9 ped/m²

which is certainly representing the free flow. Hence, we choose $\tau_L^{(1)} = 8$ peds. Threshold $\tau_L^{(1)}$ needs to be chosen to ensure low minimum and also low maximum speed in detector S . Consequently we need to choose $\tau_L^{(2)} \geq 10$ peds. Besides, the averaged density in detector L greater than 2 peds/m² implies condensed conditions (especially assuming a crowd in the exit area). Hence, we choose $\tau_L^{(2)} = 18$ peds ($18/|L| \approx 2$ peds/m²). Moreover, it can be clearly seen from Figure 3, where all experimental rounds are depicted one by one, that the chosen values for $\tau_L^{(1)}$ and $\tau_L^{(2)}$ copy quantitative phenomena for stabilized free flow and stabilized congestion.

Finally, computed experimental values of calibration quantities including variances and Chebyshev's upper boundary $\Delta^E(n, \varepsilon)$ needed for calibration are in Table 2.

Table 2: Computed calibration statistics from the experimental data. Chebyshev's upper boundary Δ^E is computed using $n = 7$ and $\varepsilon = 0.05$.

i	\bar{N}_i^E [ped]	$\text{Var}(\bar{N}_i^E)$ [ped ²]	$\Delta_E(n, \varepsilon)$ [ped]	$\text{Var}(N_i^E)$ [ped ²]
1	4.15	0.01	0.48	0.28
2	7.58	0.06	1.05	0.62

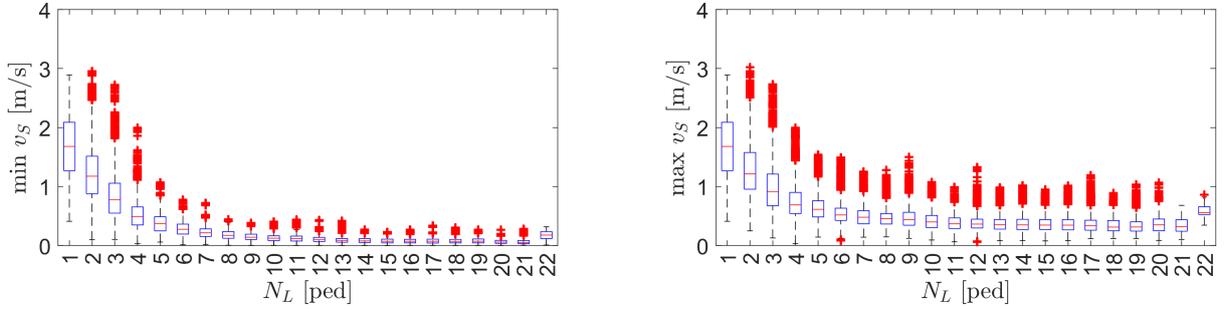


Figure 4: Dependence of minimum and maximum speed in detector S on number of pedestrians N_L .

Set-up of Calibration Episode

We already discussed setting of experimental geometry and inflow into the room. At this stage, we have to examine time length of a simulation and number of pedestrians.

To realise the maximum (needed) experimental time, i.e. the time of running the model to obtain stationary (stable in time) results, we choose three test parametric sets which is variable in s and τ_s and represent all possible parametric situations under condition $0 < \tau_s \leq s$ (low s and low τ_s , high s and low τ_s , high s and high τ_s). Let us remind that other parameters (not calibrated in this episode or already calibrated parameters are fixed at the same value during the whole calibration episode). Exact values for test parametric sets are denoted in Table 3. We performed 50 iterations with maximum experimental time as 500 s for these test parametric sets.

Table 3: Chosen test parametric sets for finding a suitable set-up of this calibration episode and needed simulation time for obtaining quality value in a simulation (using relative error $\varepsilon = 0.05$).

-	s [m]	τ_s [m]	$t_{\text{stop}}^{(1)}$ [s]	$t_{\text{stop}}^{(2)}$ [s]
PS1	0.1	0.1	82.20	-
PS2	0.25	0.1	31.85	136.40
PS3	0.25	0.25	32.70	36.50

We can find the maximum (needed) experimental time t_{stop} using relative error between stationary (obtained by the whole simulation, i.e. it ran 500 s in our case) $N_i^S := N_i(t = 500)$ and actual value of $N_i, i \in \{1, 2\}$ as

$$t_{\text{stop}}^{(i)} := \min \left\{ t \in \mathbf{R}^+ : \frac{|N_i(t) - N_i^S|}{N_i^S} \leq \varepsilon \right\} \quad (8)$$

which can be interpreted as the first time step t fulfilling that absolute difference between actual value of $N_i(t)$ and stationary value N_i^S is less than $\varepsilon \cdot 100\%$ of this stationary value. We performed this metric for all test parametric sets with $\varepsilon = 0.05$ and the results are seen in Table 3. Visualisations of time development (median and inter-quartile range) are depicted for N_2 in Figure 5. It can be seen that PS1 for low s and low τ_s is not able to create congestion (i.e. $N_L \leq 18$ ped) and from this reason there is no curve for this combination in Figure 5. Final value of t_{stop} can be defined as

$$t_{\text{stop}} := \left\lceil \max_{i \in \{1,2\}} \max_{j \in \{1,2,3\}} t_{\text{stop}}^{(i)}(j) \cdot \frac{1}{100} \right\rceil \cdot 100 \quad (9)$$

to give an extra space for possible variations in results due to stochasticity (index j represents different test parametric sets and $\lceil \cdot \rceil$ the ceiling function). Thus, we use $t_{\text{stop}} = 200$ s.

Finally, the number of pedestrians allowed to come into the model remains to be established. We use $c_{\text{ped}} = 300$ ped because of the fact that this number ensures the continuous inflow during 200 s of model run with mean time headway equal to 2 s.

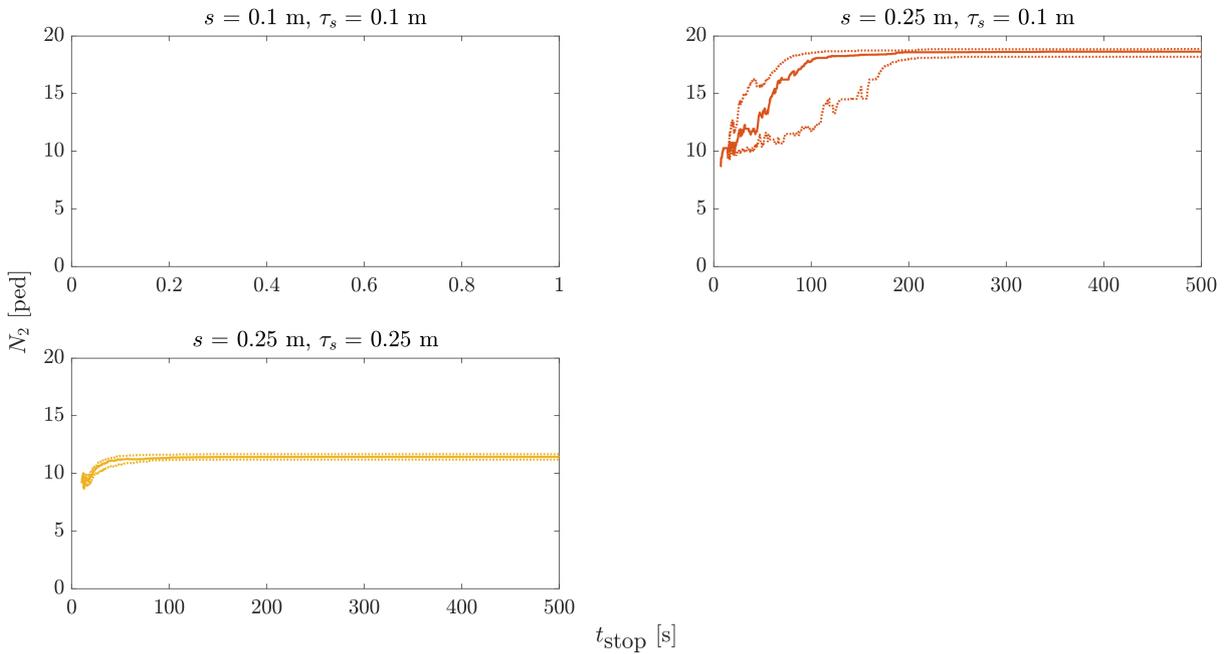


Figure 5: Time development of N_2 (median and inter-quartile range) for different maximum experimental time t_{stop} for test parametric sets from Table 3.

Number of Iterations

Having set up simulation properties which are needed to run a model to be comparable with the used experiment, we need to solve a computational-statistic issue, i.e. we need to find the number of needed iterations to conserve certain accuracy. Using Chebyshev's inequality (1) we are able to fix an error caused by a limited number of iterations of model simulations.

Table 4: Chosen test parametric sets for finding a suitable set-up of this calibration episode and varying number of iterations (using probability error $\varepsilon = 0.05$).

-	$\Delta^{(1)}(n, \varepsilon)$					$\Delta^{(2)}(n, \varepsilon)$				
n [-]	10	20	30	40	50	10	20	30	40	50
PS1	0.71	0.50	0.41	0.36	0.32	-	-	-	-	-
PS2	0.46	0.33	0.27	0.23	0.21	3.75	2.65	2.17	1.88	1.68
PS3	0.59	0.42	0.34	0.30	0.27	0.44	0.31	0.26	0.22	0.20

We use the same test parametric sets as in the previous section (see Table 3) and $\Delta(n, \varepsilon = 0.05)$ is examined for varying number of iterations $n \in \mathbf{N}$. The results can be seen in Table 4 and in Figure

6. We may conclude that the less stable test parametric set in the sense of stochasticity is PS2 (high s and high τ_s). If we choose $n = 50$, we obtain (at worst) almost the same value of error as in the experimental dataset. Hence we conclude with chosen $n = 50$ iterations.

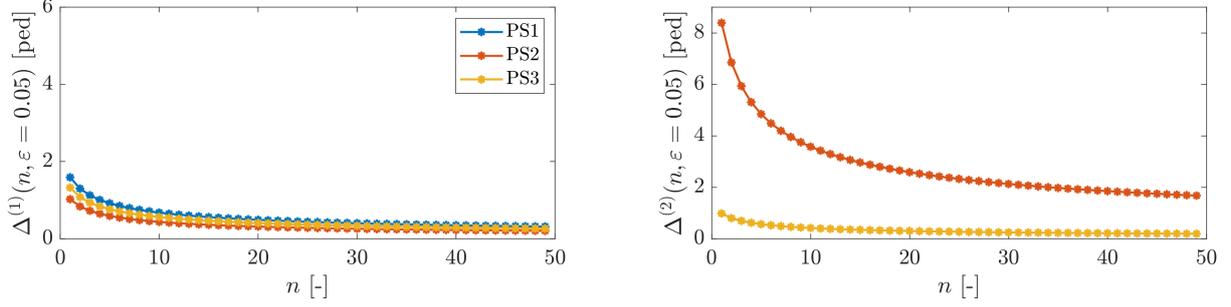


Figure 6: Chebyshev's $\Delta^{(1,2)}(n, \varepsilon)$ for $\varepsilon = 0.05$ for different number of iterations $2 \leq n \leq 50$.

Test of Episode

Since our concept of calibration assumes separate calibration episodes, we have to check if a change in to-be-calibrated parameters s, τ_s produces a change in calibration quantities N_1, N_2 , and if a change in not yet calibrated parameters (other than s, τ_s), which are fixed during this calibration episode, does not cause any change in the calibration quantities N_1, N_2 .

We use again the test parametric sets defined in Table 3 to check the first property. Their values of N_1 and N_2 are depicted in Figure 7 where can be clearly seen that the values of calibration quantities are influenced by the values of to-be-calibrated parameters s and τ_s .

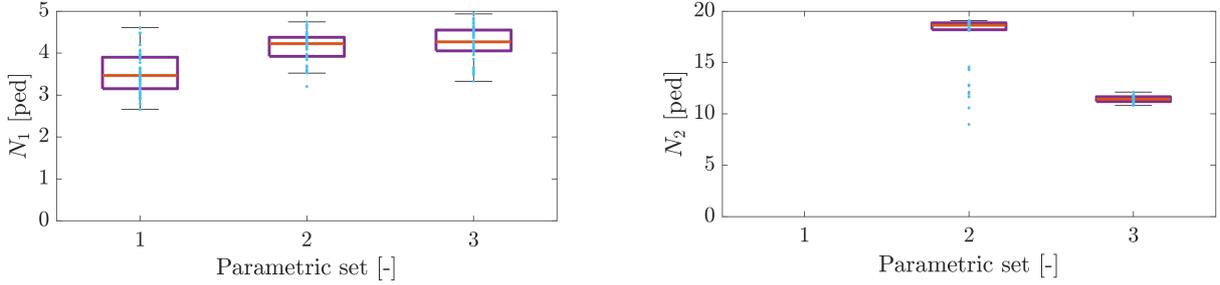


Figure 7: Values of calibration quantities N_1 and N_2 for test parametric sets defined in Table 3.

To check the second property, we need another test parametric sets. Firstly, we fix the to-be-calibrated parameters at arbitrary values, for example, $s = 0.25$ m and $\tau_s = 0.15$ m. Not yet calibrated parameters (10 in total) have to change their values. However, this would involve a huge number of parametric combinations, due to the huge number of parameters with continuous range of values. Thus, we reduce the number of combinations to save the computational time. To do it, we establish three values (low, middle, high) for each not yet calibrated parameter, and an initial setting of them, see Table 5. In every second-property-test parametric set just one parameter will be changed, the other parameters will be kept at the initial values. Then the final number of combinations will be only $3 \cdot 10 = 30$. The results can be seen in Figure 8 and we can state that changes in other parameters do not produce a significant change in calibration quantities.

Perform of Episode

Having prepared every feature of this calibration episode, the final parametric sets can be established. In accordance with our validation and verification study in section in [19], we may expect that calibrated values fulfil $s \geq 0.16$ m and $\tau_s \geq 0.15$ m (to make pedestrians to walk at a speed lower than the maximum speed). Using these findings, we design the final parametric sets for calibration of pedestrian initial size s and pedestrian minimum size τ_s as

$$\{(s, \tau_s) \in \mathbf{R}^2 : s \in \{0.1, 0.12, 0.14, \dots, 0.4\} \wedge \tau_s \in \{0.1, 0.12, 0.14, \dots, 0.4\} \wedge \tau_s \leq s\}. \quad (10)$$

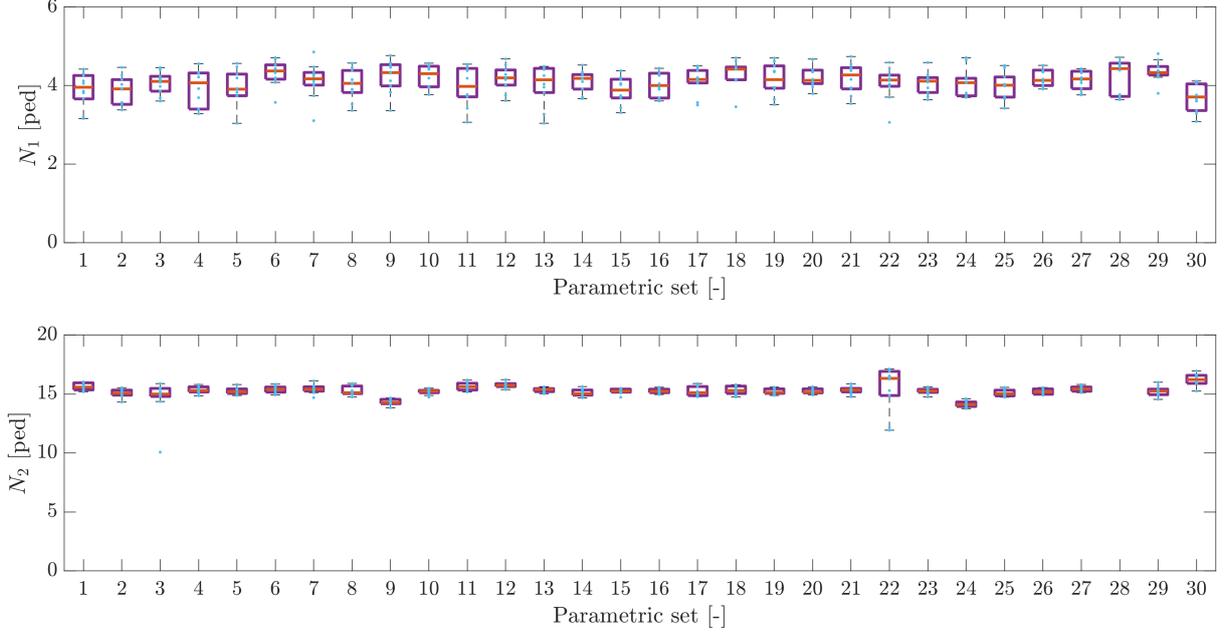


Figure 8: Values of calibration quantities N_1 and N_2 for the second test parametric sets defined in Table 5.

Table 5: List of values for each not yet calibrated parameter used during the test of episode. Initial values are bold.

Parameter	Low	Middle	High
δt [s]	0.01	0.05	0.1
δx [m]	0.01	0.05	0.1
$\delta \xi$ [rad]	0.01	0.05	0.1
δs [m]	0.01	0.05	0.1
w [m]	0	0.05	0.1
a_{crisis} [m/s ²]	30	60	100
ϑ [rad]	$\pi/64$	$\pi/$ 32	$\pi/16$
ν [rad]	$\pi/8$	$\pi/4$	$\pi/2$
a [m/s ²]	0.1	0.5	1
φ [rad]	$\pi/4$	$3\pi/4$	π

Other (not yet calibrated) model parameters are set to their fixed values defined in Table 5. In total, we have 136 final parametric sets whose values of \bar{N}_1 and \bar{N}_2 , and experimental values \bar{N}_1^E, \bar{N}_2^E , are depicted in Figure 9. It can be clearly seen that it is possible to find a parametric set with similar values as the experiment produces. Hence we can use this final parametric sets to perform this calibration episode.

At this stage, we compute Chebyshev's delta method and Euclidean distance. To decide about optimum set in the case of hypothesis testing, firstly, we need to check the normality of the data. We have 7 experimental observations available, thus the normality test does not have a high power - we just can say that our experimental dataset does not produce a distribution which is fundamentally different from a normal distribution.

We have 50 observations (iterations) for each parametric set in our model, thus we can decide about normality of data more significantly. A statistic describing multivariate skewness is denoted as A , statistic describing multivariate kurtosis is denoted as B (for more details see [12]) and for the both statistic we evaluate if hypothesis H_0 (skewness/kurtosis is the same as for normal distribution) is rejected or not at significance level $\alpha = 0.05$. We found that we cannot reject the null hypothesis in

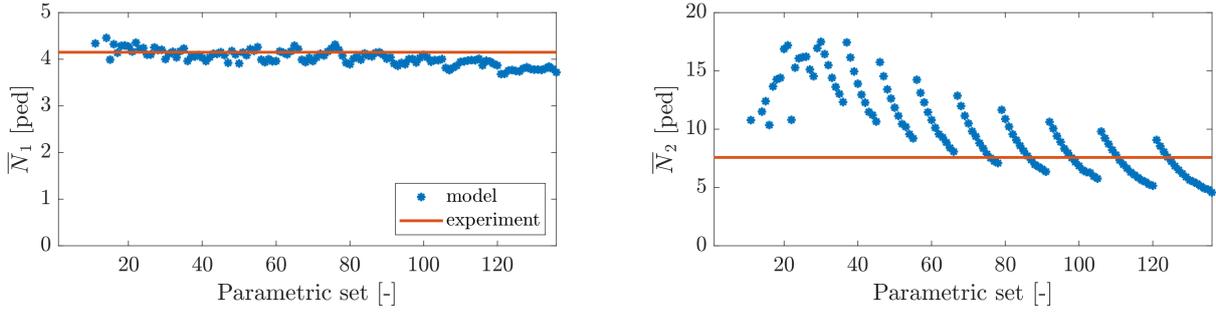


Figure 9: Model values of \bar{N}_1 , \bar{N}_2 for every parametric set and experimental values \bar{N}_1^E, \bar{N}_2^E .

the most cases. For only four parametric sets the both statistics A and B at the same time reject the null hypothesis. Therefore we can conclude that we have approximately normally distributed data.

Having fulfilled assumptions, James' test [8] is performed for every parametric set j at significance level $\alpha = 0.05$ and the final optimum set can be seen in Table 6.

In Figure 10 and Figure 11, there is a visual comparison of optimum sets obtained by these three methods. Even though Chebyshev and Euclidean distance are just approximate in a comparison to hypothesis testing, all three methods give similar optimum sets. However, only in the case of hypothesis testing, p-values denoted as p_j (if we rejected parametric set j , we would make an error with probability p_j) of every parametric set j is evaluated.

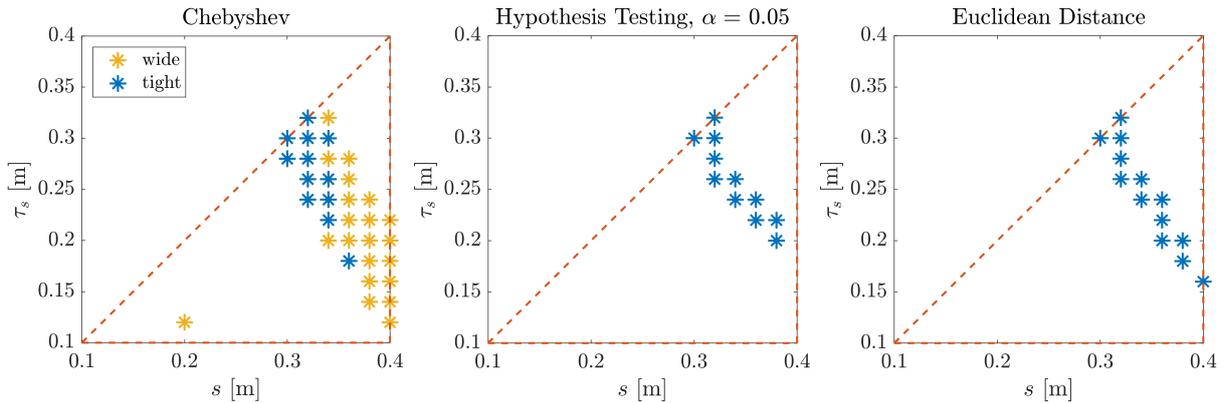


Figure 10: Final results of the calibration episode about pedestrian sizes s and τ_s obtained by different methods. Ranges of examined parameters are drawn by red dashed line.

Parametric sets chosen by hypothesis testing are consistent with our validation study from [19]. We noted there, for instance, that initially smaller pedestrians do not change their size, i.e. τ_s is almost equal to their initial size due to no crowd and no reason to resize. Initially wider pedestrians allow more resizing than the initially smaller pedestrians - although they could leave the room much bigger, their τ_s is much smaller. It results from the fact that wider pedestrians produce congestion near the exit, thus the ability to pass through narrow space using body rotation or pressing is stronger.

As we have more options for pedestrian sizes to choose, we expect that it may be more suitable for further purposes to choose the values with higher difference between s and τ_s . This choice is able to model more type of social compression during a pedestrian movement in a room and to ensure more diversity between pedestrians themselves. Therefore we choose $s = 0.38$ m and $\tau_s = 0.20$ m, which is the fifth best result according to its p-value, see Table 6 or Figure 11. Let us note that parametric sets with $\tau_s > 0.30$ m (parametric set $j = 78$ from Table 6) are not possible due to the size of the exit in experiment which is 0.60 m.

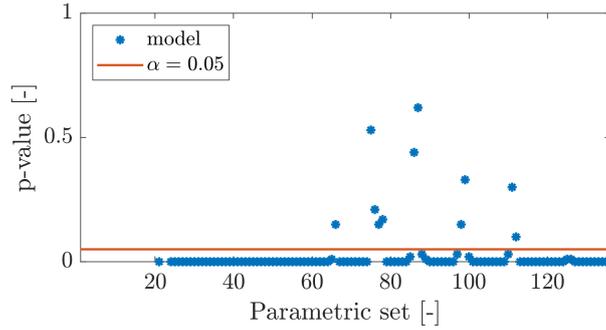


Figure 11: Final results of the calibration episode about pedestrian sizes s and τ_s obtained by hypothesis testing: p -values for every parametric set.

Table 6: Final results obtained by hypothesis testing of the calibration episode about pedestrian sizes s and τ_s .

j [-]	s [m]	τ_s [m]	p-value [-]
87	0.34	0.26	0.62
75	0.32	0.26	0.53
86	0.34	0.24	0.44
99	0.36	0.24	0.33
111	0.38	0.20	0.30
76	0.32	0.28	0.21
78	0.32	0.32	0.17
98	0.36	0.22	0.15
77	0.32	0.30	0.15
66	0.30	0.30	0.15
112	0.38	0.22	0.10

CONCLUSIONS

To summarize, we defined calibration quantities for capturing parameters of pedestrian sizes in our decision-based model, which are influenced only by appropriate model parameters. We studied maximum needed time t_{stop} to get a stationary values of calibration quantities which saved a lot of simulation time that would have been useless. Chebyshev's estimate of number of iterations had a strong impact on the quality of the results of the whole calibration process. Furthermore, hypothesis testing, which we used to find an optimum parametric set, is powerful mathematical tool which can be easily interpretable. Although Euclidean distance does not involve variances of observations and Chebyshev's estimate does not provide exact probabilities of acceptance, they found the optimum set similarly to hypothesis testing.

To conclude, we recommend to use statistical methods, including testing of hypothesis, throughout the entire calibration process of any model. For further calibration episodes, we choose $s = 0.38$ m and $\tau_s = 0.20$ m.

References

- [1] Nikolai WF Bode and Enrico Ronchi. Statistical model fitting and model selection in pedestrian dynamics research. *Collective Dynamics*, 4:1–32, 2019.
- [2] Marek Bukáček, Pavel Hrabák, and Milan Krbálek. Microscopic travel-time analysis of bottleneck experiments. *Transportmetrica A: transport science*, 14(5-6):375–391, 2018.

- [3] Mario Campanella, Serge Hoogendoorn, and Winnie Daamen. Quantitative and qualitative validation procedure for general use of pedestrian models. In *Pedestrian and Evacuation Dynamics 2012*, pages 891–905. Springer, 2014.
- [4] MC Campanella, SP Hoogendoorn, and W Daamen. A methodology to calibrate pedestrian walker models using multiple-objectives. In *Pedestrian and Evacuation Dynamics*, pages 755–759. Springer, 2011.
- [5] Winnie Daamen and Serge Hoogendoorn. Calibration of pedestrian simulation model for emergency doors by pedestrian type. *Transportation Research Record*, 2316(1):69–75, 2012.
- [6] Marion Gödel, Nikolai Bode, Gerta Köster, and Hans-Joachim Bungartz. Bayesian inference methods to calibrate crowd dynamics models for safety applications. *Safety science*, 147:105586, 2022.
- [7] Serge P Hoogendoorn and Winnie Daamen. Microscopic calibration and validation of pedestrian models: Cross-comparison of models using experimental data. In *Traffic and Granular Flow'05*, pages 329–340. Springer, 2007.
- [8] G. S. James. Tests of linear hypotheses in univariate and multivariate analysis when the ratios of the population variances are unknown. *Biometrika*, 41(1/2):19–43, 1954.
- [9] Moonsoo Ko, Taewan Kim, and Keemin Sohn. Calibrating a social-force-based pedestrian walking model based on maximum likelihood estimation. *Transportation*, 40(1):91–107, 2013.
- [10] Ruggiero Lovreglio, Enrico Ronchi, and Daniel Nilsson. Calibrating floor field cellular automaton models for pedestrian dynamics by using likelihood function optimization. *Physica A: Statistical Mechanics and its Applications*, 438:308–320, 2015.
- [11] Francisco Martinez-Gil, Miguel Lozano, Ignacio García-Fernández, and Fernando Fernández. Modeling, evaluation, and scale on artificial pedestrians: a literature review. *ACM Computing Surveys (CSUR)*, 50(5):72, 2017.
- [12] George AF Seber. *Multivariate observations*. John Wiley & Sons, 2009.
- [13] Armin Seyfried and Andreas Schadschneider. Fundamental diagram and validation of crowd models. In *International Conference on Cellular Automata*, pages 563–566. Springer, 2008.
- [14] Martijn Sparnaaij. How to calibrate a pedestrian simulation model: An investigation into how the choices of scenarios and metrics influence the calibration. Master thesis, TU Delft, 2017.
- [15] Bernhard Steffen and Armin Seyfried. Methods for measuring pedestrian density, flow, speed and direction with minimal scatter. *Physica A: Statistical mechanics and its applications*, 389(9):1902–1910, 2010.
- [16] Tomer Toledo and Haris N Koutsopoulos. Statistical validation of traffic simulation models. *Transportation Research Record*, 1876(1):142–150, 2004.
- [17] Antoine Tordeux, Jun Zhang, Bernhard Steffen, and Armin Seyfried. Quantitative comparison of estimations for the density within pedestrian streams. *Journal of statistical mechanics: theory and experiment*, 2015(6):P06030, 2015.
- [18] Jana Vacková and Marek Bukáček. Ruling principles for decision-based pedestrian model. In *Stochastic and Physical Monitoring Systems 2019*. SPMS, 2019.
- [19] Jana Vacková and Marek Bukáček. Social and physical pedestrian sizes and their impact on the decision-based modelling. In *The Fire and Evacuation Modeling Technical Conference*. FEMTC, 2020.
- [20] Jana Vacková and Marek Bukáček. Kernel estimates as general concept for the measuring of pedestrian density. *arXiv preprint arXiv:2205.10145*, 2022.