Collective Dynamics

Calibration of Decision-Based Crowd-Behaviour Model

Jana Vacková[1] · Marek Bukáček[1]

[1] FNSPE, Czech Technical University in Prague, Czech Republic
E-mail: janca.vackova@fjfi.cvut.cz, marek.bukacek@fjfi.cvut.cz

Received: 30 Octobre 2023 / Last revision received: 12 April 2024 / Accepted: 15 April 2024
DOI: 10.17815/CD.2024.155

Abstract Various methods of calibration are used depending on the model type, application, and individual preferences. While there is no universally applicable method, statistical techniques have become popular in recent decades. The introduced calibration concept consists of separate calibration episodes to avoid choosing only a few metrics to describe the whole system and a high computational time increasing exponentially with the number of parameters. These episodes are designed to be separated from each other and to cover one type of pedestrian behaviour captured by some model parameters. The design of the calibration quantities; the estimate of the needed simulation time to get stationary results; and the number of replications by Chebyshev's inequality influencing the quality of the results are discussed. Furthermore, hypothesis testing is used. This calibration process can be applied to any pedestrian model; this paper deals with a case study: its application to the crowd-behaviour phase in the author's decision-based model.

Keywords Crowd-behaviour · decision-based model · calibration · hypothesis testing

1 Introduction

The calibration process aims to find the eligible parameter values to mimic the real world represented by experimental data. The paper [1] provides an overview of common calibration methods and types. In terms of [2] or [3], model validation and model calibration are supposed to be strongly connected, however, as was discussed in [1], it is a common mistake, which we agree with: calibration is a learning process (using training data[1]); in contrast to validation is checking the model properties (using testing data). Nevertheless, there is no standardized approach to model calibration yet [1,4]. The choice of calibration

[1] We assume indirect calibration. Direct calibration (without any training data) can be generally performed.

methods varies based on the model type and its intended application, often influenced by personal preferences. Selecting the suitable quantity to represent the data's characteristics and determining the appropriate method, including the objective function, to minimize the error between the model and actual data is a complex task.

In recent decades, statistical methods for calibration became more popular, e.g. [2,5]. There are strong advantages of the statistical approach, as was discussed in [5]: statistics is a well-established branch of study with many general methods and techniques, mostly already implemented in software, and can be used for solving a wide range of problems due to their generality. Moreover, statistical methods incorporate data stochasticity and uncertainty, which should be considered in pedestrian dynamics in general. However, the most popular method is still maximum likelihood estimation, e.g. [6, 7], bringing only a point estimate of a parameter. Few publications taking into account a distribution of a parameter in question arose recently, e.g. [1], and showed that statistical methods are a very promising approach in the calibration. Nevertheless, there are only a few studies where their authors included also statistical model checking of assumptions [5].

The authors of [2,8] discussed the number of replications due to the stochasticity of the model output. Different from [9], they used a sequential approach in [2], which means that firstly the simulation time is set, then the number of replications is found using the sample standard deviation of data with pre-defined allowable error.

All mentioned approaches shaped our perspective and helped to form our calibration strategy. The calibration concept introduced in this paper consists of separate calibration episodes to avoid choosing just a few metrics and quantities to describe the whole, complex system. The episodes are designed to be separated from each other and to cover one type of pedestrian behaviour captured by (one or several) model parameters. Moreover, using this approach we avoid a high computational time depending exponentially on the number of calibrated model parameters.

Since the calibration process is not universalised [1] and is under-reported [5], the main goal is to introduce the microscopic calibration methodology based on statistical methods and report it deeply while applying to the author's decision-based model using an egress experimental data. The case study will be performed in the crowd-behaviour phase.

Experimental Data Experimental data from [10] are used here. Pedestrians (undergraduate students) randomly entered the room by one of three entrances, walked to the opposite wall, and left the room by one exit. By controlling the input flow, different conditions from free flow to congestion in the exit area were achieved. In total, our sample is made up of 2000 trajectories through 10 experimental runs.

Decision-Based Model In general, pedestrian movement is comprised of several phases: strategic (a long-term plan), tactical (short-term decisions), and operational (finding a specific position to move) [11–13]. Here, we focus on the operational phase covering the crowd behaviour; see details about the model definition in [14, 15].

Mention the rules briefly, the agent tries to find their new position using the optimum velocity. If it fails, the pedestrian needs to change the optimum direction due to a collision with room equipment or other pedestrians. To maximize the benefit for pedestrians (even though the optimum direction is not allowed), a change of course is primarily examined because it does not imply a change in the pedestrian's speed. The optimum pedestrian's course is permitted to be changed by a rotation of the pedestrian's body with a maximum angle $\varphi \in \langle 0, 2\pi \rangle$ (per a time step). If the change in the pedestrian course does not solve the conflict, the slowing down is then applied. All these rules are evaluated in the pedestrian's field of vision $\nu \in \langle 0, \pi)$.

A concept of an agent's social size varying in time according to circumstances in the pedestrian surroundings is applied and represents a social compression. This resizing can make an outflow smooth again if an arc occurs. When the reducing of pedestrian sizes is not sufficient for non-zero outflow, a crisis rule is then applied: the randomly chosen pedestrian looks in a crisis angle $\vartheta \in (0, \pi/2)$, if there is a free space, the pedestrian leaves the arc and goes through the exit, see Fig. 1. The agent sizes were shown important to capture the density at the exit area properly [16]; the appropriate parameters were already calibrated [14].

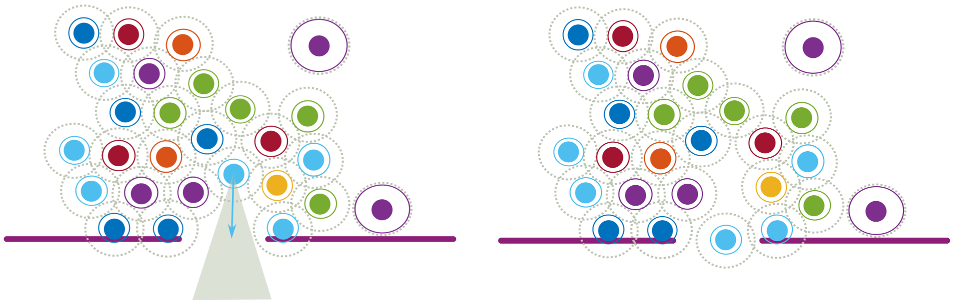

Figure 1 Dense crowd behaviour: when an arc occurs, the solution by crisis model rule is provided. Solid circles around pedestrians represent the social size. Dotted circles depict the initial size.

To summarize, the main not-yet-calibrated parameters influencing the dense-crowd behaviour in the authors' model are the field of vision ν, the crisis field of vision ϑ in the exit area, and the maximum possible change of pedestrian course φ. These parameters are crucial features to prevent this model from being stuck (they control an arc appearance).

2 Calibration Strategy

In general, we will perform an automatic quantitative calibration, mostly microscopic non-physical calibration. Although we will use the n_p-at-a-time approach, we cannot miss the global solutions since we design the episodes not to influence the not-yet-calibrated parameters, which is the benefit of our method. Each calibration episode (CE) has to be designed to be (hierarchically) separated from the others covering one type of pedestrian movement (it can be influenced just by values of already calibrated parameters). The establishment of each CE has the same following stages:

Choice of Calibration Quantity/ies It is necessary to establish the calibration quantity (CQ) reasonably to capture any specific type of pedestrian movement based on real system behaviour. Moreover, we need to assume that the CQ is designed to generate a conditionally independent random sample, as was discussed in [5]. We argue that this property can be understood as fulfilled since we observe the system in a stationary state under fixed conditions following used experimental data. To be complete, calibration statistics are estimated as follows: the expected value using sample mean; and the variance using sample variance.

Set-up of CE The similarity of the input into the model simulation needs to be ensured by setting of the experimental properties (room geometry, inflow, maximum experimental time).

Estimate of Number of Replications If it is possible, we establish the number of the model replications $n \in N$ conserving a specific predefined probability of the results for each used parametric setting. Chebyshev's inequality is eligible for this computational-statistic issue; it claims that a real mean value μ of a quantity Q is not farther from its estimated mean value $\overline{\xi}_n$ than $\Delta(n,\varepsilon) = \frac{\sigma}{\sqrt{n\varepsilon}}$ with probability at least $1-\varepsilon$, where $\varepsilon \in (0,1)$. If it is not possible to use this estimate properly, a reasonable discussion needs to follow.

Test of CE We have to check: if a change in to-be-calibrated parameters produces a change in CQs (1), and if a change in not-yet-calibrated parameters (other than to-be-calibrated parameters), which are fixed during the CE, does not cause any change in the CQs (2). To test that, hypothesis testing will be applied [17] (using few test parametric sets different from final parametric sets): analysis of variances (ANOVA) if its assumptions (normality, equal variances) are fulfilled; non-parametric Kruskal-Wallis (KW) if they are not. If any parametric set differs, Tukey's honest significant difference procedure (HSD) is applied for the detection of which one it is. Note that to test the property (2), low, medium, and high values for each of the not yet calibrated parameters is sufficient.

Perform of CE, Finding an Optimum Set Firstly, two terms need to be defined. *Tolerance set* consists of parametric sets ensuring similarity to experimental data, i.e. the parametric sets obtained by the CE. *Optimum parametric set* is the parametric set ensuring the best similarity to the experimental data. Hypothesis testing will be used to find the tolerance parametric set since it is easily interpretable and takes into account (co)variances of the data. We test under specific significance level for each parametric set $j \in N$ null hypothesis $H_0 : \mu_j = \mu_E$ against its alternative $H_A : \mu_j \neq \mu_E$, where $\mu_E \in R^k$ and $\mu_j \in R^k$ represent experimental and model real mean value respectively and k is the number of calibrating parameters. Parametric sets with p-value greater than significance level α then belong to the tolerance set (i.e. the null hypothesis is not rejected). The optimum parametric set is established by expert knowledge. Different tests can be used to check this hypothesis; James' test [18] will be applied here.

3 Calibration Episode of the Crowd Behaviour

The pedestrian behaviour in a crowd is represented in the model using the fields of vision (standard $\nu \in \langle 0, \pi)$; if an arc occurs $\vartheta \in (0, \pi/2)$) and maximum rotation $\varphi \in (0, 2\pi)$. We acknowledge that each step is discussed briefly here due to space constraints.

CQs Since the crowd behaviour needs to cover self-organized phenomena, we need to capture an occurrence of arcs mainly in the exit area which was detected [15,16] as crucial for the model. The flow brings information about any not vanished arc; the time headways bring information about potential arcs that vanished during the experimental run (simulation); and the third quartile of the time headways represents the 'almost longest' arc that has vanished. The CQs evaluated using the experimental data give the following results: $\overline{J} = 1.149$ ped/s, $\overline{\Delta T} = 0.875$ s, $\overline{\mathrm{Q3}\Delta T} = 0.982$ s.

Set-up of CE The input number of pedestrians is fixed to 300 and $t_{\text{stop}} = 200$ s to conserve the similarity with the experiment (random inflow set-up as 1.5 ped/s) and the observed number of trajectories in each round. To test this choice, we established 18 test parametric sets ($\varphi \in \{\pi/50, \pi, 2\pi\}$, $\vartheta \in \{\pi/50, \pi/5, 2\pi/5\}$, $\vartheta \in \{\pi/50, \pi/3\}$) covering the whole suitable parametric space; the results (50 replications) are depicted in Fig. 2. Each test parametric set generates the flow proportional to $1/t_{\text{stop}}$ for $t_{\text{stop}} > 200$ s; ΔT and Q3 ΔT give stationary values early.

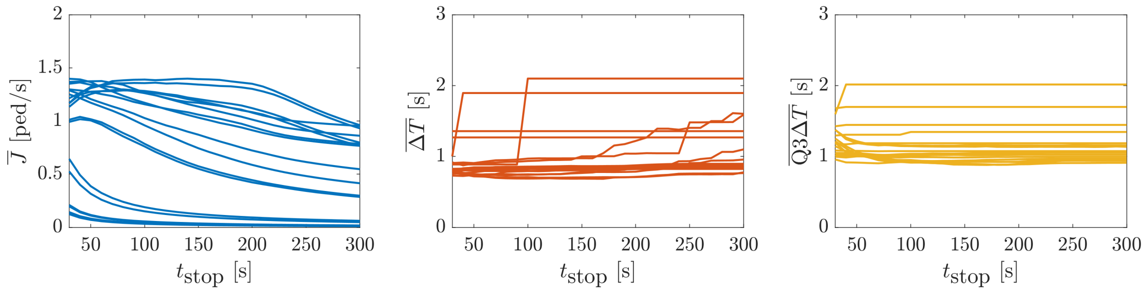

Figure 2 Time development of the CQs (sample means) for different t_{stop} for test parametric sets.

Number of Replications The same test parametric set was simulated in 50 replications (with other parameters set to some fixed values) and Chebyshev's delta is evaluated. There is a high variability for ΔT and Q3 ΔT, mainly for the parametric sets generating a high number of arcs or even forever stuck exit, i.e. they have high sample variances which cause the high Chebyshev's delta out of the desired precision as in the experiment. However, these parametric sets are not able to generate any stable results with self-vanishing arcs, i.e. a similar result as in the experiment. That implies a higher number of replications does not improve the results. Finally, 20 replications are chosen.

Test of CE A change in to-be-calibrated parameters has to produce a desired change in the designed CQs. The test parametric sets are used again with the result that ΔT seems to be almost constant in contrast to the flow J detecting the situations when the exit area is stuck; Q3 ΔT seems to be similar to ΔT, however less stable for parametric sets having almost zero outflow J which supports the assumption about higher dispersion of the time headways. Since the data have not the same variances, the KW test is used for each quantity. The null hypothesis about the equality of distributions (means) is rejected for the flow J (p-value $< 10^{-135}$), ΔT (p-value $< 10^{-24}$), and Q3 ΔT (p-value $< 10^{-34}$). Besides, the rest of the model parameters are spatial δx, temporal δt, angular $\delta \xi$, and pedestrian size δs steps, their values are fixed. Thus, the test of the property (2) is unnecessary.

Perform of CE The final parametric sets are established with the step of $\pi/10$, considering the ranges of the parameters discussed above (900 combinations). Since our task is three-dimensional due to three CQs, James' test will be used. The test [19] of its multivariate-normality assumption [18] fails, since we have 20 model replications which is not sufficient to reveal the normality. However, the James' test will be performed despite the unfulfilled assumption; not rejected parametric sets are written in Tab. 1, i.e. the maximum pedestrian's change of course needs to be around 2π per δt.

φ [rad]	ν [rad]	ϑ [rad]	p-value [-]
$19\pi/10$	$\pi/10$	$\pi/5$	0.18
$19\pi/10$	$\pi/10$	$3\pi/10$	0.08
$19\pi/10$	$\pi/10$	$2\pi/5$	0.16
$19\pi/10$	$\pi/10$	$\pi/2$	0.13
$19\pi/10$	$\pi/5$	$\pi/5$	0.09
$19\pi/10$	$\pi/5$	$\pi/2$	0.13
2π	$\pi/10$	$3\pi/10$	0.12
2π	$\pi/5$	$\pi/5$	0.07

Table 1 The optimum set of the CE about the crowd behaviour in the model.

4 Conclusions

The introduced calibration concept presents a novel approach to model calibration. It can be customized for any modeling purpose through the selection of either microscopic or macroscopic CQs, and the statistical tests addressing specific types of behaviour. The concept holds promise for the future due to its variability and mathematical background.

The tolerance set for the crowd-behaviour CE conducted in this study is detailed in Tab. 1, any of these sets can serve as the optimum parametric set. These values may indicate the potential model's weakness which will be studied in subsequent research: the model may not cover stochasticity sufficiently.

Note that if simulated data exhibit low variances, hypothesis testing succeeds when the parametric space is deeply examined, leading to similar results obtained through Euclidean distance minimization. Conversely, if the variances are higher, hypothesis testing takes into account their sizes, a crucial aspect in aligning with experimental data.

Acknowledgements This work was supported by the Ministry of Education, Youth, and Sports of the Czech Republic (MŠMT ČR) under Grant SGS21/165/OHK4/3T/14 and under project Center for Advanced Applied Science CZ.02.1.01/0.0/0.0/16-019/0000778.

Ethics Statement The participants of the experiment approved for the experimental data to be stored, used and published for academic purposes.

Author Contributions Jana Vacková: Methodology, Conceptualization, Software, Visualization, Writing - original draft / Marek Bukáček: Supervision, Methodology.

References

[1] Gödel, M., Bode, N., Köster, G., Bungartz, H.J.: Bayesian inference methods to calibrate crowd dynamics models for safety applications. Safety science **147**, 105586 (2022). doi:10.1016/j.ssci.2021.105586

[2] Toledo, T., Koutsopoulos, H.N.: Statistical validation of traffic simulation models. Transportation Research Record **1876**(1), 142–150 (2004). doi:10.3141/1876-15

[3] Schadschneider, A., Chraibi, M., Seyfried, A., Tordeux, A., Zhang, J.: Pedestrian dynamics: From empirical results to modeling. In: Crowd Dynamics, Volume 1, pp. 63–102. Springer (2018). doi:10.1007/978-3-030-05129-7

[4] Campanella, M., Hoogendoorn, S., Daamen, W.: Quantitative and qualitative validation procedure for general use of pedestrian models. In: Pedestrian and Evacuation Dynamics 2012, pp. 891–905. Springer (2014). doi:10.1007/978-3-319-02447-9

[5] Bode, N.W., Ronchi, E.: Statistical model fitting and model selection in pedestrian dynamics research. Collective Dynamics **4**, 1–32 (2019). doi:10.17815/CD.2019.20

[6] Lovreglio, R., Ronchi, E., Nilsson, D.: 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). doi:10.1016/j.physa.2015.06.040

[7] Ko, M., Kim, T., Sohn, K.: Calibrating a social-force-based pedestrian walking model based on maximum likelihood estimation. Transportation **40**(1), 91–107 (2013). doi:10.1007/s11116-012-9411-z

[8] Sparnaaij, M., Duives, D.C., Knoop, V.L., Hoogendoorn, S.P.: Multiobjective calibration framework for pedestrian simulation models: A study on the effect of movement base cases, metrics, and density levels. Journal of Advanced Transportation **2019** (2019). doi:10.1155/2019/5874085

[9] Ronchi, E., Kuligowski, E.D., Reneke, P.A., Peacock, R.D., Nilsson, D.: The process of verification and validation of building fire evacuation models (2013). URL http://nvlpubs.nist.gov/nistpubs/technicalnotes/NIST.TN.1822.pdf

[10] Bukáček, M., Hrabák, P., Krbálek, M.: Microscopic travel-time analysis of bottleneck experiments. Transportmetrica A: transport science **14**(5-6), 375–391 (2018). doi:10.1080/23249935.2017.1419423

[11] Cheng, L., Yarlagadda, R., Fookes, C., Yarlagadda, P.K.: A review of pedestrian group dynamics and methodologies in modelling pedestrian group behaviours. World **1**(1), 002–013 (2014)

[12] Kang, W., Han, Y.: A simple and realistic pedestrian model for crowd simulation and application. Preprint (2017). doi:10.48550/arXiv.1708.03080

[13] Martinez-Gil, F., Lozano, M., García-Fernández, I., Fernández, F.: Modeling, evaluation, and scale on artificial pedestrians: a literature review. ACM Computing Surveys (CSUR) **50**(5), 72 (2017). doi:10.1145/3117808

[14] Vacková, J., Bukáček, M.: Calibration of pedestrian sizes in decision-based modelling. In: The Fire and Evacuation Modeling Technical Conference. FEMTC (2022)

[15] Vacková, J., Bukáček, M.: Ruling principles for decision-based pedestrian model. In: Stochastic and Physical Monitoring Systems 2019. SPMS (2019)

[16] Vacková, J., Bukáček, M.: Social and physical pedestrian sizes and their impact on the decision-based modelling. In: The Fire and Evacuation Modeling Technical Conference. FEMTC (2020). URL https://files.thunderheadeng.com/femtc/2020_d3-11-vackova-paper.pdf

[17] Lehmann, E., Romano, J.: Testing statistical hypotheses. MR2135927 (2005). doi:10.1007/978-3-030-70578-7

[18] James, G.S.: 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). URL http://www.jstor.org/stable/2333003

[19] Seber, G.A.: Multivariate observations. John Wiley & Sons (2009)