Safety Science 147 (2022) 105586

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/safety

Bayesian inference methods to calibrate crowd dynamics models for safety applications

Marion Gödel [a,b,*], Nikolai Bode [c,**], Gerta Köster [b], Hans-Joachim Bungartz [a]

[a] Department of Informatics, Technical University of Munich, Boltzmannstraße 3, 85748 Garching, Germany
[b] Department of Computer Science and Mathematics, Munich University of Applied Sciences, Lothstraße 64, 80335 München, Germany
[c] Department of Engineering Mathematics, University of Bristol, Ada Lovelace Building, University Walk, Bristol, BS8 1TW, United Kingdom

ARTICLE INFO

Keywords:
Crowd dynamics
Pedestrian safety
Model calibration
Bayesian inference
Bottleneck
Safety engineering

ABSTRACT

Crowd simulation is a crucial tool to assess risks and engineer crowd safety at events and in built infrastructure. Simulations can be used for what-if studies, for real-time predictions, as well as to develop regulations for crowd safety. A reliable prediction requires a carefully calibrated model. Model parameters are often calibrated as point estimates, single parameter values for which the model evaluation fits given data best. In contrast, Bayesian inference provides a full posterior distribution for the fitted parameters that includes the residual uncertainty after calibration. In this work, we calibrate a microscopic model and an emulator derived from a microscopic model for crowd dynamics using point estimates and Approximate Bayesian Computation. We calibrate on data measuring the flow through a key scenario of crowd safety: a bottleneck. We vary the bottleneck width and demonstrate via three case studies the advantages and shortcomings of the two calibration techniques. In a case with a unimodal posterior, both methods yield similar results. However, one safety-relevant case study, that mimics the dynamics of evacuating people squeezing through an opening, exhibits a faster-is-slower dynamic where multiple free-flow speeds lead to the same flow. In this case, only Bayesian inference reveals the true bimodal shape of the posterior distribution. For multidimensional calibration, we illustrate that Bayesian inference allows accurate calibration by describing parameter relations. We conclude that, in practice, point estimation often seems sufficient, but Bayesian inference methods are necessary to capture important structural information about the uncertain parameters, and thus the physics of safety.

1. Introduction

Microscopic crowd simulations are a common tool to evaluate evacuation concepts for new buildings, check safety concepts for events or increase comfort and capacities in infrastructure. For a reliable prediction, careful calibration of the model parameters is essential. We understand calibration as the process of adjusting model parameters so that a predefined simulation outcome matches observed data. Since the readers of this manuscript may have diverse backgrounds, we include a short glossary of key terms used in this manuscript in Table 1.

1.1. Motivation

Often, parameters that cannot be measured directly are calibrated. Calibration can be performed in different ways, from manual adjustment by visual comparison of simulation outcomes and observations to automated alignment using quantitative simulation results. For automated calibration, point estimates such as maximum likelihood estimates are a common approach in pedestrian dynamics (Daamen and Hoogendoorn, 2012; Hoogendoorn and Daamen, 2006, 2007; Ko et al., 2013; Lovreglio et al., 2015). While point estimates are easy to implement and relatively cheap to evaluate, one drawback is that they provide, by definition, a single parameter estimate. When a parameter is influential, even relatively small changes of the parameter value likely have a significant impact on the quantity of interest. If we use a point estimate for such a parameter for our subsequent studies, neglecting the uncertainty in the result of the calibration, the results of our studies may appear more certain than they are.

We propose to apply Bayesian inference methods that provide a full posterior distribution of the uncertain parameters instead. While Bayesian inference methods are computationally more costly than most point estimates, the posterior comprises uncertainty about the data

* Corresponding author at: Department of Computer Science and Mathematics, Munich University of Applied Sciences, Lothstraße 64, 80335 München, Germany.
** Corresponding author.
E-mail addresses: marion.goedel@hm.edu (M. Gödel), nikolai.bode@bristol.ac.uk (N. Bode), gerta.koester@hm.edu (G. Köster), bungartz@in.tum.de (H.-J. Bungartz).

https://doi.org/10.1016/j.ssci.2021.105586
Received 28 June 2021; Received in revised form 6 October 2021; Accepted 10 November 2021
Available online 7 December 2021

Table 1
Key terms for calibration as they are understood in this manuscript.

Term	Description
Frequentist approach	Parameters to be calibrated are fixed but possibly unknown. The estimator for the parameter is a random variable with mean, (co-)variance and distribution.
Bayesian inference	Parameters to be calibrated are considered random variables following a probability distribution that includes any knowledge on the parameter.
Prior distribution	Probability distribution of the uncertain parameters that incorporates any information on the parameter before incorporating the observations.
Posterior distribution	Probability distribution of the uncertain parameters after the observations have been taken into account.
Calibration	The process of adjusting model parameters so that they match observations for a predefined simulation outcome.
Microscopic calibration	Observed data set used for calibration consists of microscopic quantities, often individual trajectories.
Macroscopic calibration	Observed data set used for calibration consists of observations on macroscopic quantities such as density or flow.
Distance measure	A metric used for quantitative calibration which indicates the difference between observations and the model response at a given parameter value.

set used for calibration. This enables us to handle the uncertainty in the parameter after calibration. Bayesian inference methods are not commonly employed in crowd dynamics yet and have only been used in a few publications (Bode, 2020; Bode et al., 2019; Corbetta et al., 2015; Song and Lovreglio, 2021).

In this work, we perform calibration for a bottleneck scenario. Bottlenecks are key elements in many topographies, e.g. when looking at ingress or egress situations. The constriction can lead to high densities that can potentially be harmful. Many experiments have been carried out for bottlenecks (Haghani and Sarvi, 2018) and they are often used for calibration (Hoogendoorn and Daamen, 2007; Liu et al., 2014; von Sivers, 2016). Especially the flow through bottlenecks is often of interest (Daamen and Hoogendoorn, 2010; Liao et al., 2014; Liddle et al., 2009, 2011; Rupprecht et al., 2011; Seyfried et al., 2009). For example, when corridor widths in guidelines are suggested to ensure a certain flow. To obtain reliable simulation results, we need to consider uncertainties after calibration.

We demonstrate on a bottleneck scenario how Approximate Bayesian Computation (ABC), a flexible technique to perform Bayesian model calibration, can be employed for calibrating crowd dynamics models. ABC is a likelihood-free approach which is beneficial since the likelihood of computer models is typically not known. We compare the Bayesian method to a classical point estimate in three case studies to highlight the advantages of a full posterior distribution over a single point estimate: In the first example, we show that point estimates and ABC are both suitable for calibrating a stochastic simulator when the posterior is a unimodal symmetric distribution. In the second example, ABC reveals a multi-modal posterior distribution that arises because several parameter values lead to the same value for the quantity of interest, in this case, the flow in a force-based simulator with a faster-is-slower dynamic. The point estimate, on the other hand, finds by design only a single mode. For subsequent studies, this can lead to too low egress times indicating a more efficient egress and therefore potentially harming pedestrians. In the third example, we infer multiple parameters. ABC provides a joint posterior distribution that contains additional information such as parameter correlations and how well parameters are informed by the data.

1.2. State-of-the art calibration in crowd simulation

In crowd simulation, calibration of model parameters is an essential step, yet there is no standardized approach (Lovreglio et al., 2015). As stated before, we understand calibration as a process that aims to adjust simulation response and observation based on a predefined quantity of interest by adjusting model parameters. Comparing simulation outcomes with observations is a step also taken for validation. However, the purpose of calibration and validation are different: Validation ensures that the model captures the relevant behaviours of pedestrians while calibration aims to find the best set of parameter values given data. Therefore, calibration and validation should be clearly separated, and ideally performed based on different data sets. From a machine learning perspective, calibration could be understood as the learning phase in which the model is trained to resemble labelled data well, whereas validation tests the generalization of the trained algorithm on a separate test set.

Approaches for calibration can be roughly categorized by the data used for calibration, the uncertain parameters to be calibrated, and the method. The classification into microscopic and macroscopic calibration by Schadschneider (2001) is based upon the data set: For microscopic calibration, the data set consists of individual trajectories while for macroscopic calibration, aggregate measures such as density or flow measurements form the data set. The grouping into microscopic and macroscopic stems from validation: Microscopic validation is necessary to ensure that the microscopic modelling introduces valid information about the local effects while macroscopic validation assures that the emerging patterns from observation also develop in the simulation. In crowd dynamics, microscopic and macroscopic often refer to the scale of modelling. In microscopic modelling, individual agents are simulated while macroscopic models propagate aggregate measures such as flow or density. While calibration is also relevant for macroscopic models, we focus on microscopic models in this manuscript. Therefore, the terms microscopic and macroscopic solely refer to the scale of the quantity of interest here.

Trajectories for microscopic calibration are typically obtained from video footage. Antonini et al. (2006), Berrou et al. (2007), Ko et al. (2013), Robin et al. (2009), Tang and Jia (2011) use trajectories obtained by manual annotation for calibration while Dias and Lovreglio (2018), Ruggiero et al. (2018), Hoogendoorn and Daamen (2007), Zeng et al. (2017) employ trajectory data (semi-)automatically extracted using software. For the comparison of individual trajectories, either only one agent is simulated while the others are moved according to the observed trajectories (Zeng et al., 2017), or the agents are all placed at the observed positions of the agents and then only one simulation step is simulated for all agents (Wolinski et al., 2014). The challenges of microscopic calibration lie within the definition of the distance measure (explanation in Table 1) (Guy et al., 2012; Wolinski et al., 2014) and the potentially large impact of errors in the trajectories on the calibration result (Rudloff et al., 2014). Macroscopic calibration, on the other hand, means that aggregate simulation outcomes such

as density, speed, or flow are compared (Berrou et al., 2007; Chu, 2009; Steiner et al., 2007; Wolinski et al., 2014). In this work, calibration is performed against a macroscopic quantity, the flow through a bottleneck.

When we take a look at the parameters that are calibrated, we can separate them into physical and non-physical parameters. The former are measurable parameters such as the average torso size of pedestrians in a certain population while the latter are parameters that cannot be measured directly. We argue that the free-flow speed or desired speed, which we calibrate in this study, lies in between the two extremes of physical and non-physical parameters. On the one hand, the speed of an individual can be measured and, in some experiments, participants are asked to move through a topography in order to estimate their free-flow speed. On the other hand, in experiments participants are observed and may therefore change their behaviour (observer effect) suggesting that an intrinsic free-flow speed cannot be measured without bias. When physical parameters are measured directly from available data without evaluating the model, we refer to it as direct calibration, as in Hussein and Sayed (2018), Tang and Jia (2011), Zeng et al. (2014). Indirect calibration, in contrast, requires the evaluation of the model by comparing simulation and observation, and therefore also typically requires a distance metric for quantifying the disagreement. A well-studied example for indirect calibration is the parameters associated with the interaction forces in Social Force based models (Daamen et al., 2013; Dias et al., 2018; Hoogendoorn and Daamen, 2006, 2007; Johansson et al., 2007; Seer et al., 2014a; Steiner et al., 2007; Taherifar et al., 2019; Tang and Jia, 2011; Voloshin et al., 2015; Zeng et al., 2017). These are similar to the personal space strength and obstacle repulsion which we calibrate later in this study.

Finally, we group the methods used for calibration. The biggest difference is whether the calibration is quantitative or qualitative. Qualitative calibration here means visual comparison implying manual work since it is complex to formalize visual comparison for software. While we believe manual calibration is a common approach, the literature on it is sparse. We are not aware of any publications that formalize the process of visual comparison. Within quantitative calibration, either a manual comparison or (semi-)automatized methods can be applied. Central to all of them is a metric that compares model evaluations to observations, here called distance measure. For manual quantitative calibration, typically a one-at-a-time approach is utilized where one parameter is varied while all others are fixed. This approach can easily miss global minima in the distance measure when several parameters are calibrated.

Popular automated methods are regression with least squares (Guo et al., 2012; Johansson et al., 2007; Tang and Jia, 2011; Seer et al., 2014b) and maximum likelihood estimation (Antonini et al., 2006; Campanella et al., 2011; Daamen and Hoogendoorn, 2012; Hoogendoorn and Daamen, 2006, 2007; Ko et al., 2013; Lovreglio et al., 2015; Robin et al., 2009; Zeng et al., 2014). In addition, recently a few studies have been published in which Bayesian inference methods are employed for calibration (Bode, 2020; Bode et al., 2019; Corbetta et al., 2015; Song and Lovreglio, 2021). In this paper, we compare the results of Bayesian inference to point estimation and point out the benefits and limitations of both approaches.

1.3. Bayesian inference methods for calibration

Maximum likelihood estimation as, well as regression with least squares, are examples of frequentist approaches to parameter estimation. Central to both methods is the optimization of a distance measure. For Bayesian inference, sampling methods are common. They find samples of the posterior distribution of the uncertain parameters based on a user-defined prior distribution. The prior distribution reflects the initial guess on the parameter distribution. In the absence of prior knowledge, so-called noninformative priors, close or equal to a uniform distribution, can be used. This prior distribution is then updated with the evidence inherent in available data during the inversion. The result is an informed distribution, the so-called posterior distribution. In short, sampling methods provide a full posterior distribution instead of a single point estimate. Sampling methods can be divided into likelihood-based and likelihood-free approaches. The former require a explicit likelihood function which indicates the probability of a parameter set given the data. A popular likelihood-based sampling approach are Markov chain Monte Carlo methods like Metropolis–Hasting algorithms (Metropolis et al., 1953) or Gibbs sampling (Geman and Geman, 1984). For computational models, such as individual-based crowd simulation tools, the likelihood is often unknown or intractable. Therefore, we choose a likelihood-free method, Approximate Bayesian Computation (ABC) (Tavaré et al., 1997; Beaumont et al., 2002). Bayesian approaches can also be used to obtain point estimates such as the maximum a posteriori estimator (MAP) or conditional mean (Smith, 2014).

An additional challenge for calibration is the stochasticity in most microscopic crowd simulations due to random initialization of the agent's position and speed. There are a few ways to cope with the stochasticity: either several simulations are averaged in order to reduce or remove the stochasticity in the output (Chu, 2009; Taherifar et al., 2019), or the noise is set to zero (Daamen et al., 2013), or the seed for the pseudo-random number generator is fixed which is equivalent to a single realization of the random variables. In the latter case, there are two main shortcomings. First, the single realization may deviate strongly from the average response, and second, depending on how the random numbers are drawn in the simulation, varying parameters might have an impact on more attributes than expected unless multiple seeds are used for speed, position, and so on. Approximate Bayesian Computation can cope with stochastic models so that we can avoid all workarounds.

2. Methods

2.1. Simulating flow through a bottleneck

In the following, we describe two models for the flow through a bottleneck: A model based on optimization of a utility function, the Optimal Steps Model, and an emulator for a force-based model, the Social Force Model, specifically for the bottleneck. The Python routines used for this study are publicly available.[1]

2.1.1. Bottleneck scenario

We use different data for calibration of the two models. For the calibration, we reconstruct the bottleneck experiments performed by Seyfried et al. (2009) in the Vadere framework (Kleinmeier et al., 2019). In these experiments, the flow of pedestrians was studied in 18 runs in which the width of the bottleneck was varied between 0.8 m and 1.2 m (in 0.1 m increments) and the number of participants was 20, 40, and 60 participants. Fig. 1 shows the scenario setup used for the simulation. The width of the bottleneck increases from 0.8 m on the left to 1.2 m on the right according to the experiment.

In Seyfried et al. (2009) the flow of the pedestrians through the bottleneck is measured by

$$J = \frac{\Delta N}{\Delta t} \tag{1}$$

[1] Scripts for parameter estimation with Approximate Bayesian Computation and point estimation:
uq/inversion/run_inversion_ABC_bottleneck_1d.py (case study 1),
uq/inversion/run_inv_ABC_bottleneck_SFMemulator.py (case study 2),
uq/inversion/run_inversion_ABC_bottleneck_3d.py (case study 3),
Scripts for propagation:
uq/propagation/run_propagation_abc_vs_pe_OSM.py (case study 1),
uq/propagation/run_propagation_abc_vs_pe_SFMEmulator.py (case study 2)
available at https://gitlab.lrz.de/vadere/uncertainty-quantification/.

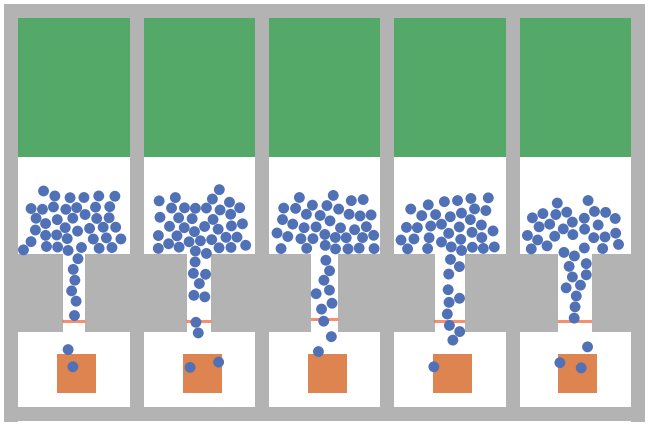


Fig. 1. Snapshot of simulation of bottleneck scenario at 12 s. Bottleneck width increases from 0.8 m (left) to 1.2 m (right). The flow through the bottleneck is measured at the end of the bottleneck (red line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

where ΔN is the number of participants in the experiments and Δt is the time difference between the last and the first participant crossing the measurement line. The measurement line is placed about one metre before the end of the constriction. We measure the flow in the simulation in the same manner. In the experimental setup, for 60 participants the flow measurements were $d_1 = [1.288, 1.674, 1.900, 2.123, 2.364]\,1/s$. We use this data set to calibrate the Optimal Steps Model.

For the calibration of the second model, the Social Force Model emulator, we use a different data set, as described in more detail in Section 2.1.3 after this model is introduced.

2.1.2. Optimal Steps Model

The Optimal Steps Model (OSM) (Kleinmeier et al., 2019; Seitz and Köster, 2012; von Sivers and Köster, 2015) is a microscopic crowd dynamics model. The basis of the OSM is a balance of goals: Reaching the next geographical target while keeping a certain distance to obstacles and other agents. Each agent has information about the topography in the form of a navigational field that encodes the geodesic distance of each point in the topography to the agent's next target. The distance to the next target is obtained by solving the eikonal equation on a regular mesh. The navigational field represents the utility of each position. Obstacles and other agents are represented by dips in the utility. The total utility includes information for all three goals.

The OSM is an agent-based model (ABM) since each agent has individual attributes. One common attribute in ABMs for crowd simulation is the free-flow speed at which an agent moves when no obstacles or other agents are present. In addition, each agent is assigned an individual stepping frequency. This frequency is used for the event-driven update of the model: For each agent, stepping events are registered in an event queue according to its stepping frequency. At each stepping event, the agent finds its next position by maximizing the utility. We use the OSM implementation from the Vadere framework, a more detailed description can be found in Kleinmeier et al. (2019).

2.1.3. Social Force Model emulator for the bottleneck

In addition to the simulations of the bottleneck scenario with the Optimal Steps Model, we demonstrate calibration with Bayesian inference using an approximation based on a Social Force Model. Helbing et al. (2000) present a Social Force Model based simulation of a bottleneck that shows a faster-is-slower dynamic which means that multiple free-flow speeds lead to the same flow. Instead of implementing the model, we use the output presented in Helbing et al. (2000). We perform a cubic interpolation on the data shown for the relationship between the desired speed and specific flow. This interpolation is used to construct our emulator. We derive the flow for the five different bottleneck widths from the specific flow. The desired speed in the Social Force Model is equivalent to the free-flow speed in the Optimal Steps Model. In addition, we add a Gaussian noise $\mathcal{N}(0, 0.01)$ to the model. The resulting relationship between free-flow speed and flow can be seen in Fig. 2. Fig. 2(b) also shows the data set used for calibrating the model. For calibrating the Social Force Model emulator, we use the artificial data set $d_2 = [0.974, 1.105, 1.235, 1.323, 1.486]\,1/s$.

Analogously, we create an emulator for the relationship between desired speed and egress time, named leaving times in Helbing et al. (2000), which we use for one exemplary propagation (we refer to simulations based on calibrated parameter values as propagation). Based on the data presented in Helbing et al. (2000), we perform cubic interpolation and add a zero-mean Gaussian noise with variance $\sigma^2 = 25$. The resulting emulator is presented in Fig. 3, the cubic interpolation in Fig. 3(a) and the emulator including the noise term in Fig. 3(b).

2.2. Methods and measures for calibration

We define the inverse problem as the task to identify the parameter x of a model f given data d for a defined norm $\|\cdot\|$ such that

$$d = f(x) + \eta \tag{2}$$

in the presence of noise η which can be measurement noise or model error, or both.

We use two techniques for the estimation of the parameter: Approximate Bayesian Computation (ABC) (Tavaré et al., 1997) and point estimation. In the Bayesian sense, the result of parameter estimation is an informed and therefore updated posterior distribution $p(x|d)$ from a given prior distribution $p(x)$, an initial guess about the uncertain parameters, based on the observations. That means the result of the inversion is not only the best value for parameter x, but a distribution for x. Bayes' theorem

$$p(x|d) = \frac{p(x)p(d|x)}{p(d)} \propto p(x) \cdot p(d|x)$$

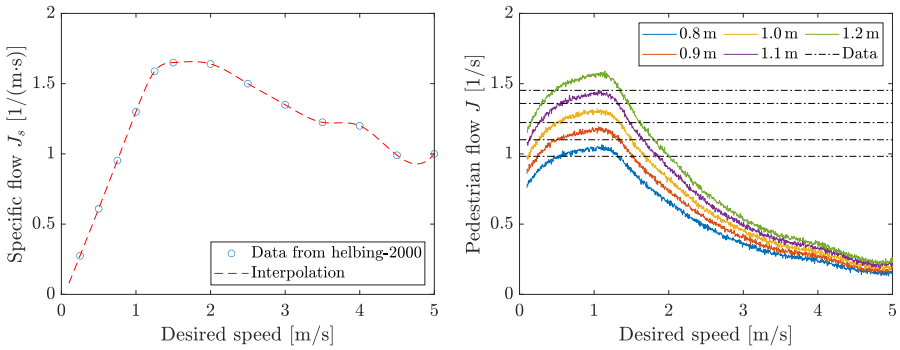
states that the posterior distribution $p(x|d)$ is proportional to the product of prior distribution $p(x)$ and the likelihood $p(d|x)$ without the evidence $p(d)$. The likelihood at x_0, $p(d|x = x_0)$, describes the probability of observing the given data d at parameter value x_0. The likelihood is often unknown in applications since it reflects the distributions of error which can usually only be assumed.

2.2.1. Point estimate

As a reference, we use a point estimate. Point estimation such as maximum likelihood estimation has already been used in pedestrian dynamics for parameter calibration (Daamen and Hoogendoorn, 2012; Hoogendoorn and Daamen, 2006, 2007; Ko et al., 2013; Lovreglio et al., 2015). While Bayesian inference methods such as ABC obtain a posterior distribution of the uncertain parameter, point estimation provides only the single best parameter value. We derive the point estimate from the model evaluations obtained during ABC rejection: We choose the parameter value x_c closest to the data in terms of the distance measure , f_d, the posterior mode $x_{pe} = \arg\min_{x_c} f_d(x_c)$. This is known as the maximum a posteriori (MAP) estimate.

2.2.2. Inversion method: Approximate Bayesian computation (ABC)

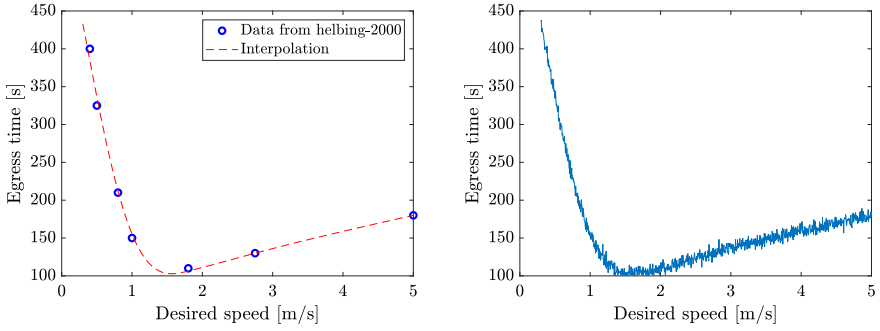
Approximate Bayesian Computation is a so-called likelihood-free method i.e. the likelihood is calculated implicitly instead of being explicitly fed to the method. This method is often applied to ecological or biological models (Beaumont, 2010). ABC is suited well for computer models for which likelihood functions are not available and in which stochastic terms might be present. ABC requires prior distributions for the uncertain input parameters, a distance measure $f_d(x)$ and a tolerance ϵ.



(a) Interpolation of specific flow data from Helbing et al. (2000).

(b) Emulator for widths from 0.8 m to 1.2 m (including noise).

Fig. 2. Social Force emulator for five different bottleneck widths, constructed with flow data from Helbing et al. (2000) including an additive zero-mean Gaussian noise with variance $\sigma^2 = 10^{-4}$. In (b), additionally, the data used for the calibration is shown.



(a) Interpolation of egress times from Helbing et al. (2000).

(b) Emulator including Gaussian noise.

Fig. 3. Social Force emulator for leaving times or egress times for a room with 200 people, constructed with egress times from Helbing et al. (2000) including an additive zero-mean Gaussian noise with variance $\sigma^2 = 25$.

There are several sampling strategies for ABC (Toni et al., 2009). We choose ABC rejection because it has a lower number of parameters, and is easy to implement and parallelize since it is not an iterative method. In the rejection sampler, candidates x_c are drawn based on the prior distribution $p(x)$. At each candidate, the distance measure is computed by evaluating the underlying model at the candidate $f(x_c)$. The distance measure describes the distance between the candidate and the data d. We use an Euclidean distance measure $f_d(x_c) = \|d - f(x_c)\|_2^2$. Candidates for which $f_d(x_c) = 0$ are samples of the posterior. We approximate the posterior by keeping all candidates for which the distance measure is lower than the tolerance, $f_d(x_c) < \epsilon$, as posterior samples $x_s \sim \hat{p}(x|d)$. This weaker requirement allows us to reduce the number of candidates N and therefore the number of model evaluations. The approximation is asymptotically exact given a suitable distance measure i.e. for $\epsilon \to 0$ and $N \to \infty$ we obtain the true posterior $p(x|d)$.

2.2.3. Measuring the residual uncertainty after calibration

The result of Bayesian inference is a posterior distribution of the uncertain parameters. This distribution is a basis for subsequent studies to quantify the uncertainty in the simulation outcome. The impact of the variation in the posterior is evaluated by looking at the distribution of the quantity of interest.

We analyse the remaining uncertainty in the parameter after calibration by propagating the posterior from Bayesian inference and the point estimate, respectively. Propagation of a posterior distribution means evaluating the model f at the posterior samples obtained with ABC rejection. In the case of the point estimate, we obtain the best parameter value x_{pe} at which we run the simulation to obtain $f(x_{pe})$. We use the M posterior samples obtained with ABC and propagate these through the model. In addition, we evaluate our model M times at the point estimate. Since there is some stochasticity in the model, more specifically in its initialization, multiple evaluations at the same parameter value yield different outcomes. As a result, we obtain $5 \cdot M$ flow values for the five bottleneck widths for each method.

For the first case study, we fit a generalized linear model to the flow results, separately for the posterior samples and the point estimate. The confidence interval of the fitted slope parameter is then used as a measure of the variation. A large confidence interval indicates a large variation after calibration and vice versa.

3. Results and discussion

We perform calibration with Approximate Bayesian Computation and the point estimate in three cases: First, we calibrate the free-flow speed in the Optimal Steps Model for five bottlenecks against flow measurements from an experiment. Second, we infer the desired speed in the Social Force Model emulator from artificial flow measurements. Finally, we use Bayesian inference and point estimation to estimate free-flow speed, personal space strength, and obstacle repulsion in the Optimal Steps Model from the flow measured in an experiment. Bayesian inference provides a set of posterior samples. For each use case, we compare the histogram of the posterior samples to the point estimate. The prior distributions used for Bayesian inference are listed in Table 2. The parameters of the uniform prior distributions were determined as follows. The range for the free-flow speed mean is based on Weidmann (1993, p. 83). Both obstacle repulsion and personal space strength are non-physical parameters, which complicates the choice of the prior. The intervals were found by manual pre-calibration. For both, we choose large intervals in order to assure that there is no posterior density outside the chosen interval. Bayesian inference can only find posterior density in areas where the prior is non-zero.

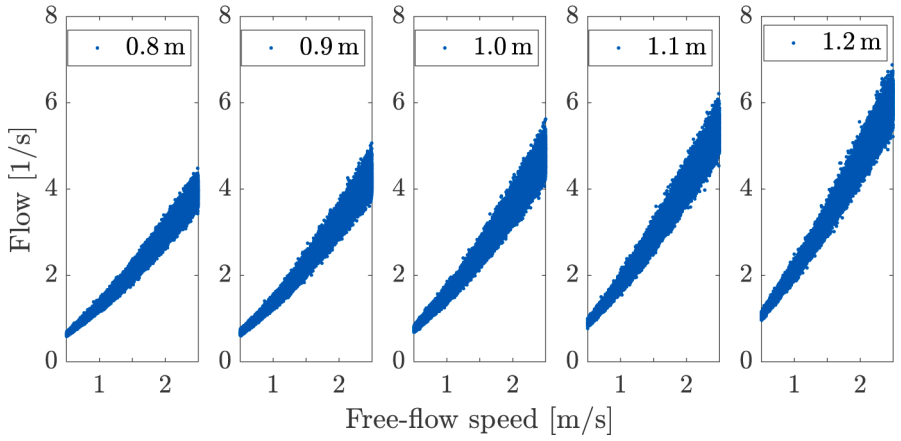


Fig. 4. Relationship between free-flow speed mean and flow for the five bottlenecks with different widths (10^5 model evaluations). Bottleneck width increases from 0.8 m (left) to 1.2 m (right).

Table 2
Prior distributions for the uncertain parameters for the three case studies.

Section	Model	Parameter	Prior
3.1	Optimal Steps Model	Free-flow speed mean	$\mathcal{U}(0.5, 2.5)$
3.2	Social Force Model emulator	Desired speed	$\mathcal{U}(0.5, 2.5)$
		Free-flow speed mean	$\mathcal{U}(0.5, 2.5)$
3.3	Optimal Steps Model	Obstacle repulsion	$\mathcal{U}(1.0, 12.0)$
		Personal space strength	$\mathcal{U}(1.0, 12.0)$

Table 3
Confidence intervals for the slope of linear regressions for observations, and propagated point estimate and full ABC posterior.

	Data	Point estimate	ABC posterior
Confidence interval (CI)	[2.251, 2.951]	[1.651, 2.990]	[1.595, 3.080]
Size of CI	0.699	1.338	1.485

3.1. Unimodal posterior (Optimal Steps Model): point estimate is sufficient

For the first case study, we simulate the five bottlenecks of widths from 0.8 m to 1.2 m in Fig. 1 with the Optimal Steps Model. Fig. 4 shows the monotonically increasing input–output relationship between the uncertain parameter, the free-flow speed, and the quantity of interest, the flow. We observe that, as expected, the flow increases with increasing speed and also with increasing width of the bottleneck. In addition, the model evaluations show the stochasticity in the system. Repeated evaluations for a given parameter value yield different results. The size of the variation in the quantity of interest, the flow, increases with the free-flow speed mean.

ABC is based on the distance measure which evaluates the distance between a model evaluation at a candidate for the uncertain parameter and the observations. In Fig. 5(a), the distance measure is shown for the five bottlenecks simulated with the OSM. Since the trend of the input–output relationship is strictly monotonically increasing for each bottleneck width, the distance measure for calibrating a single bottleneck must have a single minimum. We choose a tolerance of $\epsilon = 0.0685$ so that we keep 1% of the candidates as samples of the posterior (Beaumont et al., 2002). Fig. 5(b) is a histogram of the posterior samples. Their mean and mode are about 1.14 m/s and their standard deviation is 0.027 m/s.

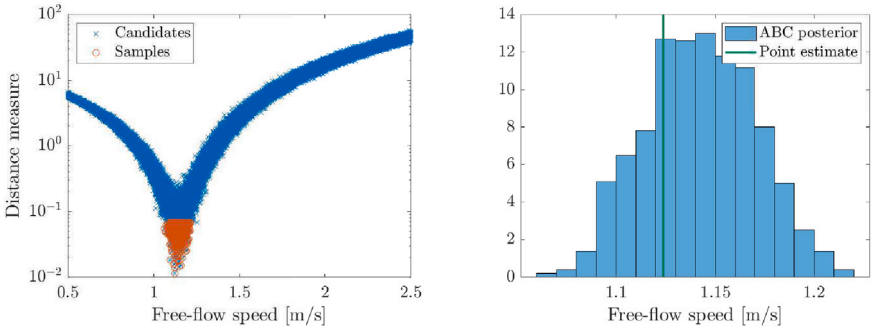
We now use the posterior obtained from ABC and propagate it through the system to quantify the uncertainty in the simulation outcome. In addition, we propagate the point estimate through the system. Since our model has some stochasticity in the initialization process, we perform the same number of propagations for both the ABC posterior and the point estimate. Fig. 6 depicts the histogram of the flow for each bottleneck. Even though we use a single parameter value for the point estimate, there is a significant variation in the flow values obtained from the simulations. The width of the histogram of the flow samples after propagating the point estimate 1000 times is similar to the width of the histogram of flow samples obtained from propagating the full ABC posterior.

The histograms in Fig. 6 also reveal that the propagation of both, the point estimate and the ABC posterior, deviate from the observed data set when adapting to all five bottlenecks at once. This could have several reasons: First, the relationship between bottleneck width and flow may differ between experiments and simulation. Second, the measurement error in the observational data is so large that it affects the relationship between bottleneck width and flow significantly. Third, there may be underlying effects in the experiments such as a learning effect or a shared identity among the participants that are not reflected in the simulation, leading to slightly different dynamics.

In order to evaluate the variation in the distribution of the quantity of interest for all five bottlenecks, we calculate the confidence intervals of the slope of the linear fit as a metric for the variation. For comparability of the confidence intervals for the observed data set, ABC posterior and point estimate, we perform a linear fit through the five flow values. For point estimate and ABC posterior, we perform a fixed number of $N_{av} = 200$ linear fits to a set of five randomly selected flows, one for each width. For the experimental data set, we perform a single fit, as there is only one data set. For each fit, we calculate the confidence interval of the estimated slope parameter. Then, we average the confidence intervals over all iterations and present the average confidence interval. In Fig. 7, the regression is shown exemplarily for the propagated ABC posterior.

When we evaluate the size of the confidence interval of the slope from the regression for different tolerances ϵ, as in Fig. 8, the results depend strongly on the tolerance. The width of the ABC posterior and therefore the width of the histogram of the flow samples after propagating the posterior depend highly on the tolerance ϵ. According to theory, with infinitely small tolerance and infinite sample size N, we obtain the true posterior. In reality, however, the infinite sample size is not feasible and the tolerance is at least limited by the numerical precision. When using real data for the calibration, there may also be an error between the best model evaluation and the data. In this case, the choice of the tolerance is even more restricted and the optimal tolerance is not clear (Alahmadi et al., 2020).

We compare the results for a tolerance of $\epsilon = 0.0685$ for ABC to the results using the point estimate and the data in Table 3. The size of the confidence interval using the point estimate and the full posterior are of the same magnitude, with the ABC confidence interval being a bit larger. Both are significantly larger than the confidence interval



(a) Distance measure. (b) Histogram of posterior samples.

Fig. 5. Results of Bayesian inference with ABC threshold $\epsilon = 0.068$ (acceptance rate = 1%).

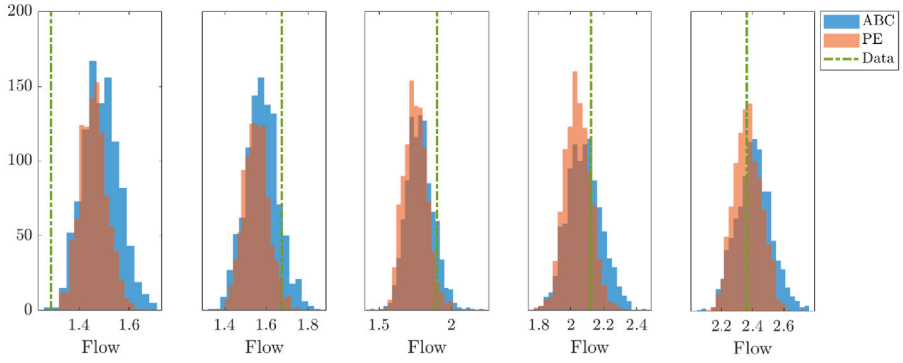


Fig. 6. Histogram of flow values obtained from the propagation of point estimate (PE, 1000 repetitions) and ABC posterior (ABC, $\epsilon = 0.0685$). Bottleneck width increases from 0.8 m (left) to 1.2 m (right).

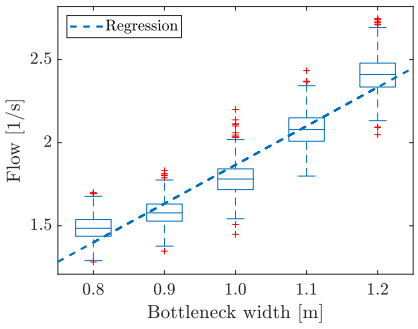


Fig. 7. Regression for propagated ABC posterior ($\epsilon = 0.0685$). A box plot of the flow samples resulting from propagation is shown together with the linear regression.

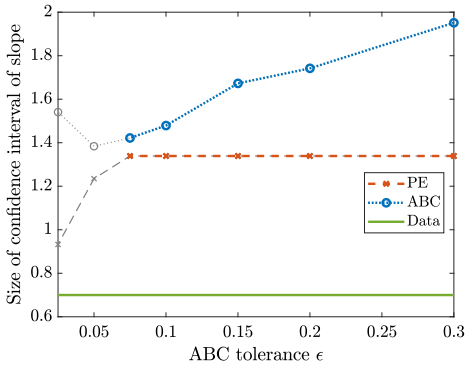


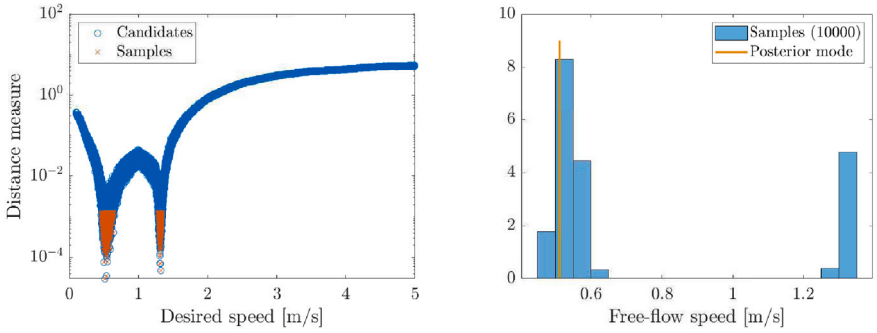
Fig. 8. Size of the confidence interval of slope (linear fit) for different ABC tolerances (ϵ). For tolerances below 0.0685, less than 1000 posterior samples are propagated (grey markers).

obtained from the data. This should always be the case. Otherwise, the simulation with the calibrated parameter would present a lower variance than the observations used for the calibration.

In summary, for this monotonically increasing relationship between the uncertain input parameter and quantity of interest through a stochastic simulator, the results of ABC and the point estimate are similar after propagation. Therefore, in these cases, a point estimate is sufficient. For practical purposes, this is advantageous, because Bayesian inference methods are computationally expensive. However, the next two subsections demonstrate, in which cases it can be problematic to rely on point estimates only.

3.2. *Bimodal posterior (Social Force Model emulator): ABC reveals bimodal posterior*

For the second case study, we focus on a setup of five bottlenecks simulated with a Social Force Model emulator. The emulator is built for an egress scenario in which 200 pedestrians exit a room. The SFM simulations exhibit a faster-is-slower phenomenon: the flow increases with the desired speed up to a certain point (here about 1.5 m/s) and then decreases. This effect is also captured by the emulator. That means the emulator function that describes the relationship between desired speed and flow is not bijective, but only surjective - a value



(a) Distance measure (sum of squares). (b) Posterior obtained with ABC.

Fig. 9. Results of Bayesian inference with ABC tolerance $\epsilon = 0.00128$.

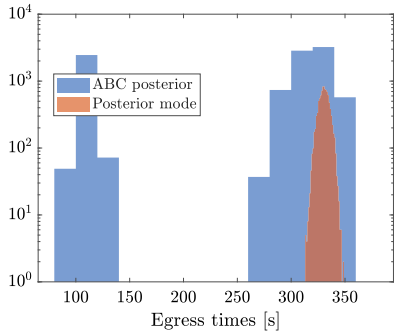


Fig. 10. Egress times when propagating the point estimate and the full ABC posterior in the Social Force Model emulator for egress times.

of the function can be reached by multiple parameter values. Hence, the parameter is not identifiable and can therefore not be uniquely estimated.

In order to calibrate the desired speed to a set of five flow measurements from the emulator, we calculate the distance measure for the SFM emulator as shown in Fig. 9. The distance measure displays two local minima corresponding to the two intersections between the data points and the model evaluations in Fig. 2(b). We choose the tolerance $\epsilon = 0.00128$ so that we keep 1% of the candidates as posterior samples.

Bayesian inference reveals a bimodal posterior distribution while the point estimate, by design, finds only one of the two parameter locations (compare Fig. 9(b)). The impact of this limitation becomes clear when we perform simulations based on the calibration. In this case, when we propagate the single point estimate or the bimodal posterior distribution through the model, we observe a similar distribution of our original quantity of interest, the flow. Then, the difference between point estimate and full ABC posterior depends mainly on the choice of the tolerance ϵ, as in the previous example with the unimodal distribution. The variation in the distribution of the flow values after propagation is mainly driven by the stochasticity in the model.

However, if we evaluate another quantity of interest, the situation changes profoundly: The distribution of the quantity of interest for the point estimate may differ drastically from the one produced by propagating a full ABC posterior. As an example, we use the second emulator for the egress time presented in Section 2.1.3. The egress time is the time that it takes for 200 pedestrians to leave the room. If we evaluate the egress times for the full ABC posterior, they vary between about 110 s and 330 s, whereas for the point estimate we obtain only about 330 s (compare Fig. 10). In this case, we obtain a conservative estimate for the egress times. If the point estimate finds the other extremum at about 1.2 m/s, leaving times would vary around 110 s indicating too efficient egress compared to the full posterior.

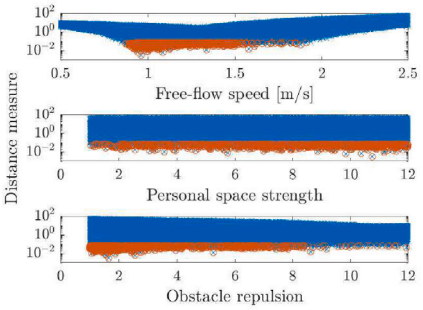
This example highlights that Bayesian inference methods are suitable to capture even unusual posterior distributions of the uncertain parameters, such as bimodal distributions, where point estimates must fail by design. In our experience, calibration is typically performed against one quantity of interest, often a flow–density relationship, and then the calibrated parameters are used for various studies. In this case, the results after calibration through point estimation may give a false sense of certainty.

3.3. Multivariate posterior (Optimal Steps Model): ABC reveals parameter relations

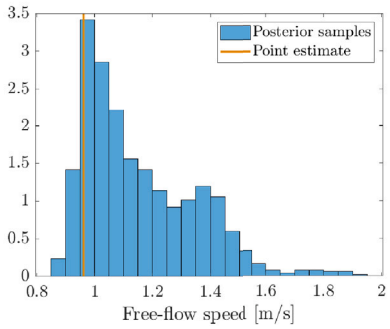
Finally, we discuss an example for a multidimensional calibration. As in the first case study, we use the Optimal Steps Model to simulate five bottlenecks of different widths ranging from 0.8 m to 1.2 m. In addition to the free-flow speed, we also calibrate the personal space strength that reflects the natural distance that two pedestrians who are not in a group naturally keep from each other. As the third parameter, we calibrate the obstacle repulsion that assures that the agents keep a certain distance to walls. Again, we use the flow measurements from the bottleneck experiment described in Seyfried et al. (2009).

First, we evaluate the distance measure which is a scalar function that depends on the three uncertain parameters. In Fig. 11(a), we show the evaluations of the distance measure against each parameter. These evaluations were obtained during the calibration, which means that all parameters are varied simultaneously which causes a large variation. Based on the distance measure, we set the tolerance to $\epsilon = 0.07342246$ so that we keep 1% of the candidates as posterior samples. We obtain a three-dimensional posterior distribution. In Fig. 11, the univariate marginal distribution that is the distribution of a single variable is shown for each parameter together with the point estimate for the parameter. As for the distance measure, the multivariate posterior has too many dimensions for a single plot.

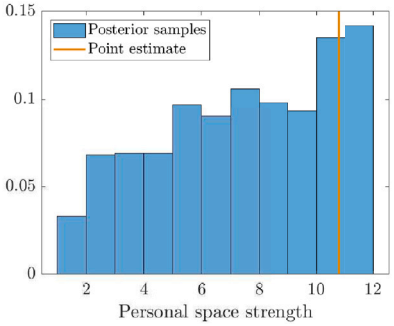
Additionally, we evaluate the bivariate marginal posterior distributions (Fig. 12). These are the joint distributions of two out of the three parameters. The shape of the univariate posterior distributions for the free-flow speed (Fig. 11(b)) and the obstacle repulsion (Fig. 11(d)) vary significantly from their uniform prior distributions. We conclude that these two parameters are well informed by the data. For the third parameter, the personal space strength, however, the posterior distribution (Fig. 11(c)) is closer to its prior distribution and hence the parameter is less informed which also means that it has less impact regarding this data set. There is a strong correlation between the first and third parameter, the free-flow speed and the obstacle repulsion, concerning the data (Fig. 12(b)). Both parameters have a clear impact on the flow: The flow increases with the free-flow speed and it decreases with increasing obstacle repulsion. With a higher obstacle repulsion, pedestrians keep a higher distance to the walls. Consequently, the effective bottleneck width decreases with increasing obstacle repulsion. The correlation between the parameters shows that



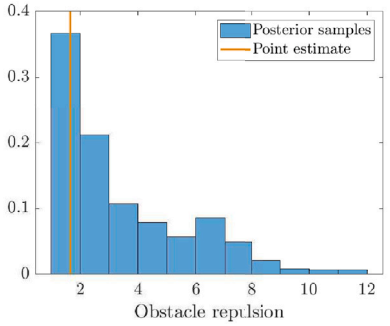
(a) Distance measure at candidates.



(b) Univariate posterior for free-flow speed.

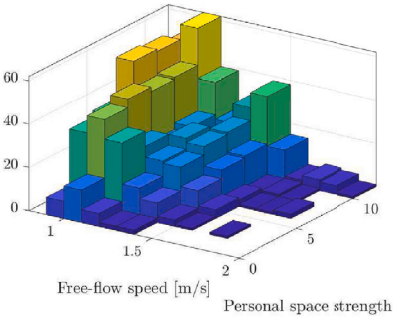


(c) Univariate posterior for personal space strength.

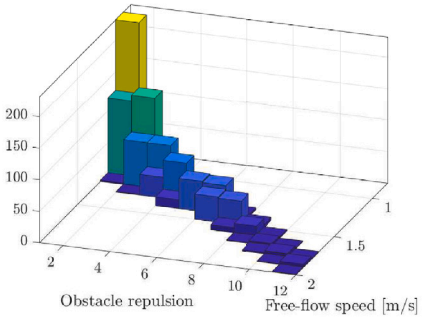


(d) Univariate posterior for obstacle repulsion.

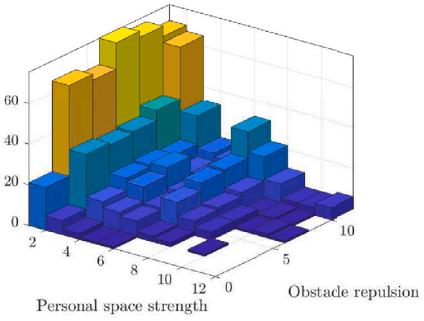
Fig. 11. Results of Bayesian inference with ABC threshold of 0.07342246 (acceptance rate 1%).



(a) Bivariate posterior for free-flow speed and personal space strength (Pearson correlation coefficient: 0.163).



(b) Bivariate posterior for free-flow speed and obstacle repulsion (Pearson correlation coefficient: 0.978).



(c) Bivariate posterior for personal space strength and obstacle repulsion (Pearson correlation coefficient: 0.217).

Fig. 12. Bivariate posterior histograms obtained with ABC ($\epsilon = 0.07342246$).

the data is best represented when the parameter value for the obstacle repulsion is not chosen independently from the free-flow speed but ideally when both are small at the same time.

Note that the basis for ABC is the distance measure which compares all five data points of the experiments, corresponding to five bottleneck widths, with simulations. Consequently, all evaluations are relative to the distance measure. This makes it more difficult to draw conclusions on the parameters. It is also important to underline that the parameter relations do not hold in general, but only with respect to the particular data set we used for calibration.

This example is intended as a demonstration of principle. In our example, we use five measures of the flow to calibrate three parameters. We suggest using more data in practical applications for reliable calibration.

Often experimental data is used to calibrate parameters that are subsequently transferred to a different setting. Our results highlight that this approach is limited by the amount of information contained in the data. In the last case study, the personal space strength is not well informed. A safety engineer would be ill-advised to apply the personal space strength parameter obtained from the five bottlenecks to a safety scenario where one expects a strong influence of people's need for personal space.

4. Conclusion and outlook

In this work, we showed the benefits and limitations of using Bayesian inference methods for calibrating pedestrian simulation models to empirical data at the example of Approximate Bayesian Computation. We chose a bottleneck scenario because of its importance to safety engineering and the availability of experimental data for the calibration.

First, we calibrated the free-flow speed parameter of the Optimal Steps Model. We compared the results to a common approach for calibration, the point estimate. When propagating both, the point estimate and the full posterior through the model, the variation in the quantity of interest depended mainly on a central parameter of ABC, the user-defined tolerance ϵ. Qualitatively, there was no significant difference between point estimate and Bayesian inference and therefore a point estimate appeared sufficient.

Second, we adapted the desired speed in a Social Force emulator in an egress scenario with a faster-is-slower dynamic. Bayesian inference revealed a bimodal posterior distribution that the point estimate could not recover. Consequently, a part of the posterior distribution was lost when we relied on point estimation. We evaluated the egress times for the ABC posterior and the point estimate to emphasize the importance of considering the full posterior distribution for subsequent studies. For the point estimate, the egress times only varied around a single value instead of displaying the correct bimodal shape. This means that the physics of safety is misjudged, that egress times are miscalculated, and, even worse, that a harmful side effect of rushing evacuees may be completely overlooked by the safety engineer in charge.

Finally, we performed a multidimensional calibration with three parameters, the free-flow speed, the personal space strength, and the obstacle repulsion in the Optimal Steps Model. The uni- and bivariate posterior distribution obtained with ABC showed that two parameters, free-flow speed and obstacle repulsion, were well informed by the data, while the third parameter, the personal space strength, was not informed to the same degree. This exposes a need for caution when transferring calibrated parameters from one safety scenario to the next. In addition, the multivariate posterior revealed a strong correlation between free-flow speed and obstacle repulsion.

In summary, our results show that using Bayesian inference methods for calibration make the application of crowd models more reliable even though Bayesian inference is computationally more demanding than point estimation and the choice of prior distributions for parameters as well as the choice of tolerances can be complex. In practice, noninformative priors are often assumed due to a lack of knowledge. Contrary to point estimation, Bayesian inference can unveil the shape of the posterior, even from a noninformative prior distribution. When the posterior distribution is symmetric and unimodal, a point estimate may be sufficient, provided that models include a stochastic component. Nevertheless, if there is reason to doubt that the posterior is symmetric and unimodal, we recommend performing Bayesian inference for calibration. In general, the posterior distributions obtained through Bayesian inference contain additional information about the parameter and the calibration process that are pertinent to correctly assessing safety risks and thus need to be considered for subsequent studies.

CRediT authorship contribution statement

Marion Gödel: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration, Funding acquisition. **Nikolai Bode:** Conceptualization, Methodology, Writing – review & editing. **Gerta Köster:** Writing – review & editing, Supervision, Funding acquisition. **Hans-Joachim Bungartz:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

M.G. would like to thank the German Academic Exchange Service (DAAD) and the research office FORWIN at Munich University of Applied Sciences for providing funding for the research stay at University of Bristol for the collaboration with N.B. All authors have read and agreed to the published version of the manuscript.

Funding

M.G. was supported by the German Federal Ministry of Education and Research through the project S^2 UCRE (grant no. 13N14463). G.K. was supported by the German Federal Ministry of Education and Research through the project OPMoPS (grant no. 13N14562). The APC was funded by the research office FORWIN at Munich University of Applied Sciences HM.

References

Alahmadi, A.A., Flegg, J.A., Cochrane, D.G., Drovandi, C.C., Keith, J.M., 2020. A comparison of approximate versus exact techniques for Bayesian parameter inference in nonlinear ordinary differential equation models. R. Soc. Open Sci. 7 (3), 191315. http://dx.doi.org/10.1098/rsos.191315.

Antonini, G., Bierlaire, M., Weber, M., 2006. Discrete choice models of pedestrian walking behavior. Transp. Res. B 40 (8), 667–687. http://dx.doi.org/10.1016/j.trb.2005.09.006.

Beaumont, M.A., 2010. Approximate Bayesian computation in evolution and ecology. Annu. Rev. Ecol. Evol. Syst. 41 (1), 379–406. http://dx.doi.org/10.1146/annurev-ecolsys-102209-144621.

Beaumont, M.A., Zhang, W., Balding, D.J., 2002. Approximate Bayesian computation in population genetics. Genetics 162 (4), 2025–2035, URL: https://www.genetics.org/content/162/4/2025.

Berrou, J.L., Beecham, J., Quaglia, P., Kagarlis, M.A., Gerodimos, A., 2007. Calibration and validation of the legion simulation model using empirical data. In: Waldau, N., Gattermann, P., Knoflacher, H., Schreckenberg, M. (Eds.), Pedestrian and Evacuation Dynamics 2005. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 167–181.

Bode, N., 2020. Parameter calibration in crowd simulation models using approximate Bayesian computation. Collect. Dyn. 5, http://dx.doi.org/10.17815/cd.2020.68.

Bode, N.W., Chraibi, M., Holl, S., 2019. The emergence of macroscopic interactions between intersecting pedestrian streams. Transp. Res. B 119, 197–210. http://dx.doi.org/10.1016/j.trb.2018.12.002.

Campanella, M.C., P.Hoogendoorn, S., Daamen, W., 2011. A methodology to calibrate pedestrian walker models using multiple-objectives. In: Peacock, R.D., Kuligowski, E.D., Averill, J.D. (Eds.), Pedestrian and Evacuation Dynamics. Springer US, Boston, MA, pp. 755–759.

Chu, C.-Y., 2009. A computer model for selecting facility evacuation design using cellular automata. Comput.-Aided Civ. Infrastruct. Eng. 24 (8), 608–622. http://dx.doi.org/10.1111/j.1467-8667.2009.00619.x.

Corbetta, A., Muntean, A., Vafayi, K., 2015. Parameter estimation of social forces in pedestrian dynamics models via a probabilistic method. Math. Biosci. Eng. 12 (2), http://dx.doi.org/10.3934/mbe.2015.12.337.

Daamen, W., Campanella, M., Hoogendoorn, S.P., 2013. Calibration of nomad parameters using empirical data. In: Kozlov, V.V., Buslaev, A.P., Bugaev, A.S., Yashina, M.V., Schadschneider, A., Schreckenberg, M. (Eds.), Traffic and Granular Flow '11. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 109–120.

Daamen, W., Hoogendoorn, S., 2010. Capacity of doors during evacuation conditions. Procedia Eng. 3, 53–66. http://dx.doi.org/10.1016/j.proeng.2010.07.007, 1st Conference on Evacuation Modeling and Management.

Daamen, W., Hoogendoorn, S., 2012. Calibration of pedestrian simulation model for emergency doors by pedestrian type. Transp. Res. Rec. 2316 (1), 69–75. http://dx.doi.org/10.3141/2316-08.

Dias, C., Iryo-Asano, M., Nishiuchi, H., Todoroki, T., 2018. Calibrating a social force based model for simulating personal mobility vehicles and pedestrian mixed traffic. Simul. Model. Pract. Theory 87, 395–411. http://dx.doi.org/10.1016/j.simpat.2018.08.002, URL: https://www.sciencedirect.com/science/article/pii/S1569190X18301096.

Dias, C., Lovreglio, R., 2018. Calibrating cellular automaton models for pedestrians walking through corners. Phys. Lett. A 382 (19), 1255–1261. http://dx.doi.org/10.1016/j.physleta.2018.03.022.

Geman, S., Geman, D., 1984. Stochastic relaxation, gibbs distributions, and the Bayesian restoration of images. IEEE Trans. Pattern Anal. Mach. Intell. PAMI-6 (6), 721–741. http://dx.doi.org/10.1109/TPAMI.1984.4767596.

Guo, R.-Y., Huang, H.-J., Wong, S., 2012. Route choice in pedestrian evacuation under conditions of good and zero visibility: Experimental and simulation results. Transp. Res. B 46 (6), 669–686. http://dx.doi.org/10.1016/j.trb.2012.01.002.

Guy, S.J., van den Berg, J., Liu, W., Lau, R., Lin, M.C., Manocha, D., 2012. A statistical similarity measure for aggregate crowd dynamics. ACM Trans. Graph. 31 (6), 190:1–190:11. http://dx.doi.org/10.1145/2366145.2366209.

Haghani, M., Sarvi, M., 2018. Crowd behaviour and motion: Empirical methods. Transp. Res. B 107, 253–294. http://dx.doi.org/10.1016/j.trb.2017.06.017.

Helbing, D., Farkas, I., Vicsek, T., 2000. Simulating dynamical features of escape panic. Nature 407, 487–490. http://dx.doi.org/10.1038/35035023.

Hoogendoorn, S.P., Daamen, W., 2006. Microscopic parameter identification of pedestrian models and implications for pedestrian flow modeling. Transp. Res. Rec. 1982 (1), 57–64. http://dx.doi.org/10.1177/0361198106198200108.

Hoogendoorn, S.P., Daamen, W., 2007. Microscopic calibration and validation of pedestrian models: Cross-comparison of models using experimental data. In: Schadschneider, A., Pöschel, T., Kühne, R., Schreckenberg, M., Wolf, D.E. (Eds.), Traffic and Granular Flow'05. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 329–340.

Hussein, M., Sayed, T., 2018. A methodology for the microscopic calibration of agent-based pedestrian simulation models. In: 2018 21st International Conference on Intelligent Transportation Systems. ITSC, pp. 3773–3778. http://dx.doi.org/10.1109/ITSC.2018.8569395.

Johansson, A., Helbing, D., Shukla, P., 2007. Specification of the social force pedestrian model by evolutionary adjustment to video tracking data. Adv. Complex Syst. 10, 271–288. http://dx.doi.org/10.1142/S0219525907001355.

Kleinmeier, B., Zönnchen, B., Gödel, M., Köster, G., 2019. Vadere: An open-source simulation framework to promote interdisciplinary understanding. Collect. Dyn. 4, http://dx.doi.org/10.17815/CD.2019.21.

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

Liao, W., Seyfried, A., Zhang, J., Boltes, M., Zheng, X., Zhao, Y., 2014. Experimental study on pedestrian flow through wide bottleneck. Transp. Res. Procedia 2, 26–33. http://dx.doi.org/10.1016/j.trpro.2014.09.005, The Conference on Pedestrian and Evacuation Dynamics 2014 (PED 2014), 22-24 October 2014, Delft, The Netherlands.

Liddle, J., Seyfried, A., Klingsch, W., Rupprecht, T., Schadschneider, A., Winkens, A., 2009. An experimental study of pedestrian congestions: Influence of bottleneck width and length. arXiv 0911.4350 (v2), URL: http://arxiv.org/abs/0911.4350.

Liddle, J., Seyfried, A., Steffen, B., Klingsch, W., Rupprecht, T., Winkens, A., Boltes, M., 2011. Microscopic insights into pedestrian motion through a bottleneck, resolving spatial and temporal variations. arXiv 1105.1532 (v1), URL: http://arxiv.org/abs/1105.1532.

Liu, S., Lo, S., Ma, J., Wang, W., 2014. An agent-based microscopic pedestrian flow simulation model for pedestrian traffic problems. IEEE Trans. Intell. Transp. Syst. PP (99), 1–10.

Lovreglio, R., Ronchi, E., Nilsson, D., 2015. Calibrating floor field cellular automaton models for pedestrian dynamics by using likelihood function optimization. Physica A 438, 308–320. http://dx.doi.org/10.1016/j.physa.2015.06.040.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. J. Chem. Phys. 21, 1087–1092.

Robin, T., Antonini, G., Bierlaire, M., Cruz, J., 2009. Specification, estimation and validation of a pedestrian walking behavior model. Transp. Res. B 43 (1), 36–56. http://dx.doi.org/10.1016/j.trb.2008.06.010.

Rudloff, C., Matyus, T., Seer, S., 2014. Comparison of different calibration techniques on simulated data. In: Weidmann, U., Kirsch, U., Schreckenberg, M. (Eds.), Pedestrian and Evacuation Dynamics 2012. Springer International Publishing, Cham, pp. 657–672.

Ruggiero, L., Charitha, D., Xiang, S., Lucia, B., 2018. Investigating pedestrian navigation in indoor open space environments using big data. Appl. Math. Model. 62, 499–509. http://dx.doi.org/10.1016/j.apm.2018.06.014.

Rupprecht, T., Klingsch, W., Seyfried, A., 2011. Influence of geometry parameters on pedestrian flow through bottleneck. In: Peacock, R.D., Kuligowski, E.D., Averill, J.D. (Eds.), Pedestrian and Evacuation Dynamics. Springer US, pp. 71–80. http://dx.doi.org/10.1007/978-1-4419-9725-8_7.

Schadschneider, A., 2001. Cellular automaton approach to pedestrian dynamics - theory. In: Schreckenberg, M., Sharma, S.D. (Eds.), Pedestrian and Evacuation Dynamics. Springer, pp. 75–86.

Seer, S., Brändle, N., Ratti, C., 2014a. Kinects and human kinetics: A new approach for studying pedestrian behavior. Transp. Res. C 48, 212–228. http://dx.doi.org/10.1016/j.trc.2014.08.012.

Seer, S., Rudloff, C., Matyus, T., Brändle, N., 2014b. Validating social force based models with comprehensive real world motion data. Transp. Res. Procedia 2, 724–732. http://dx.doi.org/10.1016/j.trpro.2014.09.080, The Conference on Pedestrian and Evacuation Dynamics 2014 (PED 2014), 22-24 October 2014, Delft, The Netherlands.

Seitz, M.J., Köster, G., 2012. Natural discretization of pedestrian movement in continuous space. Phys. Rev. E 86 (4), 046108. http://dx.doi.org/10.1103/PhysRevE.86.046108.

Seyfried, A., Passon, O., Steffen, B., Boltes, M., Rupprecht, T., Klingsch, W., 2009. New insights into pedestrian flow through bottlenecks. Transp. Sci. 43, 395–406. http://dx.doi.org/10.1287/trsc.1090.0263.

Smith, R.C., 2014. Uncertainty Quantification: Theory, Implementation, and Applications. In: Computational Science and Engineering, Society for Industrial and Applied Mathematics, URL: http://www.cambridge.org/de/academic/subjects/mathematics/mathematical-modelling-and-methods/uncertainty-quantification-theory-implementation-and-applications.

Song, X.B., Lovreglio, R., 2021. Investigating personalized exit choice behavior in fire accidents using the hierarchical Bayes estimator of the random coefficient logit model. Anal. Methods Accid. Res. 29, 100140. http://dx.doi.org/10.1016/j.amar.2020.100140.

Steiner, A., Philipp, M., Schmid, A., 2007. Parameter estimation for a pedestrian simulation model. In: Swiss Transport Research Conference. URL: https://digitalcollection.zhaw.ch/handle/11475/14141.

Taherifar, N., Hamedmoghadam, H., Sree, S., Saberi, M., 2019. A macroscopic approach for calibration and validation of a modified social force model for bidirectional pedestrian streams. Transp. A: Transp. Sci. 15 (2), 1637–1661. http://dx.doi.org/10.1080/23249935.2019.1636156.

Tang, M., Jia, H., 2011. An approach for calibration and validation of the social force pedestrian model. In: Proceedings 2011 International Conference on Transportation, Mechanical, and Electrical Engineering. TMEE, pp. 2026–2031. http://dx.doi.org/10.1109/TMEE.2011.6199614.

Tavaré, S., Balding, D.J., Griffiths, R.C., Donnelly, P., 1997. Inferring coalescence times from DNA sequence data. Genetics 145 (2), 505–518, URL: https://www.genetics.org/content/145/2/505.

Toni, T., Welch, D., Strelkowa, N., Ipsen, A., Stumpf, M.P., 2009. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J. R. Soc. Interface 6 (31), 187–202. http://dx.doi.org/10.1098/rsif.2008.0172.

Voloshin, D., Rybokonenko, D., Karbovskii, V., 2015. Optimization-based calibration for micro-level agent-based simulation of pedestrian behavior in public spaces. Procedia Comput. Sci. 66, 372–381. http://dx.doi.org/10.1016/j.procs.2015.11.043, 4th International Young Scientist Conference on Computational Science.

von Sivers, I., 2016. Modellierung sozialpsychologischer Faktoren in Personenstromsimulationen - Interpersonale Distanz und soziale Identitäten. (Ph.D. thesis). Technische Universität München, URL: https://mediatum.ub.tum.de/doc/1303742/1303742.pdf.

von Sivers, I., Köster, G., 2015. Dynamic stride length adaptation according to utility and personal space. Transp. Res. B 74, 104–117. http://dx.doi.org/10.1016/j.trb.2015.01.009.

Weidmann, U., 1993. Transporttechnik Der FussgäNger, second ed. In: Schriftenreihe des IVT, vol. 90, Institut für Verkehrsplanung, Transporttechnik, Strassen- und Eisenbahnbau (IVT) ETH, Zürich, http://dx.doi.org/10.3929/ethz-b-000242008.

Wolinski, D., J. Guy, S., Olivier, A.-H., Lin, M., Manocha, D., Pettré, J., 2014. Parameter estimation and comparative evaluation of crowd simulations. Comput. Graph. Forum 33 (2), 303–312. http://dx.doi.org/10.1111/cgf.12328.

Zeng, W., Chen, P., Yu, G., Wang, Y., 2017. Specification and calibration of a microscopic model for pedestrian dynamic simulation at signalized intersections: A hybrid approach. Transp. Res. C 80, 37–70. http://dx.doi.org/10.1016/j.trc.2017.04.009.

Zeng, W., Nakamura, H., Chen, P., 2014. A modified social force model for pedestrian behavior simulation at signalized crosswalks. Procedia - Soc. Behav. Sci. 138, 521–530. http://dx.doi.org/10.1016/j.sbspro.2014.07.233, The 9th International Conference on Traffic and Transportation Studies (ICTTS 2014).