Transportation Research Part C 195 (2027) 106173

Contents lists available at ScienceDirect

Transportation Research Part C

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

An experiment-informed framework for crowd evacuation simulation via adaptive calibration

Wenhan Wu [a], Jingwei Ge [b,*], Li Li [c]

[a] Senseable City Laboratory, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA
[b] Department of Data and Systems Engineering, University of Hong Kong, Hong Kong, China
[c] Department of Automation, Tsinghua University, Beijing, 100084, China

ARTICLE INFO

Keywords:
Parameter calibration
Crowd evacuation
Adaptive sampling
Social force model
Finite-budget optimization

ABSTRACT

Computational modeling has become indispensable for studying crowd evacuation, enabling systematic analysis of pedestrian dynamics across scenarios that are difficult to observe, reproduce, or ethically test in real emergencies. Among microscopic simulation approaches, the social force model remains one of the most widely used, yet its empirical calibration is still fragmented across scenario-specific applications, selected parameter subsets, and narrow target observables. Here, we develop an experiment-informed framework for crowd evacuation simulation via adaptive calibration, in which scenario reconstruction, forward simulation, multi-scale discrepancy evaluation, and adaptive sampling are integrated into a closed-loop workflow. Using competitive room-evacuation experiments as a case study, we show that the adaptive sampler progressively concentrates evaluations in low-loss regions of the admissible parameter space and identifies trial-specific parameter vectors that improve agreement with the experimental observations used for calibration. The calibrated simulations reproduce the main experimental evacuation statistics across both the multi-scale calibration objective and competitiveness-related quantities that are not directly optimized. The calibrated parameter ensemble further reveals that prescribed behavioral conditions are encoded through coordinated variation across self-driven adjustment, psychological repulsion, and physical-contact mechanisms, rather than through a single dominant parameter. These results support the framework as a systematic approach for incorporating experimental evidence into social force simulations, enabling offline calibration, assessment of empirical consistency, and model-dependent parameter interpretation.

1. Introduction

Crowd evacuation is a critical process in public safety, shaping the consequences of emergencies in transport hubs, stadiums, schools, commercial buildings, and other densely occupied facilities (Helbing and Johansson, 2009). Understanding how pedestrians move, interact, and compete for exits under emergency conditions is essential for evaluating evacuation capacity and improving emergency-response strategies (Dong et al., 2020). However, the empirical basis for such understanding is intrinsically limited (Boltes et al., 2019), since high-quality recordings of real evacuations are scarce: emergencies are rare and unforeseeable, and pedestrian trajectories must often be reconstructed post hoc from partial surveillance footage under ethical, legal, and technical constraints (Wu

* Corresponding author.
E-mail addresses: wenhanwu@mit.edu (W. Wu), jingweig@hku.hk (J. Ge), li-li@tsinghua.edu.cn (L. Li).

https://doi.org/10.1016/j.trc.2026.106173
Received 17 May 2026; Received in revised form 5 October 2026; Accepted 16 November 2026
0968-090X/© 2026 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

and Zheng, 2024). As a result, controlled evacuation experiments and simulation-based pipelines have become indispensable tools for studying emergency evacuation behavior.

Controlled evacuation experiments provide the main empirical surrogate for studying pedestrian behavior under emergency conditions. They typically recruit volunteers to move through controlled rooms, corridors, stairways, bottlenecks, or reduced-scale arenas, while overhead videos or sensors record trajectories, exit times, densities, speeds, and local interaction patterns (Haghani, 2020). Such experiments have been widely used to investigate fundamental evacuation mechanisms, including bottleneck discharge, clogging and intermittent outflow, stair and corridor movement, obstacle effects near exits, and social group behavior (Helbing et al., 2005; Pastor et al., 2015; Huo et al., 2016; Shiwakoti et al., 2019; Xie et al., 2020). Recent studies have further extended this empirical line to new experimental settings, such as distancing-constrained bottleneck flow, mixed-reality exit-choice experiments, vertical evacuation scenarios, and more complex bottleneck structures (Echeverría-Huarte et al., 2022; Tao et al., 2024; Zhang et al., 2024; Sun et al., 2025). Nevertheless, controlled experiments cannot fully reproduce the psychological stress, uncertainty, injury risk, and strong contact dynamics of real emergencies within ethical constraints (Gwynne et al., 2017). They are also costly to scale, because each experiment requires recruited participants, controlled facilities, safety supervision, and measurement infrastructure (Haghani and Sarvi, 2018; Haghani, 2020). The feasible range of scenarios is therefore limited by participant availability, site geometry, experimental realism, and the practical difficulty of systematically varying many boundary and behavioral conditions. These limitations mean that experiments alone cannot cover the broad space of evacuation situations required for safety assessment, motivating simulation models as a complementary tool for controlled, repeatable, and low-cost exploration.

Numerical simulations address this need by allowing scenario layouts, initial configurations, and behavioral assumptions to be varied systematically in a computational environment. Mainstream evacuation models can be broadly classified into three classes: force-based models, cellular automata models, and agent-based models (Zheng et al., 2009). Among them, the social force model (Helbing and Molnár, 1995; Helbing et al., 2000) has become one of the most widely used formulations for studying pedestrian and evacuation dynamics, offering a compact and interpretable description of microscopic pedestrian motion. However, the practical value of this model depends on whether its parameters are empirically grounded. These parameters encode key behavioral and mechanical components of evacuation, including desired motion, velocity adaptation, interpersonal avoidance, body compression, and sliding friction (Chen et al., 2017). If they are inherited from previous studies or adjusted only partially, simulations may match one aggregate outcome while using parameter combinations that are not meaningful for the target scenario. This is especially problematic in dense evacuations, where self-driven, repulsive, and contact forces interact nonlinearly, meaning that one parameter change can affect multiple observables (Kretz et al., 2018). Parameter calibration therefore functions as a necessary link between experiments and simulation-based inference: a calibrated model should reproduce individual escape times and complementary signatures across multiple observational scales. Without such grounding, simulations may reproduce observed dynamics but provide a limited basis for interpreting the underlying mechanisms or supporting broader model analysis.

To realize this, we propose an experiment-informed framework bridging empirical evacuation experiments with social force simulations via adaptive full-vector calibration under a finite computational budget. The framework integrates four coupled modules: scenario reconstruction, forward simulation, discrepancy evaluation, and adaptive sampling. For each experimental run, it reconstructs the simulation context, evaluates candidate parameter vectors through forward simulation, compares simulated and empirical outcomes across multiple observational scales, and adaptively reallocates the sampling budget toward low-loss regions of the admissible parameter space. Through this closed-loop process, the framework produces a re-runnable, empirically constrained computational counterpart of each recorded experimental run, defined by its reconstructed scenario and entry conditions, calibrated parameter vector, and fixed observable-extraction procedures, thereby establishing an operational link between the recorded experiment and the corresponding social force simulation. In this sense, it primarily provides an empirically constrained basis for simulation and model assessment, upon which broader computational analyses can subsequently be developed. We examine the framework using the competitive room-evacuation experiments of Garcimartín et al. (2017) as a case study. The dataset provides individual-level trajectories from repeated evacuation drills under controlled conditions, allowing us to evaluate whether calibrated parameters reproduce empirical evacuation statistics and preserve systematic differences among experimental settings. The results indicate that the calibrated simulations recover the main spatio-temporal movement patterns observed in the experiments and that the calibrated parameter vectors provide an interpretable representation of how behavioral conditions are encoded within the social force model. More broadly, since controlled evacuation experiments are resource-intensive yet yield limited realizations, the proposed framework can extend the computational utility of empirical observations by providing an empirically grounded basis for repeated simulation and subsequent model analysis.

The rest of this paper is organized as follows. Section 2 reviews scenario-oriented and method-oriented studies on social force parameter calibration. Section 3 introduces the social force model and formulates the parameter calibration problem. Section 4 describes the experiment-informed simulation framework and its adaptive sampling procedure. Section 5 applies the framework to a competitive evacuation experiment to assess search performance, simulation reproducibility, and parameter interpretation. Section 6 concludes and discusses the implications and limitations of the framework.

2. Related work

In early applications of the social force model (Helbing and Molnár, 1995; Helbing et al., 2000), parameters were assigned from socio-psychological literature, empirical investigations, and engineering handbooks, enabling the model to reproduce collective phenomena such as lane formation, bottleneck oscillation, and evacuation dynamics. However, because most parameters are not directly observable and may vary across different scenarios and behavioral conditions, later studies began to calibrate parameters using

empirical data. This section reviews two strands of this literature: scenario-oriented calibration studies involving specific pedestrian facilities, evacuation contexts, and model extensions, and method-oriented calibration studies examining how model parameters can be estimated, interpreted, and validated from trajectory data.

2.1. Scenario-oriented calibration for social force models

A large body of work has developed calibration procedures for specific application scenarios. Johansson et al. (2007) provided an early example by using video-tracking data from escalator scenes and an evolutionary optimization algorithm to identify suitable interaction parameters. Li et al. (2015) applied differential evolution to estimate model parameters for classroom evacuation during the 2013 Ya'an earthquake, while Sticco et al. (2021) focused mainly on contact-related coefficients to reproduce high-density crowd gatherings at store entrances. Bassoli and Vincenzi (2021) calibrated the social force model for pedestrian flow on footbridges using response surface methodology, and Li et al. (2024) combined YOLOv5-based pedestrian detection with an improved Newton downhill method for metro-station pedestrian flow simulation. Other studies have extended social force calibration to more specialized pedestrian contexts or model variants. Dias et al. (2018) calibrated model parameters for interactions between pedestrians and personal mobility vehicles. Zeng et al. (2014) developed and estimated a modified social force model for pedestrian behavior at signalized crosswalks. Xie et al. (2021) identified parameters in a group-associated formulation using single-exit room evacuation data. Qi et al. (2022) optimized model parameters for preschool children evacuation using multi-scenario experimental data.

2.2. Method-oriented calibration for social force models

A second strand of research has focused on methods for estimating and interpreting social force parameters from trajectory data. Tang and Jia (2011) utilized least-squares regression to fit repulsion-related parameters and reproduce pedestrian-flow characteristics. Ko et al. (2012) proposed a maximum-likelihood framework based on observed trajectories, formulating calibration as a statistical estimation problem. Seer et al. (2014) adopted nonlinear regression to estimate parameter values and their distributions from real-world bidirectional pedestrian traffic data. Corbetta et al. (2015) developed Bayesian probabilistic calibration to estimate parameter values and uncertainty distributions, and further linked parameter estimation with force-model comparison. Other studies have examined the sensitivity, transferability, and interpretability of calibrated parameters. Rudloff et al. (2013) showed that calibration results can depend on trajectory quality, objective-function choice, and estimation method. Ma et al. (2019) indicated that desired-speed assignment strategies can substantially affect model performance under different density conditions. Kretz et al. (2018) emphasized a more fundamental difficulty: many parameters lack direct measurable counterparts, and observed behaviors often result from the coupled effects of multiple interacting parameters.

2.3. Summary of research gaps

In summary, existing studies have established that empirical calibration is essential for applying social force models to real pedestrian dynamics. Across different application contexts and estimation methods, they show that calibrated parameters are scenario-dependent quantities shaped by data quality, objective design, uncertainty, and parameter coupling. However, several limitations remain. First, many studies calibrate only selected parameters, such as desired speed, repulsion terms, contact coefficients, or extension-specific parameters, while other components remain fixed according to literature values or prior assumptions. Second, calibration is often tied to a single scenario or a narrow set of observables, which may improve local fidelity but does not necessarily ensure that simulated trajectories reproduce evacuation dynamics across multiple observational scales. Third, existing methods rarely treat the social force model as a full high-dimensional parameterized simulator whose admissible parameter space must be searched under a finite computational budget. Therefore, systematic full-vector calibration of the social force model remains insufficiently explored, especially when calibrated simulations are expected to support both empirical reproduction and mechanism-level interpretation.

3. Preliminaries

This section introduces the basic modeling and calibration concepts used throughout the study. Pedestrian motion is described using the social force model, in which each simulation is governed by a six-parameter calibration vector $\boldsymbol{\theta}$. The calibration task is then defined as the search for a parameter vector within an admissible domain that enables the simulated trajectories to reproduce the corresponding experimental runs as closely as possible. We first present the mathematical formulation of the social force model and define the parameters considered in calibration. We then formulate parameter calibration as a finite-budget optimization problem under three sources of uncertainty.

3.1. Mathematical form of the social force model

The social force model of Helbing and Molnár (1995), Helbing et al. (2000) is a continuum physical-force formulation in which each pedestrian is a self-propelled disk governed by a continuous equation of motion. Let $\mathbf{x}_i(t)$ and $\mathbf{v}_i(t)$ denote the position and velocity of pedestrian i, we have

$$\frac{d\mathbf{x}_i(t)}{dt} = \mathbf{v}_i(t) \tag{1}$$

where the change of $\mathbf{v}_i(t)$ is described by the following acceleration equation:

$$m_i \frac{d\mathbf{v}_i(t)}{dt} = \mathbf{f}_{id} + \sum_{j(\neq i)} \mathbf{f}_{ij} + \sum_{W} \mathbf{f}_{iW} \tag{2}$$

The first term $\mathbf{f}_{id}$ is a self-driven force that drives pedestrian i toward the destination:

$$\mathbf{f}_{id} = m_i \frac{v_i^0(t)\mathbf{e}_i^0(t) - \mathbf{v}_i(t)}{\tau_i} \tag{3}$$

where pedestrian i of mass m_i prefers to move with a desired speed $v_i^0(t)$ in a desired direction $\mathbf{e}_i^0(t)$, and therefore tends to adapt the actual velocity $\mathbf{v}_i(t)$ within a characteristic time τ_i.

The second term $\mathbf{f}_{ij}$ is an interaction force with other pedestrians j:

$$\mathbf{f}_{ij} = A_i \exp\left[\left(r_{ij} - d_{ij}\right)/B_i\right]\mathbf{n}_{ij} + kg\left(r_{ij} - d_{ij}\right)\mathbf{n}_{ij} + \kappa g\left(r_{ij} - d_{ij}\right)\Delta v_{ji}^t \mathbf{t}_{ij} \tag{4}$$

Here, $A_i \exp\left[\left(r_{ij} - d_{ij}\right)/B_i\right]\mathbf{n}_{ij}$ is a repulsive interaction force that describes the psychological tendency of pedestrians i and j to stay away from each other, where A_i and B_i are constants, r_{ij} is the sum of their radii r_i and r_j, $d_{ij} = \left\|\mathbf{x}_i - \mathbf{x}_j\right\|$ is the distance between their centers of mass, and $\mathbf{n}_{ij}$ is the normalized vector pointing from pedestrian j to i. The body force $kg\left(r_{ij} - d_{ij}\right)\mathbf{n}_{ij}$ counteracts body compression and the sliding friction force $\kappa g\left(r_{ij} - d_{ij}\right)\Delta v_{ji}^t \mathbf{t}_{ij}$ impedes relative tangential motion, where k and κ are large constants, $\mathbf{t}_{ij}$ is the tangential direction, and Δv_{ji}^t is the tangential velocity difference. Note that $g(x)$ is zero if pedestrians i and j do not touch each other; otherwise, it equals the argument x.

The third term $\mathbf{f}_{iW}$ is an interaction force with wall W, taking a similar functional form:

$$\mathbf{f}_{iW} = A_i \exp\left[\left(r_i - d_{iW}\right)/B_i\right]\mathbf{n}_{iW} + kg\left(r_i - d_{iW}\right)\mathbf{n}_{iW} - \kappa g\left(r_i - d_{iW}\right)\left(\mathbf{v}_i \cdot \mathbf{t}_{iW}\right)\mathbf{t}_{iW} \tag{5}$$

where $A_i \exp\left[\left(r_i - d_{iW}\right)/B_i\right]\mathbf{n}_{iW}$, $kg\left(r_i - d_{iW}\right)\mathbf{n}_{iW}$ and $\kappa g\left(r_i - d_{iW}\right)\left(\mathbf{v}_i \cdot \mathbf{t}_{iW}\right)\mathbf{t}_{iW}$ represent the repulsive interaction force, body force, and sliding friction force between pedestrian i and walls, respectively. Here, d_{iW} is the Euclidean distance to wall W, $\mathbf{n}_{iW}$ is the normalized vector perpendicular to it, and $\mathbf{t}_{iW}$ is the direction tangential to it.

3.2. Parameter calibration as a finite-budget problem

In most studies, the social force model requires eight scalar parameters to be specified. The two parameters describing pedestrian anthropometric properties are fixed as $m_i = 70\,\mathrm{kg}$ and $r_i = 0.20\,\mathrm{m}$ to represent typical adult pedestrians and are not treated as calibration parameters. The remaining six parameters govern the pedestrian's desired motion, interpersonal repulsion, and contact interactions, and therefore need to be inferred from empirical data. To simplify calibration, these individual-level parameters are assumed homogeneous within each experimental run. This keeps the calibration problem tractable in a six-dimensional parameter space and allows the resulting parameter vector to represent the effective dynamics of each run. Accordingly, we define the parameter vector as:

$$\boldsymbol{\theta} = (v^0, \tau, A, B, k, \kappa) \in \Theta \subset \mathbb{R}^6 \tag{6}$$

where Θ represents the admissible domain of these model parameters. The calibrated vector contains only pedestrian-motion and interaction parameters, whereas facility characteristics and experimental conditions, such as exit width, room geometry, crowd size, and initial density, enter through the reconstructed scenario and entry conditions. When the framework is applied to a different experimental condition, the corresponding scenario is reconstructed and the parameter vector is recalibrated against its own empirical observations.

The calibration seeks a parameter vector $\boldsymbol{\theta}_g^\star$ that enables the numerical simulation to reproduce the empirical trajectory dataset $\mathcal{D}_g$ of experimental group g as closely as possible. Let $\mathcal{S}(\boldsymbol{\theta}; \mathcal{D}_g)$ denote the simulated trajectories obtained by integrating Eq. (2) with parameter vector $\boldsymbol{\theta}$ under the initial and boundary conditions associated with $\mathcal{D}_g$, and let $\mathcal{M}\left(\mathcal{S}(\boldsymbol{\theta}; \mathcal{D}_g), \mathcal{D}_g\right)$ denote a scalar discrepancy between $\mathcal{S}(\boldsymbol{\theta}; \mathcal{D}_g)$ and $\mathcal{D}_g$. The calibration target is given by:

$$\boldsymbol{\theta}_g^\star = \arg\min_{\boldsymbol{\theta}\in\Theta} \mathcal{M}\left(\mathcal{S}(\boldsymbol{\theta}; \mathcal{D}_g), \mathcal{D}_g\right) \tag{7}$$

Here, $\theta_g^\star$ should be interpreted as a calibration optimum: its location is affected by two sources of inferential uncertainty. 1) Realization uncertainty arises because the empirical dataset $\mathcal{D}_g$ represents a single realization of the experimental process; therefore, the expected loss over repeated realizations cannot be reliably estimated, and the observed realization $\mathcal{D}_g$ is used directly in Eq. (7). 2) Metric uncertainty arises from the construction of the discrepancy function $\mathcal{M}$, including the choice and relative weighting of its constituent observables, and alternative metric specifications may lead to different minimizers.

Distinct from these inferential uncertainties, the finite computational budget affects only the numerical approximation of $\boldsymbol{\theta}_g^\star$: even with the experimental realization and discrepancy function fixed, the continuous parameter domain $\Theta \subset \mathbb{R}^6$ cannot be exhaustively searched. Accordingly, the finite-budget estimate replaces the minimization over Θ by the best loss observed across a finite sequence of sampled parameter vectors. Let N_s denote the sampling budget and $\{\boldsymbol{\theta}^{(n)}\}_{n=1}^{N_s}$ the sequence generated by the sampling procedure. The best sampled index is then defined as:

$$\widehat{n}_g = \arg\min_{n\in\{1,\ldots,N_s\}} \mathcal{M}\left(\mathcal{S}(\boldsymbol{\theta}^{(n)}; \mathcal{D}_g), \mathcal{D}_g\right) \tag{8}$$

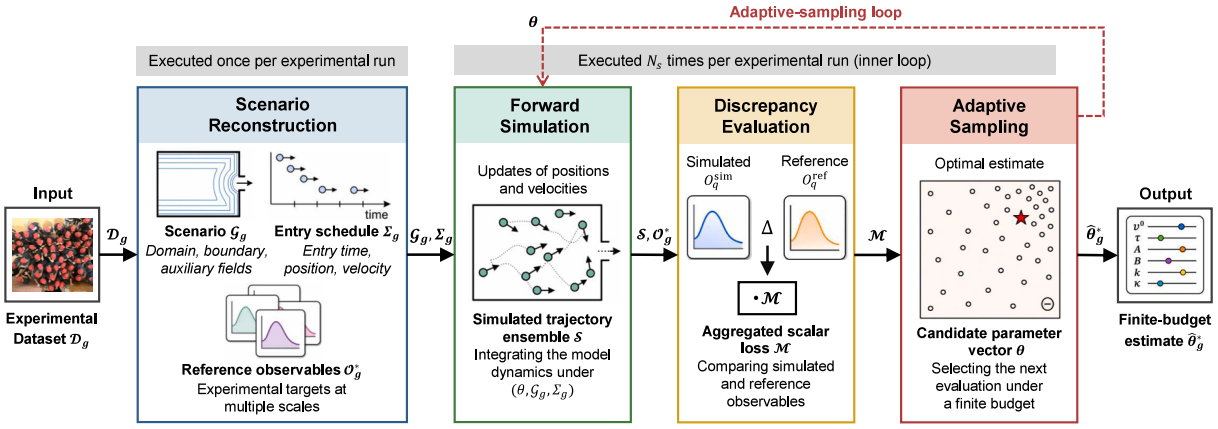

Fig. 1. Schematic diagram of the experiment-informed framework for crowd evacuation simulation. The framework consists of four coupled modules: scenario reconstruction, forward simulation, discrepancy evaluation, and adaptive sampling. Scenario reconstruction is performed once for each experimental dataset, while the remaining modules form a closed adaptive-sampling loop under a finite evaluation budget.

The finite-budget estimate of the calibration optimum is written as:

$$\hat{\theta}_g^\star = \theta^{(\hat{n}_g)} \tag{9}$$

The quality of the approximation $\hat{\theta}_g^\star \approx \theta_g^\star$ depends on the coverage and adaptivity of the sampling procedure. In Section 4, we construct such a procedure and embed it in a framework that converts an experimental realization into a calibrated simulator.

4. Experiment-informed simulation framework

The realization of the finite-budget target $\hat{\theta}_g^\star$ requires four operations with distinct input–output types, organized as four modules of the proposed framework in Fig. 1. (1) Scenario reconstruction converts the experimental dataset $\mathcal{D}_g$ into a reusable simulation context. (2) Forward simulation evolves the system under the simulator's dynamics from that context for any candidate parameter vector. (3) Discrepancy evaluation reduces the simulated ensemble to the scalar loss $\mathcal{M}$ that ranks candidates. (4) Adaptive sampling selects the next candidate under the finite budget. The four modules are coupled in a closed loop. Scenario reconstruction is evaluated once and cached, whereas the forward simulator, discrepancy evaluator, and adaptive sampler form an inner adaptive-sampling loop. In each unit evaluation, the adaptive sampler supplies a candidate parameter vector θ, the forward simulator generates the corresponding trajectory ensemble, the discrepancy evaluator returns the scalar loss $\mathcal{M}$, and this loss is used to guide the next candidate selection.

4.1. Scenario reconstruction

Scenario reconstruction is a one-shot mapping from an experimental dataset $\mathcal{D}_g$ to three cached artifacts that the remaining modules consume:

$$\mathcal{R} : \mathcal{D}_g \longmapsto (\mathcal{G}_g, \Sigma_g, \mathcal{O}_g^\star) \tag{10}$$

where $\mathcal{G}_g$ is the scenario specification, Σ_g represents the entry schedule, and $\mathcal{O}_g^\star$ indicates the set of reference observables. The mapping is evaluated once per experimental run, and subsequent sample evaluations read from the cache without re-processing $\mathcal{D}_g$.

The first artifact $\mathcal{G}_g$ encapsulates the static structures of the experimental setup that the simulator's dynamics consume. It bundles three components into a triple as follows:

$$\mathcal{G}_g = (\Omega_g, \partial\Omega_g, \phi_g) \tag{11}$$

where Ω_g and $\partial\Omega_g$ are the simulation domain and its boundary, and ϕ_g is a possibly empty collection of auxiliary fields such as distance fields or potential maps.

The second artifact Σ_g encodes the entry schedule of individuals into the simulation, as given by a set of triples:

$$\Sigma_g = \left\{(t_i^s, \mathbf{x}_i^s, \mathbf{v}_i^s)\right\}_{i=1}^{N_g} \tag{12}$$

Here, each triple gives the entry time t_i^s, position $\mathbf{x}_i^s$, and velocity $\mathbf{v}_i^s$ of pedestrian i, and N_g is the total number of pedestrians. The values of t_i^s and $\mathbf{x}_i^s$ are taken from the first recorded trajectory point, while $\mathbf{v}_i^s$ is estimated by forward differencing the first two points. The resulting Σ_g defines the deterministic input stream for the forward simulator.

In addition to $\mathcal{G}_g$ and Σ_g that feed the forward simulator, the third artifact $\mathcal{O}_g^\star$ is the target that the discrepancy module compares against simulated output:

$$\mathcal{O}_g^\star = \left\{O_q(\mathcal{D}_g)\right\}_{q=1}^{Q} \tag{13}$$

where Q is the number of reference observables, and O_q denotes the qth observable map applied once to D_g. The same maps are applied identically to simulated trajectories in the forward-simulation module, thus the discrepancy compares quantities computed by procedures that are by construction identical between simulation and experiment.

4.2. Forward simulation

Forward simulation maps a parameter vector θ with the scenario artifacts $(\mathcal{G}_g, \Sigma_g)$ to a simulated trajectory ensemble $S(\theta; D_g)$. The output shares its file layout with D_g so that the reconstruction module's extraction routine applies without branching to both simulated and experimental trajectories.

The module integrates the simulator's equations of motion by using a semi-implicit Euler discretization. At each time, the velocity is updated by:

$$\mathbf{v}_i(t + \Delta t) = \mathbf{v}_i(t) + \frac{\mathbf{F}_i(t)}{m_i} \Delta t \tag{14}$$

where Δt is the time step, and $\mathbf{F}_i(t)$ is the combined force on pedestrian i given by the simulator's dynamics. The position update is then described as:

$$\mathbf{x}_i(t + \Delta t) = \mathbf{x}_i(t) + \mathbf{v}_i(t + \Delta t)\, \Delta t \tag{15}$$

Note that physically motivated state bounds are enforced by rescaling whenever the explicit update exceeds them.

Each integration step proceeds in several stages. The module first activates each pedestrian whose entry time has been reached, evaluates the dynamical forces on each active individual, updates their velocities and positions via Eqs. (14) and (15), and removes any agent that has escaped from the exit. A consequence of the deterministic integration is that a fixed triple $(\theta, \mathcal{G}_g, \Sigma_g)$ produces the same trajectory ensemble on each re-invocation. This determinism lets the adaptive sampling module isolate its own variance from any variance introduced by the simulator.

4.3. Discrepancy evaluation

Discrepancy evaluation simplifies the simulated trajectory ensemble to the scalar loss that the sampling module uses to rank candidate parameters. The module is the map that returns a single non-negative number summarizing the disagreement between simulated and reference observables:

$$\mathcal{E} : \big(S(\theta; D_g),\, \mathcal{O}_g^\star\big) \longmapsto \mathcal{M}\big(S(\theta; D_g), D_g\big) \in \mathbb{R}_{\geq 0} \tag{16}$$

The map $\mathcal{E}$ proceeds in three stages. The first stage extracts the simulated counterparts of the reference observables by applying the same routines that produced $\mathcal{O}_g^\star$ to S:

$$O_q^{\mathrm{sim}}(\theta; D_g) := O_q\big(S(\theta; D_g)\big), \quad q = 1, \ldots, Q \tag{17}$$

This places simulated and experimental quantities in comparable forms. The second stage collapses each matched pair into a constituent metric:

$$M_q\big(S(\theta; D_g), D_g\big) := \delta_q\big(O_q^{\mathrm{sim}}(\theta; D_g),\, O_q(D_g)\big) \tag{18}$$

where δ_q captures the discrepancy at one observational scale or modality. The third stage aggregates the constituent metrics into the scalar loss:

$$\mathcal{M}\big(S(\theta; D_g), D_g\big) = \mathcal{A}\big(M_1, \ldots, M_Q\big) \tag{19}$$

where the aggregation operator $\mathcal{A}$ applies scale normalization if the constituent metrics carry heterogeneous units.

4.4. Adaptive sampling

As the discrepancy module returns a scalar loss $\mathcal{M}$ for each candidate, the remaining design problem is choosing where in Θ to evaluate it under a finite budget. For this, adaptive sampling constructs the finite-budget sequence $\{\theta^{(n)}\}_{n=1}^{N_s}$, from which the estimate $\widehat{\theta}_g^\star$ in Eq. (9) is selected. The module takes as inputs the admissible domain Θ, the total sampling budget N_s, and four configuration parameters: the number of adaptive rounds R, the number of parent sub-cubes retained per round J, the split width s giving the number of dimensions bisected per parent at each round, and an early-stopping tolerance ϵ on the best observed loss. It returns $\widehat{\theta}_g^\star$ together with the full record $\mathcal{H}$ of evaluated parameter–loss pairs $(\theta, \mathcal{M})$. Internally, to equalize sampling effort across the six parameters despite their heterogeneous physical units, every operation is performed on the normalized unit cube $[0, 1]^6$. Here, we define the affine normalization map $T : \Theta \to [0, 1]^6$ by $\mathbf{u} = T(\theta)$, and apply the inverse map T^{-1} before each call to the simulator.

The sampler is designed to satisfy two requirements. First, candidate parameter vectors must be distributed across Θ so that no region of volume comparable to the neighborhood of $\theta_g^\star$ is systematically missed. Pseudo-random samples in $[0, 1]^6$ have expected star discrepancy of order $N_s^{-1/2}$, which leaves unsampled volume comparable in size to the neighborhood that needs to be bracketed around $\theta_g^\star$ at sampling budgets of a few hundred. In contrast, low-discrepancy sequences are deterministic constructions designed

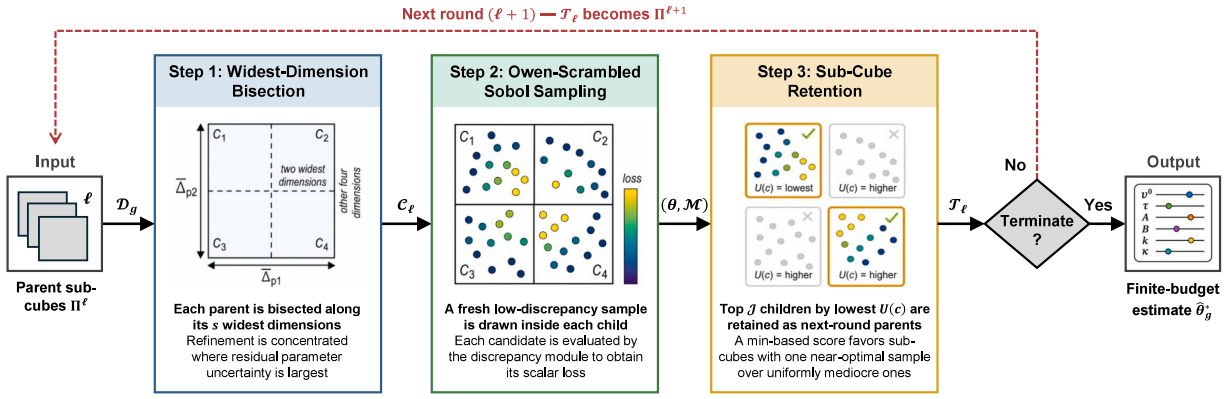

Fig. 2. Adaptive sampling procedure. In each round, parent sub-cubes Π^ℓ are bisected along the s widest dimensions to produce child sub-cubes C_ℓ. Owen-scrambled Sobol samples are evaluated within each child, and the top J children with the lowest retention score $U(c)$ are retained as next-round parents. The loop continues until the round limit is reached or the early-stopping tolerance is met.

to minimize star discrepancy, reducing this bound to $\mathcal{O}((\log N_s)^6/N_s)$ for the same sample count (Niederreiter, 1992). Second, the sampler must exploit the loss values observed on earlier samples to concentrate later samples in lower-loss regions of Θ. A single-stage low-discrepancy sequence carries no loss information, hence an adaptive wrapper around the underlying sequence is required.

We therefore adopt a three-step within-round procedure. In each round, every parent sub-cube is first partitioned into axis-aligned children (Step 1), a fresh Owen-scrambled Sobol sequence is then drawn inside each child and evaluated (Step 2), and the children with the lowest observed loss are retained as the parents for the next round (Step 3). The partition preserves the axis-aligned structure on which the low-discrepancy guarantee depends, while sub-cube-level aggregation provides the loss-based signal that drives retention.

Algorithm 1 Adaptive sampling procedure.

Require: admissible domain Θ, budget N_s, rounds R, retained parents J, split width s, tolerance ε
Ensure: finite-budget best estimate $\hat{\theta}_g^\star$, sampling record $\mathcal{H}$

map Θ to $[0,1]^6$; set $\Pi^1 \leftarrow \{[0,1]^6\}$, $\mathcal{H} \leftarrow \emptyset$, $\mathcal{M}^\star \leftarrow +\infty$
for $\ell = 1$ **to** R **do**
  compute $\bar{\Delta}_p$, the extent along dimension p averaged over the parents in Π^ℓ
  identify the s dimensions $\{p_1, \ldots, p_s\}$ with the largest $\bar{\Delta}_p$
  $C_\ell \leftarrow \emptyset$ ▷ children of round ℓ
  for each parent $\pi \in \Pi^\ell$ **do**
    bisect π along $\{p_1, \ldots, p_s\}$ into 2^s children, append them to C_ℓ
  end for
  $N \leftarrow \lfloor N_s/(R\,|C_\ell|) \rfloor$ ▷ $|C_\ell| = 2^s$ at $\ell = 1$, $2^s J$ thereafter
  for each child $c \in C_\ell$ **do**
    generate N Owen-scrambled Sobol points in $[0,1]^6$ and map onto c
    for each mapped point u **do**
      $\theta \leftarrow$ inverse normalisation of u
      obtain S from the forward simulator using $(\theta, \mathcal{G}_g, \Sigma_g)$
      obtain $\mathcal{M}$ from the discrepancy evaluator using $(S, \mathcal{O}_g^\star)$
      append $(\theta, \mathcal{M})$ to $\mathcal{H}$
      if $\mathcal{M} < \mathcal{M}^\star$ **then** $\mathcal{M}^\star \leftarrow \mathcal{M}$, $\hat{\theta}_g^\star \leftarrow \theta$
      end if
    end for
    $U(c) \leftarrow$ min of $\mathcal{M}$ values observed in c
  end for
  if $\mathcal{M}^\star < \varepsilon$ **then**
    break
  end if
  sort C_ℓ by U ascending; retain the top J as $\mathcal{T}_\ell$
  $\Pi^{\ell+1} \leftarrow \mathcal{T}_\ell$
end for
return $\hat{\theta}_g^\star$, $\mathcal{H}$

Step 1: Widest-dimension bisection. At each round $\ell \in \{1, \dots, R\}$, the admissible region Π^{ℓ} is the set of parent sub-cubes retained from round $\ell - 1$, with $\Pi^1 = \{[0,1]^6\}$ taken as a single parent at the first round. Let $\bar{\Delta}_p$ denote the extent along dimension p, averaged across the parents in Π^{ℓ}. The s dimensions with the largest $\bar{\Delta}_p$ are then bisected within every parent, producing 2^s children per parent. We write C_{ℓ} for the set of all children produced at round ℓ, so that $|C_{\ell}| = 2^s$ at round 1 and $|C_{\ell}| = 2^s J$ at every subsequent round. Specifically, the dimensions with the largest average extent carry the greatest residual parameter uncertainty, therefore refinement along them reduces the worst-case uncertainty most per split. The value of s itself controls a complementary trade-off: smaller s shrinks Π^{ℓ} slowly but keeps more samples per child, while larger s shrinks faster but spreads the per-round budget more thinly. At the extreme $s = 6$, every parent yields $2^6 = 64$ children per round, leaving each child with too few samples for its quasi-random structure to contribute useful coverage.

Step 2: Owen-scrambled Sobol sampling. Each child $c \in C_{\ell}$ receives $\lfloor N_s/(2^s R) \rfloor$ samples at round 1 and $\lfloor N_s/(2^s J R) \rfloor$ samples at every subsequent round. The samples come from a two-layer randomization of a low-discrepancy base sequence: The base layer is the deterministic Sobol sequence in dimension six, whose star discrepancy of its first N points, denoted $D_N^{\star}$, satisfies $D_N^{\star} = \mathcal{O}((\log N)^6/N)$ (Niederreiter, 1992). On top of this base, Owen scrambling applies nested uniform random permutations at every digit level of the base-b digit expansion of the Sobol points, where b is the construction base of the Sobol sequence (Owen, 1997). As a result, the scrambled sequence preserves the discrepancy bound of the underlying Sobol sequence and additionally makes each scrambled point a marginally uniform random variable on $[0,1]^6$. Each scrambled point $\mathbf{u}$ is finally affine-mapped into its child sub-cube c and passed to the forward-simulation and discrepancy-evaluation modules, which return the simulated trajectories and one scalar loss $\mathcal{M}\big(S(T^{-1}(\mathbf{u}); \mathcal{D}_g), \mathcal{D}_g\big)$.

Step 3: Sub-cube retention. Given the loss values produced in Step 2, each child $c \in C_{\ell}$ is scored by the lowest loss observed inside it, as given by:

$$U(c) = \min_{\mathbf{u} \in \text{samples}(c)} \mathcal{M}\big(S(T^{-1}(\mathbf{u}); \mathcal{D}_g), \mathcal{D}_g\big) \tag{20}$$

where $\text{samples}(c)$ is the set of mapped points falling in c. Since $\theta_g^{\star}$ is located at a single point in Θ, a sub-cube containing one low-loss sample is a better candidate for refinement than a sub-cube with uniformly moderate samples. This is because a mean-based score would invert this preference by favoring sub-cubes with many mediocre samples over those with a single near-optimal one. Given this scoring, all children in C_{ℓ} are pooled and sorted by U in ascending order, and the top J are retained as $\Pi^{\ell+1} = \mathcal{T}_{\ell}$, where $\mathcal{T}_{\ell}$ is the set of the J retained children. Across rounds, the outer loop terminates when R rounds have elapsed or when the best-so-far loss falls below the tolerance ε.

If no early stopping occurs, after R rounds the admissible region Π^{R+1} contains J sub-cubes of total volume $J/2^{sR}$ relative to the original cube. Beyond this volume contraction, the pair (R, J) governs a trade-off between refinement depth and per-child sample count: increasing R deepens the contraction but reduces the per-round budget $\lfloor N_s/R \rfloor$, while increasing J widens the explored region but reduces the per-child budget $\lfloor N_s/(2^s J R) \rfloor$. The two limiting values of J illustrate this trade-off: At $J = 1$, retention collapses to a greedy single-branch search, which amplifies the risk of premature convergence on a non-optimal sub-cube. At $J = 2^s$, every round-1 child is retained, forgoing the first-round filter, and selection still occurs at $\ell \geq 2$ since $2^s J$ candidates compete for J slots. The complete procedure of adaptive sampling is given as Algorithm 1, and Fig. 2 illustrates the three-step within-round workflow of widest-dimension bisection, Owen-scrambled Sobol sampling, and sub-cube retention.

The computational cost of Algorithm 1 consists mainly of sample generation and transformation, sub-cube partitioning, and sub-cube ranking. Generating and transforming the N_s d-dimensional Owen-scrambled Sobol samples requires $O(N_s d)$ operations. Across the R rounds, identifying the split dimensions, constructing the child sub-cubes, and ranking them by $U(c)$ require $O(Rd \log d)$, $O\Big(\sum_{\ell=1}^{R} |C_{\ell}| d\Big)$, and $O\Big(\sum_{\ell=1}^{R} |C_{\ell}| \log |C_{\ell}|\Big)$ operations, respectively. The overall algorithmic complexity is given by:

$$T_{\text{alg}} = O\left(N_s d + Rd \log d + \sum_{\ell=1}^{R} |C_{\ell}|\big(d + \log |C_{\ell}|\big)\right) \tag{21}$$

The storage complexity is $O(N_s d)$ for the sampling record $\mathcal{H}$. Under the present settings, $|C_1| = 2^s = 4$ and $|C_2| = 2^s J = 8$. Since d, R, J, and s are fixed and small, the overall algorithmic complexity is dominated by $O(N_s d)$ and therefore increases approximately linearly with the sampling budget.

5. Numerical experiments

5.1. Implementation for a competitive-evacuation case study

In this work, we validate the proposed calibration framework using the competitive room-evacuation experiments of Garcimartín et al. (2017) as a case study. The original experiment recorded evacuation drills in which volunteers left a room through a narrow door under three levels of prescribed competitiveness. These levels were imposed by verbal instructions: in the low-competitiveness condition, participants were instructed to evacuate without intentional physical contact; in the medium condition, soft physical contact was allowed; and in the high condition, moderate pushing was allowed. The experimental setup involved about one hundred volunteers, a rectangular area of approximately 6×4 m, and a door of width 0.69 m. Pedestrian positions were extracted from overhead video recordings by tracking red hats worn by the participants. The experimental dataset contained 10 low-, 9 medium-, and 13

high-competitiveness trials with available video and trajectory data. To ensure equal sample sizes across competitiveness levels and facilitate consistent comparison, 8 trials were randomly sampled without replacement from each class, yielding a balanced subset of 24 trials. The fixed single-exit geometry is selected as a case study, allowing the prescribed competitiveness condition to be isolated from environmental variation while generating typical evacuation phenomena such as congestion and intermittent discharge near the exit. By keeping the physical setting unchanged across trials, differences among the calibrated parameter vectors can be interpreted more cleanly as competitiveness-associated patterns. Next, we describe how the four framework modules are instantiated for this dataset.

For each selected experiment, the scenario-reconstruction module converts the recorded evacuation data into three cached inputs for calibration. The scenario specification $\mathcal{G}_g = (\Omega_g, \partial\Omega_g, \phi_g)$ is reconstructed from the experimental geometry as a rectangular floor domain with a single narrow exit. The accessible area defines Ω_g, the walls and exit segment define $\partial\Omega_g$, and the auxiliary field ϕ_g is a distance-to-exit field computed once over the domain. The desired direction in the social force simulator is given by the normalized negative gradient, $\mathbf{e}_i^0(t) = -\nabla\phi_g(\mathbf{x}_i(t))/|\nabla\phi_g(\mathbf{x}_i(t))|$. The entry schedule $\Sigma_g = \{(t_i^s, \mathbf{x}_i^s, \mathbf{v}_i^s)\}_{i=1}^{N_g}$ can be read from the tracked trajectories, assigning each simulated pedestrian the initial time, position, and velocity observed in the corresponding trial. The reference observables $\mathcal{O}_g^\star$ are also computed once from the experimental trajectories and consist of the ranked exit-time sequence, the time-lapse sequence between consecutive exits, and the density–speed point cloud. These cached observables provide the experimental references used to compute the constituent metrics by comparison with the same quantities from simulated trajectories.

Given the reconstructed scenario artifacts $(\mathcal{G}_g, \Sigma_g)$, the forward-simulation module maps each candidate parameter vector $\boldsymbol{\theta} = (v^0, \tau, A, B, k, \kappa)$ to a simulated trajectory ensemble $S(\boldsymbol{\theta}; \mathcal{D}_g)$. The simulator initializes agents according to Σ_g and drives them through the reconstructed domain $\mathcal{G}_g$ using the distance-field direction defined in the reconstruction step. The simulation time step $\Delta t = 0.02\,\mathrm{s}$ is set equal to the temporal interval of the experimental data, so that simulated and recorded trajectory points are aligned frame by frame. During simulation, all pedestrians' velocities and positions are updated using the combined self-driven, interpersonal, and wall-interaction forces parameterized by $\boldsymbol{\theta}$, and individuals are removed once they cross the exit segment. The simulated trajectories are written in the same structure as the experimental trajectories, allowing identical routines to extract passage times, time lapses, and density–speed observations from both data sources.

Discrepancy evaluation aggregates each simulated trajectory ensemble into the scalar loss $\mathcal{M}\big(S(\boldsymbol{\theta}; \mathcal{D}_g), \mathcal{D}_g\big)$. The three observables used here are a case-specific instantiation of the general framework and characterize complementary scales of the evacuation process: M_1 captures macroscopic evacuation efficiency through the ranked exit-time sequence, M_2 characterizes microscopic bottleneck intermittency through the inter-exit time-lapse sequence, and M_3 represents mesoscopic local movement through the density–speed relation. Note that additional or alternative observables can be incorporated for other scenarios or research objectives. The first metric, M_1, is the RMSE between the ranked exit times of the simulated and experimental runs:

$$M_1\big(S(\boldsymbol{\theta}; \mathcal{D}_g), \mathcal{D}_g\big) = \sqrt{\frac{1}{N_g}\sum_{i=1}^{N_g}\Big[t_g^e\big(i; S(\boldsymbol{\theta}; \mathcal{D}_g)\big) - t_g^e\big(i; \mathcal{D}_g\big)\Big]^2} \tag{22}$$

where $t_g^e(i)$ denotes the exit time of the ith pedestrian after sorting by passage order. The second metric, M_2, measures the intermittency of bottleneck passage. The inter-exit time lapse is defined as $\Delta t_g^e(i) = t_g^e(i+1) - t_g^e(i)$. Previous evacuation experiments have reported a power-law tail in its empirical survival function, characterized by $P(T \geq \Delta t) \propto \Delta t^{-\alpha}$ (Garcimartín et al., 2016). The discrepancy is therefore given by:

$$M_2\big(S(\boldsymbol{\theta}; \mathcal{D}_g), \mathcal{D}_g\big) = \Big|\alpha_g\big(S(\boldsymbol{\theta}; \mathcal{D}_g)\big) - \alpha_g\big(\mathcal{D}_g\big)\Big| \tag{23}$$

where α_g denotes the power-law tail exponent estimated from the inter-exit time-lapse sequence. The third metric, M_3, calculates the RMSE between the simulated and experimental density–speed relation curves fitted by the Weidmann equation (Weidmann, Ulrich, 1993):

$$M_3\big(S(\boldsymbol{\theta}; \mathcal{D}_g), \mathcal{D}_g\big) = \sqrt{\frac{1}{|\mathcal{P}_\rho|}\sum_{\rho\in\mathcal{P}_\rho}\Big[v_g^W\big(\rho; S(\boldsymbol{\theta}; \mathcal{D}_g)\big) - v_g^W\big(\rho; \mathcal{D}_g\big)\Big]^2} \tag{24}$$

where $v_g^W(\rho)$ denotes the fitted Weidmann speed at density ρ, and $\mathcal{P}_\rho$ is the set of density points used for curve comparison. Given that the three metrics have different units, they are nondimensionalized as $\widetilde{M}_1 = M_1/\left\langle t_g^e\right\rangle_i$, $\widetilde{M}_2 = M_2$, and $\widetilde{M}_3 = M_3/v_f$, where $\left\langle t_g^e\right\rangle_i$ is the mean exit time of experimental group g, M_2 requires no additional scaling because it is already dimensionless, and v_f is the free-flow speed of the Weidmann relation. The final calibration loss is written as follows:

$$\mathcal{M}\big(S(\boldsymbol{\theta}; \mathcal{D}_g), \mathcal{D}_g\big) = \sum_{q=1}^{3} w_q\, \widetilde{M}_q\big(S(\boldsymbol{\theta}; \mathcal{D}_g), \mathcal{D}_g\big) \tag{25}$$

where $(w_1, w_2, w_3) = (0.35, 0.30, 0.35)$ are a priori near-equal weights controlling the relative contributions of the three constituent metrics. Starting from equal weighting, w_2 is slightly reduced since M_2 relies on a finite-sample tail-exponent estimate, and the difference is allocated equally to M_1 and M_3. Sensitivity analyses confirm that the main calibration patterns remain robust under balanced multi-metric reweighting, while parameter-level interpretation becomes more sensitive when the objective is strongly dominated by one metric.

Table 1
Admissible domain Θ for social force parameters.

Parameter	Helbing default	Lower bound	Upper bound
$v^0\,(m/s)$	1.0	0.2	2.0
$\tau\,(s)$	0.5	0.1	1.0
$A\,(N)$	2000	200	4000
$B\,(m)$	0.08	0.04	0.12
$k\,(kg/s^2)$	1.2×10^5	1.2×10^2	1.2×10^5
$\kappa\,(kg/m/s)$	2.4×10^5	2.4×10^2	2.4×10^5

The adaptive sampler then searches the admissible domain Θ (see Table 1) for social force parameters. For each selected experiment and random seed, the sampler evaluates a finite sequence of candidate parameter vectors and selects the best estimate $\widehat{\theta}_g^{\star}$ with the lowest observed loss. The total sampling budget is fixed at $N_s = 512$ evaluations per experiment–seed pair, a practical power-of-two setting that accommodates the Sobol-based sampling structure. This value is chosen to demonstrate the framework's performance by balancing search effort and computational cost. These evaluations are distributed over $R = 2$ adaptive rounds, with $J = 2$ parent sub-cubes retained after each round. At each round, the split width is set to $s = 2$; therefore, each retained parent is bisected along the two dimensions with the largest current extent before Owen-scrambled Sobol samples are drawn within the resulting child sub-cubes. All sampling operations are carried out in the normalized unit cube $[0,1]^6$ and mapped back to the physical parameter domain before forward simulation. Finally, the sampler stops early when the best observed loss falls below $\varepsilon = 0.05$.

5.2. Adaptive search for optimal social force parameters

This subsection examines whether our adaptive sampler can efficiently identify trial-specific optima in the six-dimensional parameter space of the social force model. For each of the 24 experimental trials, we search for an optimal parameter vector $\widehat{\theta}_g^{\star} = (v^0, \tau, A, B, k, \kappa)$ that minimizes the discrepancy between simulated and observed evacuation dynamics. Given that the admissible domain Θ is high-dimensional and the simulation-based objective is expensive to evaluate, parameter calibration cannot be treated simply as an exhaustive or uniform search. Instead, the sampler is designed to first explore the admissible domain and then progressively reallocate evaluations toward empirically favorable low-loss regions. The purpose of this trial-specific evaluation is to determine whether each calibrated simulator can faithfully reproduce the experimental run from which it was constructed.

Fig. 3 presents the two-stage adaptive search process for three representative trials, with each row corresponding to one competitiveness class. In Stage-1, the full parameter domain is divided through widest-dimension bisection, and Owen-scrambled Sobol samples are drawn within the resulting sub-cubes. The loss values obtained from these evaluations are then used to identify the most promising sub-region, which is retained and further subdivided in Stage-2. The first two columns of Fig. 3 visualize the selected sub-cubes after the first and second rounds of sub-division in the (k, κ) and (A, B) projections, while the third column compares the Stage-1 and Stage-2 sample distributions in the (v^0, τ) projection. Although the three representative trials occupy different regions of the parameter space, all three rows exhibit consistent qualitative patterns. Collectively, the selected sub-cubes and the kernel-density overlays show that Stage-2 samples become concentrated within the low-loss region selected from Stage-1, confirming that the adaptive retention rule shifts the computational budget from global exploration toward local refinement.

Fig. 4 summarizes the sample distribution produced by the adaptive parameter search. Each row corresponds to one representative trial from a competitiveness class, and the three columns show the projections onto (v^0, τ), (A, B), and (k, κ). The evaluated samples do not uniformly fill the normalized parameter space; instead, they form non-uniform clouds and density contours, reflecting that later evaluations are redirected toward the sub-regions retained by the adaptive search. The red star marks the parameter vector with the lowest loss among all evaluated samples. Its position is not determined directly by the sampling density, because the true optimum is unknown and the sampler must balance exploration of the admissible domain with refinement of promising regions. Nevertheless, the adaptive rule allocates more samples to sub-regions that are more likely to contain low-loss solutions. The differences in both the sampled regions and the best-sample locations across rows indicate that the sampler adapts to trial-specific loss landscapes rather than converging to a single parameter configuration shared across different competitiveness classes.

To quantify whether the observed search behavior translates into effective optimization, we evaluate the adaptive sampler at three levels in Fig. 5. First, Fig. 5(a) compares the sampler on the six-dimensional Rosenbrock and Zakharov benchmark functions to assess its effectiveness independently of the evacuation simulator. The Rosenbrock function is characterized by a long curved valley around the global optimum (Rosenbrock, 1960), whereas the Zakharov function has coupled and differently weighted coordinates (Jamil and Yang, 2013). Standard Monte Carlo (MC), Sobol, and Latin hypercube sampling (LHS) are used for comparison (Eidehall and Petersson, 2008; Singhee and Rutenbar, 2010). All three distribute evaluations uniformly across the admissible domain, but they do not use the observed loss to guide subsequent evaluations. Across the evaluation budget, the adaptive sampler consistently attains lower running-best loss than all three methods on both functions, while the comparison methods remain close to one another. The consistent result across the two benchmark structures supports the effectiveness of reallocating a finite evaluation budget toward low-loss regions. Second, across the three competitiveness classes in Fig. 5(b), the relative loss decreases consistently with the evaluation budget, showing similar convergence across different behavioral conditions. Finally, Fig. 5(c) quantifies the reduction in calibration loss relative to the Helbing default parameterization, which serves as an uncalibrated reference. All trials show substantial reductions in loss, with most exceeding 60%, and high-competitiveness trials generally exhibit larger reductions. These results characterize the

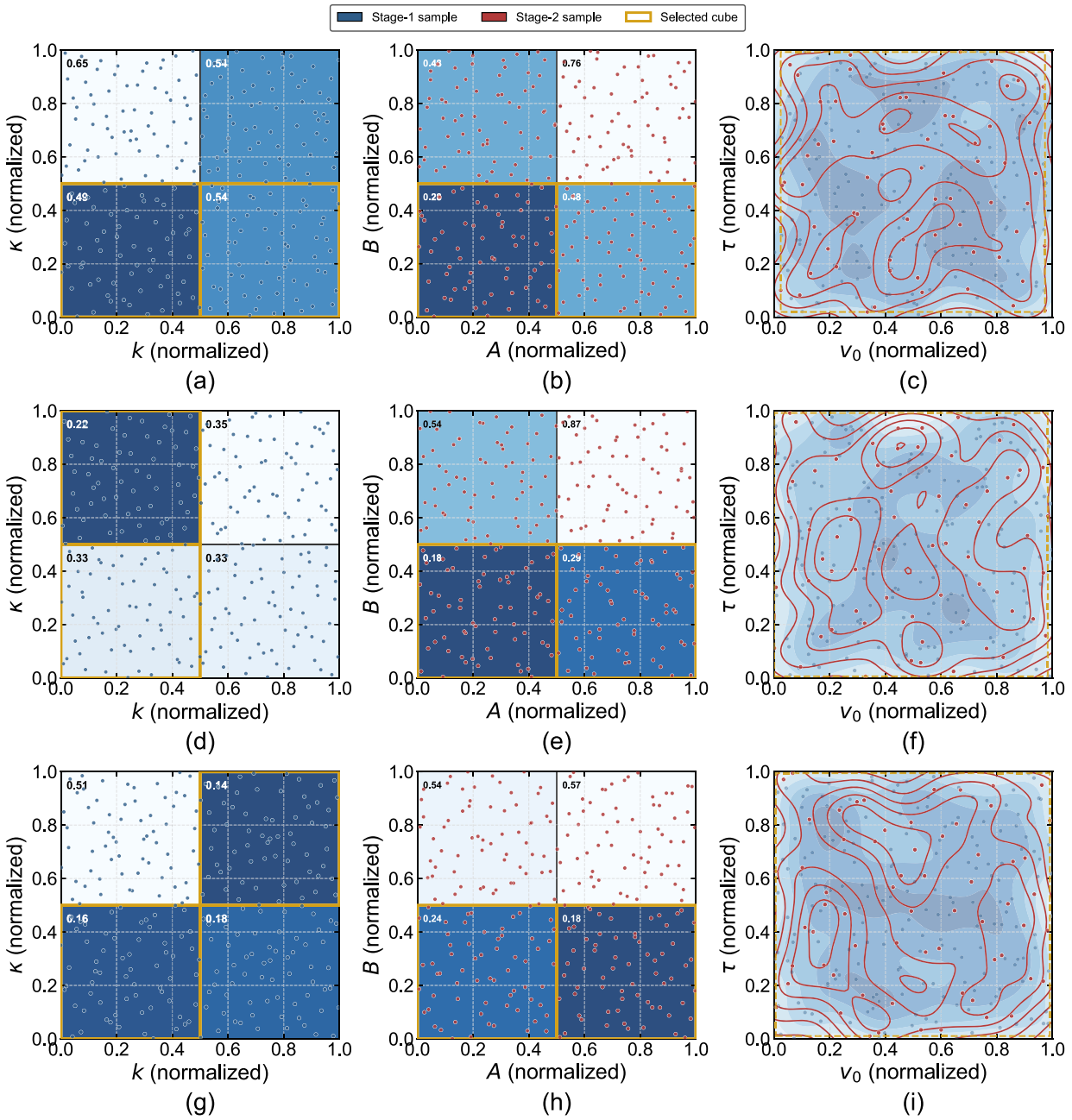

Fig. 3. Two-stage adaptive search process in the parameter space. From top to bottom, rows correspond to low-, medium-, and high-competitiveness representative trials. From left to right, columns show the Stage-1 sub-cubes in (k, κ), the Stage-2 sub-cubes in (A, B), and the overlaid Stage-1 and Stage-2 sample densities in (v^0, τ). The blue and red samples respectively denote Stage-1 and Stage-2 evaluations, and the yellow outline indicates the selected sub-cube in the corresponding adaptive stage.

in-sample improvement achieved through trial-specific calibration, showing how parameter adjustment reduces discrepancies with the observed evacuation dynamics.

5.3. *Numerical simulations reproduce evacuation dynamics*

In this part, we further investigate whether these parameters produce valid simulations of the experimental evacuation process, because a lower calibration loss alone does not guarantee that the simulated trajectories preserve the observable statistics of the real experiments. Consequently, we evaluate reproduction at two complementary levels: 1) agreement with the three observational scales

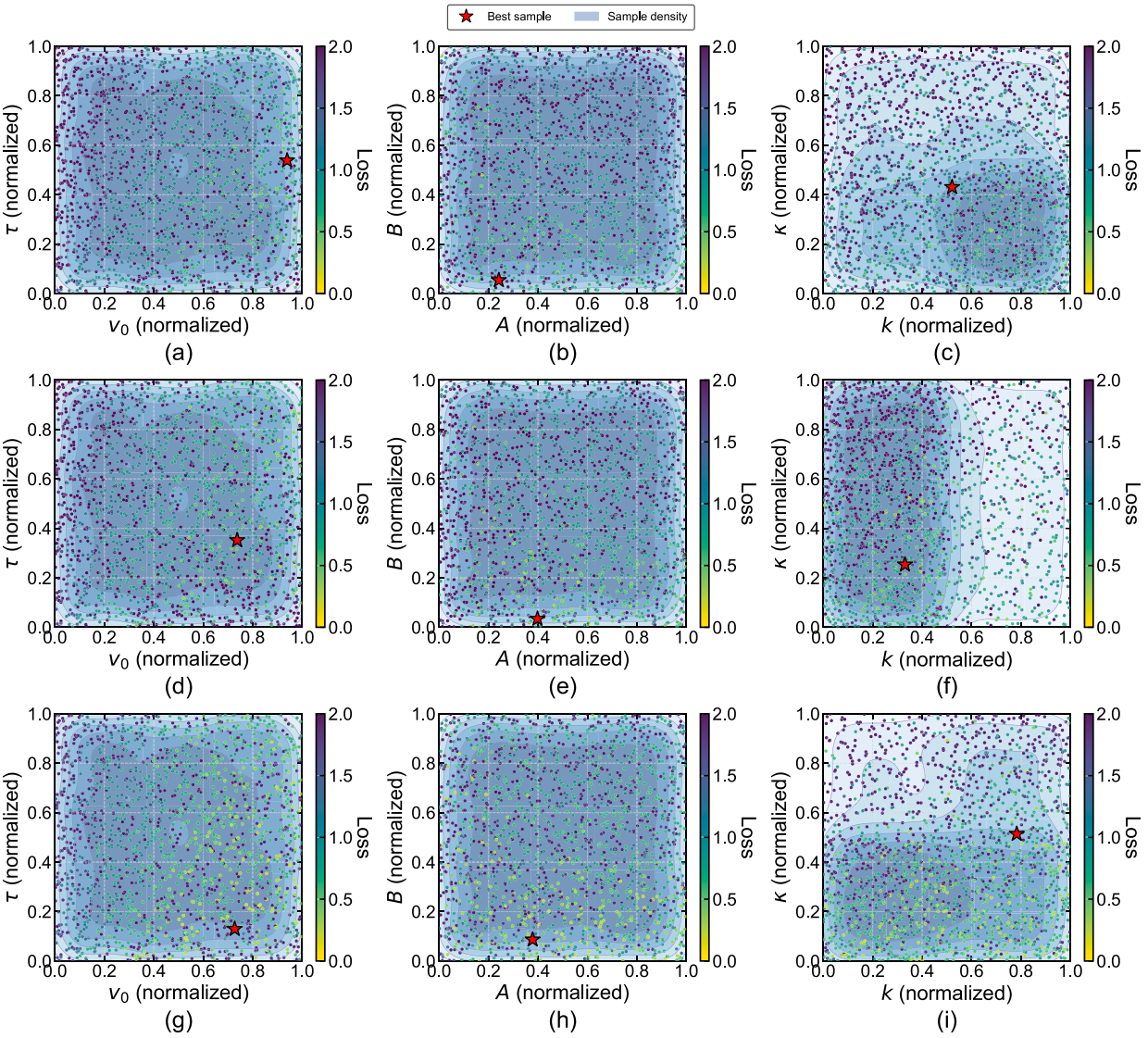

Fig. 4. Sample distribution after adaptive parameter search. From top to bottom, rows correspond to low-, medium-, and high-competitiveness representative trials. From left to right, columns show the parameter projections onto (v^0, τ), (A, B), and (k, κ). The scatter points denote evaluated parameter vectors, with color indicating the total loss. The blue contours represent the density of the sampled parameter cloud, and the red star marks the best sample for the corresponding trial.

included in the calibration objective, and 2) recovery of the between-class differences in competitiveness-related quantities that are not directly optimized during calibration.

The calibrated simulations reproduce the main empirical patterns across the three observational scales of the calibration objective in Fig. 6. The cumulative passage curves track the observed exit progression, the time-lapse survival functions reflect the heavy-tailed decay of inter-passage intervals, and the density–speed relations follow a similar downward trend with local density. Overall, the simulated results broadly overlap with the experimental data across the three scales, indicating that the calibrated simulator preserves both temporal discharge dynamics and local movement statistics rather than merely fitting a single scalar loss. Notably, the simulated results typically span a slightly broader range than the experimental observations, and this broader envelope can be attributed to the different ways in which local variability is propagated in experiments and simulations. In real evacuations, bodily coordination, visual anticipation, local yielding, and continuous posture adjustment can absorb small perturbations before they accumulate. In numerical simulations, by contrast, pedestrians are discretely initialized rigid-disk individuals governed by simplified avoidance and contact rules, and small differences are more likely to be amplified into visible trajectory-level deviations.

To assess whether the calibrated simulator also preserves characteristics not explicitly optimized, Fig. 7 further evaluates whether the calibrated simulations reproduce the quantitative characterization of competitiveness, with pairwise differences among competi-

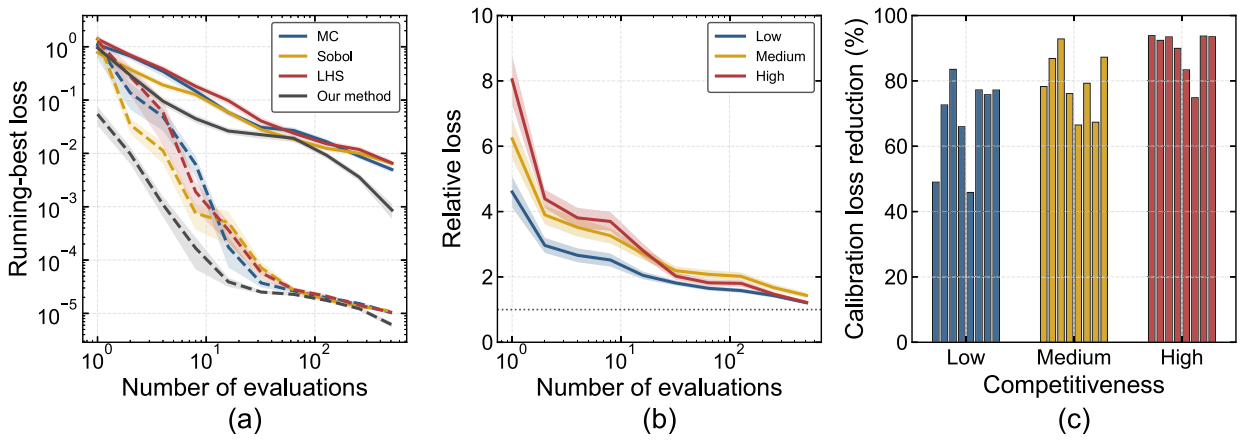

Fig. 5. Performance evaluation and calibration outcomes of the adaptive parameter search. (a) Normalized loss of MC, Sobol, LHS, and the adaptive sampler on the six-dimensional Rosenbrock (solid lines) and Zakharov (dashed lines) benchmark functions. (b) Relative loss for the low-, medium-, and high-competitiveness classes. (c) In-sample reduction in calibration loss relative to the uncalibrated Helbing default parameterization for each trial.

Table 2
Values of calibrated social force parameters across competitiveness levels.

Parameter	Low	Medium	High
$v^0\,(m/s)$	1.76 ± 0.17	1.53 ± 0.36	1.46 ± 0.38
$\tau\,(s)$	0.45 ± 0.21	0.42 ± 0.27	0.35 ± 0.18
$A\,(N)$	1779 ± 1342	1238 ± 746	818 ± 455
$B\,(m)$	0.052 ± 0.009	0.057 ± 0.016	0.058 ± 0.014
$k\,(kg/s^2)$	$(6.16 \pm 2.82)\times10^4$	$(8.06 \pm 2.25)\times10^4$	$(8.28 \pm 2.05)\times10^4$
$\kappa\,(kg/m/s)$	$(9.24 \pm 8.19)\times10^4$	$(5.85 \pm 5.01)\times10^4$	$(4.43 \pm 3.67)\times10^4$
$\mathcal{M}^\star$	0.48 ± 0.30	0.25 ± 0.09	0.18 ± 0.08

tiveness classes assessed using two-sided Mann–Whitney U tests. None of these three quantities enters the calibration objective, so this comparison provides an out-of-objective consistency check on the calibrated simulations. The experimental and simulated data exhibit consistent competitiveness-related trends: the average density, initial speed, and crowd pressure (Moussaïd et al., 2011) increase as competitiveness grows from low to high, although the difference in average density between the medium- and high-competitiveness cases is not statistically significant. This agreement indicates that the calibrated simulations capture the main quantitative signature of competitiveness beyond the quantities directly used for calibration. However, the simulated values are generally lower than the experimental values, and the between-class differences are weaker, especially for initial speed and crowd pressure. This attenuation is likely related to the assumptions of behavioral homogeneity and rigid-disk contact mechanics. In the experiments, individuals can express competitiveness through heterogeneous strategies, shoulder rotation, and body elastic compression. However, the calibrated simulations represent each class using a shared behavioral parameterization and disk-like agents with limited deformability. These assumptions constrain local packing and contact intensity, allowing the simulations to retain the direction of the competitiveness-related trends while underestimating the magnitude of these quantities. These residual discrepancies also indicate that parameter calibration can compensate for model mismatch only within the degrees of freedom provided by the social force formulation, whereas structural features absent from the model cannot be recovered through parameter adjustment alone.

In summary, the calibrated simulator preserves the main temporal, local, and class-dependent patterns of the experiments, effectively reproducing the experimental evacuation statistics across both the calibration-objective scales and the quantitative characterization of competitiveness. The existing discrepancies, such as the broader simulated trajectory envelope and weaker between-class separation, mainly reflect systematic limitations of the simplified modeling assumptions rather than random calibration failure.

5.4. Calibrated parameters encode competitiveness mechanisms

Having shown that the calibrated simulator reproduces the experimental evacuation statistics, we next examine how the prescribed competitiveness conditions are associated with the fitted parameter patterns within the social force model. We first analyze the variation of each parameter across competitiveness levels and then examine their joint structure to determine whether the observed conditions are associated with a single dominant model parameter or with coordinated changes among multiple model components. The calibrated parameters are treated here as model-dependent effective coefficients rather than as direct measurements of the participants' psychological or biomechanical properties.

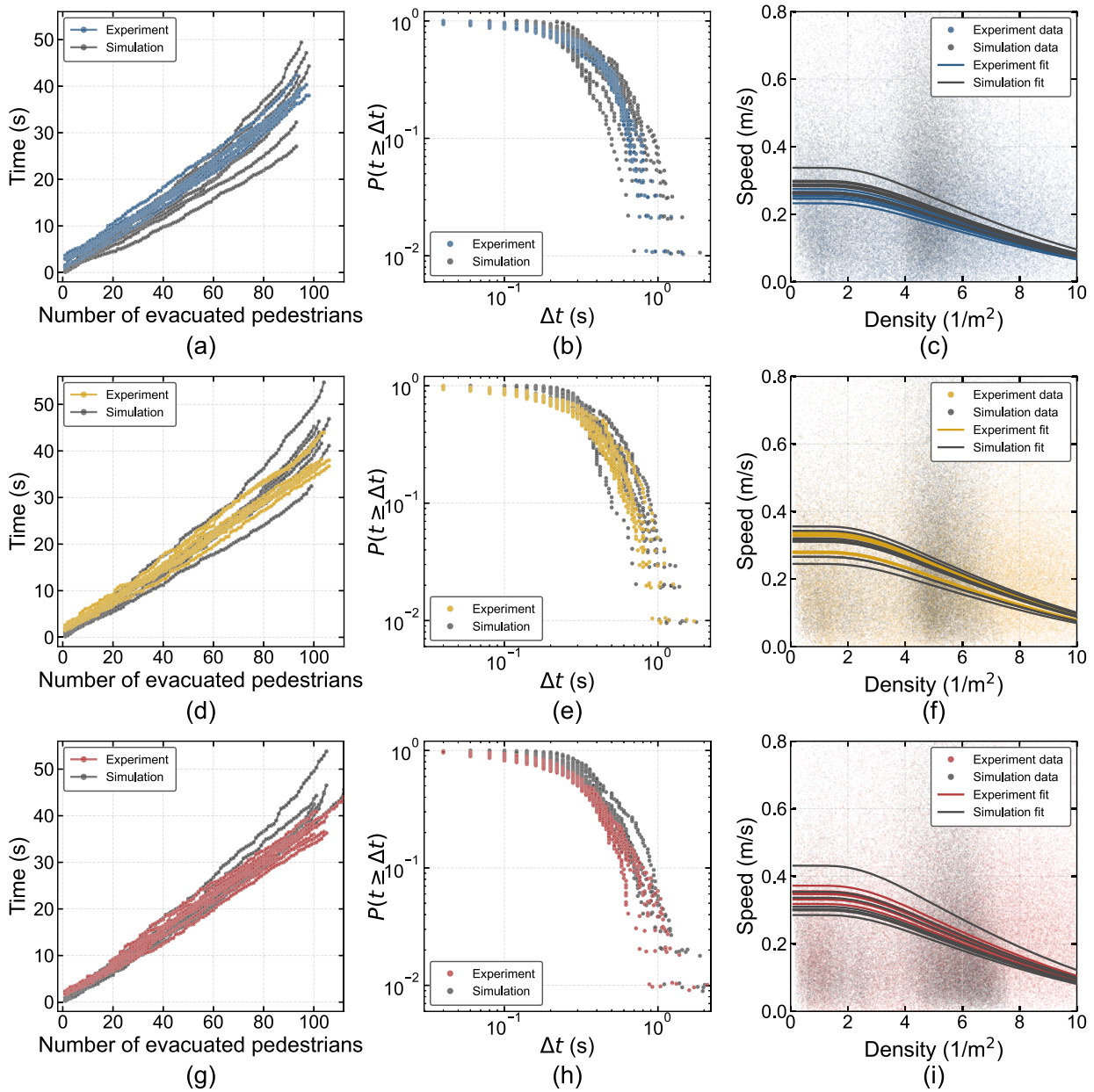

Fig. 6. Comparison between experimental and simulated data on the three observational scales of the calibration objective. From top to bottom, rows correspond to low-, medium-, and high-competitiveness classes. From left to right, columns show the cumulative passage curve, the time-lapse survival function, and the density–speed fundamental diagram.

Fig. 8 indicates that the six calibrated parameters vary systematically across competitiveness classes, with the corresponding numerical values reported in Table 2. A permutation-based multivariate regression reveals a significant ordered trend in the six-dimensional calibrated parameter configuration with increasing competitiveness ($F_{(1,22)} = 2.56$, $P < 0.05$). This trend is unlikely to be explained by unequal within-class dispersion, as a permutational analysis of multivariate dispersions shows no significant difference among the three classes ($F_{(2,21)} = 0.52$, $P > 0.05$). For self-driven adjustment, both v^0 and τ decrease from low to high competitiveness. The decline in v^0 cautions against equating competitiveness with a higher desired speed, since visible blockage and strong local interactions may reduce the attainable speed in congested evacuation. Instead, high competitiveness is associated with a shorter relaxation time, corresponding to faster modeled velocity adjustment. For psychological repulsion, A decreases markedly while B remains relatively stable or slightly increases, suggesting that the calibrated model represents the high-competitiveness condition through a reduced interpersonal-repulsion term and greater modeled tolerance of close proximity. For physical contact, k increases whereas κ decreases, implying stronger resistance under body compression but weaker resistance to tangential sliding, thereby fa-

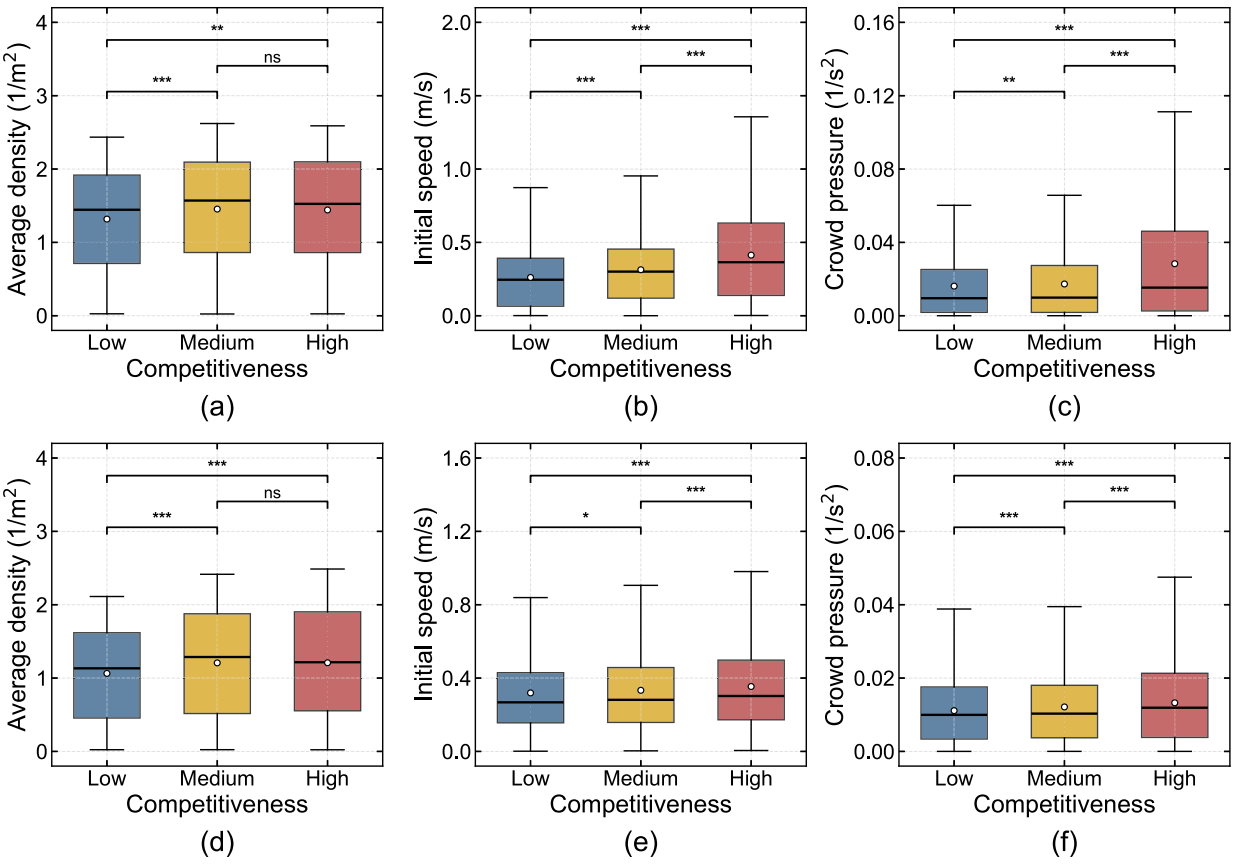

Fig. 7. Comparison between experimental and simulated data on the quantitative characterization of competitiveness. From top to bottom, rows show the experimental observations and calibrated simulations. From left to right, columns correspond to average density, initial speed, and crowd pressure. The central thick lines and open circles represent the median lines and mean values, respectively. $^{***}P < 0.001$; $^{**}P < 0.01$; $^{*}P < 0.05$; ns, $P > 0.05$.

cilitating simulated movement through small temporary gaps near the bottleneck. These trends provide preliminary evidence that high-competitiveness behavior is represented by a coordinated parameter pattern of faster velocity adjustment, weaker psychological avoidance, greater compression feedback, and reduced resistance to lateral sliding. We further analyze loss-weighted ensembles of near-optimal parameter vectors at low-loss scales of 5%, 7.5%, and 10%. Across these settings, the median six-dimensional vector displacement remains below 0.040 of the normalized search range, all six parameter-change directions are preserved, and the contrast-vector cosine similarity remains above 0.967. These results confirm the robustness of the principal competitiveness-related parameter patterns within the explored low-loss region.

The marginal trends in Fig. 8 show that competitiveness is encoded through multiple model parameters, but they do not show whether these parameters jointly form a coherent structure. We therefore analyze the calibrated parameter ensemble in the multivariate space in Fig. 9. The unsupervised Varimax-rotated PCA (Jolliffe and Cadima, 2016) in Fig. 9(a) and (b) reveals the intrinsic covariance structure of the six parameters without using class labels. The first rotated factor is mainly associated with opposite loadings of A and κ against B, with a weaker contribution from τ, whereas the second rotated factor mainly separates the self-driven parameters τ and v^0 from the contact stiffness k. The displacement of class centroids in this unsupervised factor space indicates that competitiveness is partly reflected in the natural covariation of the calibrated parameters. The supervised LDA (Fisher, 1936) in Fig. 9(c) and (d) then identifies the parameter combinations that best discriminate the competitiveness classes. The standardized discriminant coefficients show that A, τ, and v^0 make the dominant contributions to class separation, while k and κ provide secondary but non-negligible contributions associated with physical-contact mechanics. In contrast, B remains close to zero on both discriminant axes, indicating that interaction range plays a weaker role in distinguishing competitiveness levels. These results show that the parameter changes identified in Fig. 8 form a coordinated multivariate signature whose contributions are unequally distributed

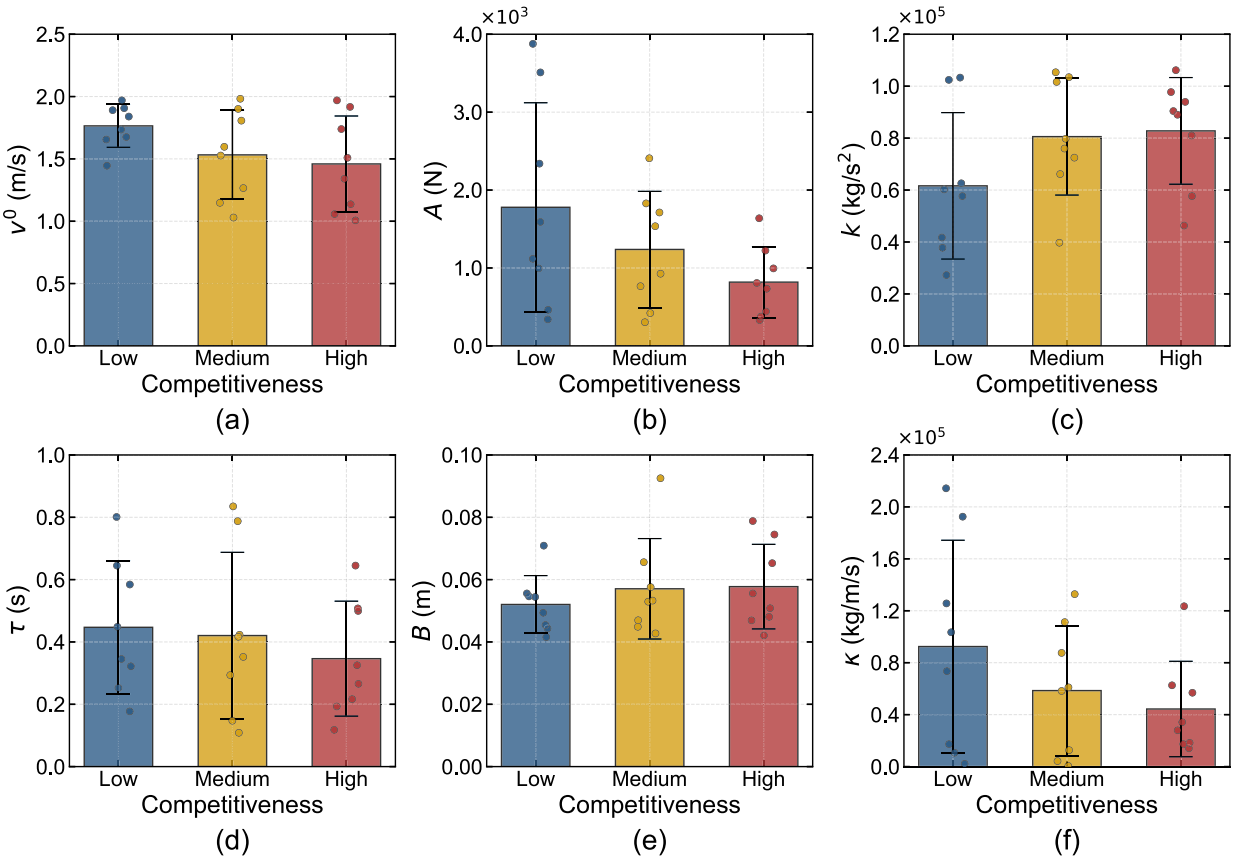

Fig. 8. Comparison of calibrated social force parameters across competitiveness levels. (a) Desired speed v^0. (b) Repulsive strength A. (c) Body compression coefficient k. (d) Characteristic time τ. (e) Interaction range B. (f) Sliding friction coefficient κ. The horizontal markers denote class means, error bars indicate within-class standard deviations, and scatter points represent trial values.

across parameters. Competitiveness is discriminated jointly by self-driven adjustment, psychological repulsion, and physical contact, with no single parameter dominating.

To sum up, the calibrated parameters provide an interpretable encoding of the experimental competitiveness levels within the social force model. The progression from low to high competitiveness is reflected not only by systematic shifts in individual parameters, but also by a coherent multivariate structure in the calibrated parameter space. Therefore, the proposed framework does not merely return fitted parameter values; it reveals how prescribed behavioral conditions are encoded within the model through coordinated changes in self-driven adjustment, psychological repulsion, and physical-contact mechanisms.

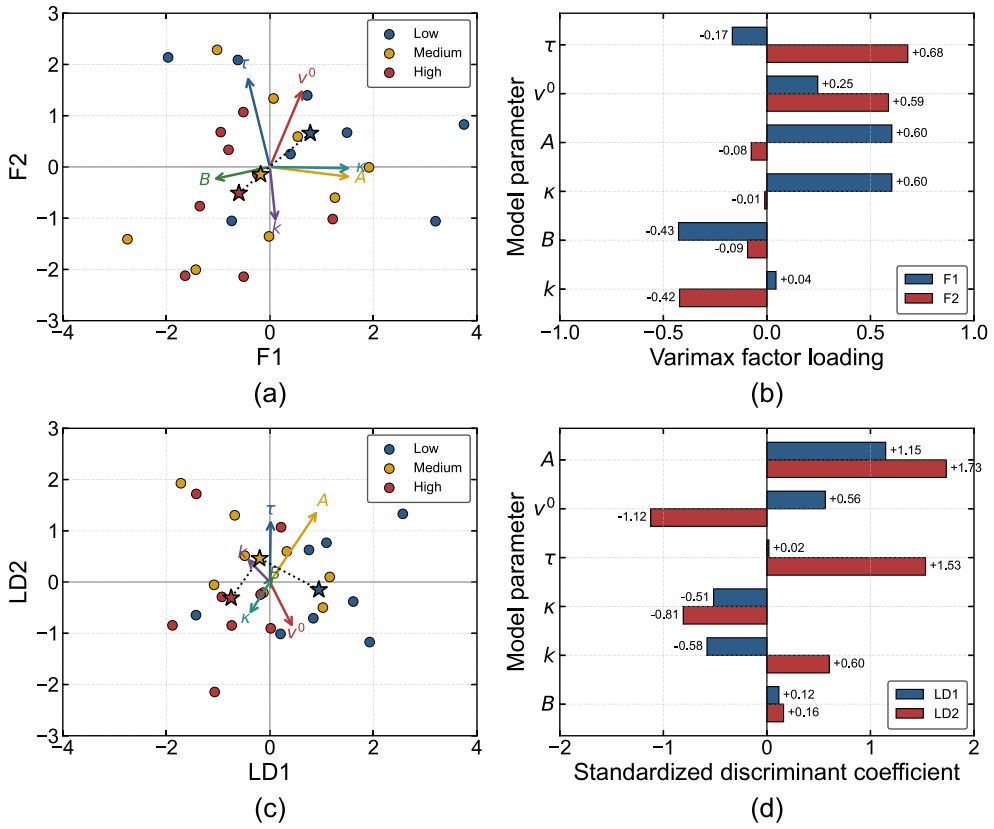

Fig. 9. Multivariate structure of the calibrated parameter ensemble. (a) Unsupervised Varimax-rotated PCA biplot with class centroids and parameter loadings. (b) Varimax factor loadings on the two rotated factors. (c) Supervised LDA biplot with class centroids and parameter vectors. (d) Standardized discriminant coefficients on the two discriminant axes.

6. Conclusions

In this work, we propose an integrated framework that uses adaptive calibration to align social force simulations with empirical evacuation experiments. The framework is motivated by a central gap in evacuation research: controlled experiments provide detailed but limited observations of pedestrian behavior, whereas numerical simulations can extend these observations only when their parameters are systematically constrained by the same empirical evidence. To address this, the framework integrates scenario reconstruction, forward simulation, discrepancy evaluation, and adaptive sampling into a closed-loop calibration procedure. By reconstructing each experimental run as a reusable simulation context, evaluating candidate parameter vectors against empirical trajectories across multiple observational scales, and adaptively concentrating the sampling budget toward low-loss regions, the framework converts limited experimental data into calibrated simulators that can be systematically re-evaluated once the original run has been recorded. At the algorithmic level, the adaptive sampler jointly searches the high-dimensional parameter vector of the social force model, allowing the calibration to account for the coupled effects of desired motion, interpersonal repulsion, and physical-contact mechanisms. At the interpretive level, the resulting calibrated parameters serve as model-dependent structured representations of how behavioral conditions are encoded, going beyond black-box reproduction of the observed data.

The proposed framework has two main practical implications for evacuation modeling and crowd safety assessment. First, it provides a way to extend costly controlled experiments beyond the conditions that can be safely tested with human participants. It complements analyses based on prescribed parameter distributions or worst-case assumptions by providing an empirically constrained basis for parameterization. Once an experimental run has been reconstructed and calibrated, the resulting simulator can serve as an empirically grounded reference for subsequent investigations of facility layouts, initial crowd configurations, or behavioral assumptions. This is particularly valuable for extreme scenarios involving high compression, strong pushing, or potentially dangerous contact dynamics (Moussaïd et al., 2016; Xie et al., 2024), which cannot be ethically reproduced in human experiments but can be explored computationally under controlled hypothetical conditions. The applicability of such analyses should, however, be assessed against the conditions and purpose of the intended application. Second, it supports mechanism-level interpretation and model diagnosis. As the full parameter vector is calibrated jointly, the resulting parameter combinations can be examined as structured representations of the behavioral and mechanical assumptions required for the model to reproduce empirical dynamics (Kretz et al., 2018). This makes it possible to assess whether simulated outcomes arise from plausible model mechanisms, to identify which model components are

most responsible for reproducing observed patterns, and to diagnose when a visually plausible simulation is actually inconsistent with empirical evidence (Gödel et al., 2022). In this sense, these capabilities provide a practical bridge from limited experimental observations to broader scenario analysis, mechanism interpretation, and evidence-based evacuation assessment.

Finally, several limitations should be acknowledged. First, the current framework assumes that all pedestrians within the same experimental run share a homogeneous parameter vector. This reduces the dimensionality of the estimation problem, but ignores individual heterogeneity in real crowds (Wu et al., 2022a,b). Subsequent research could extend the framework toward individual-level or distributional parameter calibration, with appropriate regularization or hierarchical inference to control the increasing dimensionality. Second, the adaptive sampler provides a finite-budget approximation, without a global-optimum guarantee (Jones et al., 1998). Its performance may depend on the sampling budget, adaptive rounds, retained sub-cubes, split strategy, and metric weights. Future work may quantify uncertainty around calibrated solutions, compare the sampler with other high-dimensional optimization methods, and develop budget-allocation rules that adapt to the observed loss landscape. Third, the calibrated parameters are derived from a single experimental geometry and dataset, deliberately holding the physical environment fixed to isolate competitiveness-related variation. While this controlled design facilitates clearer interpretation within the present case study, the identified parameter patterns should not be assumed to generalize unchanged across different facility or crowd configurations. Accordingly, the transferable element is primarily the calibration workflow, while the applicability of a calibrated parameter vector to other conditions should be evaluated separately. Fourth, calibration within any fixed mechanistic formulation is constrained by the model's structural assumptions, and fitted parameters may partly compensate for discrepancies that the model cannot represent explicitly. The calibrated values should therefore be interpreted as model-dependent effective coefficients rather than uniquely identifiable psychological or biomechanical quantities. Future work could assess this structural uncertainty across alternative model formulations. Despite these limitations, we show that systematic full-vector calibration can make social force simulations more empirically grounded and interpretable, providing a basis for computationally extending the utility of evacuation experiments.

CRediT authorship contribution statement

Wenhan Wu: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization; **Jingwei Ge:** Writing – review & editing, Writing – original draft, Supervision, Software, Project administration, Methodology, Data curation, Conceptualization, Validation; **Li Li:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Funding acquisition.

Acknowledgements

This work was supported by the Research and Application of Traffic Signal Control Optimization Based on New Energy Vehicle Data and Large-Small Model Collaboration (No. Z261100003326006).

Data availability

Data will be made available on request.

References

Bassoli, E., Vincenzi, L., 2021. Parameter calibration of a social force model for the crowd-induced vibrations of footbridges. Front. Built Environ. 7. https://doi.org/10.3389/fbuil.2021.656799

Boltes, M., Zhang, J., Tordeux, A., Schadschneider, A., Seyfried, A., 2019. Empirical Results of Pedestrian and Evacuation Dynamics. Springer US. pp. 671–699. https://doi.org/10.1007/978-1-4939-8763-4_706

Chen, X., Treiber, M., Kanagaraj, V., Li, H., 2017. Social force models for pedestrian traffic – state of the art. Transp. Rev. 38 (5), 625–653. https://doi.org/10.1080/01441647.2017.1396265

Corbetta, A., Muntean, A., Vafayi, K.a., 2015. Parameter estimation of social forces in pedestrian dynamics models via a probabilistic method. Math. Biosci. Eng. 12 (2), 337–356. https://doi.org/10.3934/mbe.2015.12.337

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. https://doi.org/10.1016/j.simpat.2018.08.002

Dong, H., Zhou, M., Wang, Q., Yang, X., Wang, F.-Y., 2020. State-of-the-art pedestrian and evacuation dynamics. IEEE Trans. Intell. Transp. Syst. 21 (5), 1849–1866. https://doi.org/10.1109/tits.2019.2915014

Echeverría-Huarte, I., Shi, Z., Garcimartín, A., Zuriguel, I., 2022. Pedestrian bottleneck flow when keeping a prescribed physical distance. Phys. Rev. E 106 (4), 044302. https://doi.org/10.1103/physreve.106.044302

Eidehall, A., Petersson, L., 2008. Statistical threat assessment for general road scenes using monte carlo sampling. IEEE Trans. Intell. Transp. Syst. 9 (1), 137–147. https://doi.org/10.1109/tits.2007.909241

Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. Ann. Eugen. 7 (2), 179–188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x

Garcimartín, A., Parisi, D.R., Pastor, J.M., Martín-Gómez, C., Zuriguel, I., 2016. Flow of pedestrians through narrow doors with different competitiveness. J. Stat. Mech.: Theory Exp. 2016 (4), 043402. https://doi.org/10.1088/1742-5468/2016/04/043402

Garcimartín, A., Pastor, J.M., Martín-Gómez, C., Parisi, D., Zuriguel, I., 2017. Pedestrian collective motion in competitive room evacuation. Sci. Rep. 7 (1). https://doi.org/10.1038/s41598-017-11197-x

Gwynne, S.M.V., Kuligowski, E.D., Boyce, K.E., Nilsson, D., Robbins, A.P., Lovreglio, R., Thomas, J.R., Roy-Poirier, A., 2017. Enhancing egress drills: preparation and assessment of evacuee performance. Fire Mater. 43 (6), 613–631. https://doi.org/10.1002/fam.2448

Gödel, M., Bode, N., Köster, G., Bungartz, H.-J., 2022. Bayesian inference methods to calibrate crowd dynamics models for safety applications. Saf. Sci. 147, 105586. https://doi.org/10.1016/j.ssci.2021.105586

Haghani, M., 2020. Empirical methods in pedestrian, crowd and evacuation dynamics: Part I. experimental methods and emerging topics. Saf. Sci. 129, 104743. https://doi.org/10.1016/j.ssci.2020.104743

Haghani, M., Sarvi, M., 2018. Crowd behaviour and motion: empirical methods. Transp. Res. B: Methodol. 107, 253–294. https://doi.org/10.1016/j.trb.2017.06.017
Helbing, D., Buzna, L., Johansson, A., Werner, T., 2005. Self-organized pedestrian crowd dynamics: experiments, simulations, and design solutions. Transp. Sci. 39 (1), 1–24. https://doi.org/10.1287/trsc.1040.0108
Helbing, D., Farkas, I., Vicsek, T., 2000. Simulating dynamical features of escape panic. Nature 407 (6803), 487–490. https://doi.org/10.1038/35035023
Helbing, D., Johansson, A., 2009. Pedestrian, Crowd and Evacuation Dynamics. Springer New York. pp. 6476–6495. https://doi.org/10.1007/978-0-387-30440-3_382
Helbing, D., Molnár, P., 1995. Social force model for pedestrian dynamics. Phys. Rev. E 51 (5), 4282–4286. https://doi.org/10.1103/physreve.51.4282
Huo, F., Song, W., Chen, L., Liu, C., Liew, K.M., 2016. Experimental study on characteristics of pedestrian evacuation on stairs in a high-rise building. Saf. Sci. 86, 165–173. https://doi.org/10.1016/j.ssci.2016.02.025
Jamil, M., Yang, X.-S., 2013. A literature survey of benchmark functions for global optimisation problems. Int. J. Math. Model. Numer. Optim. 4 (2), 150–194.
Johansson, A., Helbing, D., Shukla, P.K., 2007. Specification of the social force pedestrian model by evolutionary adjustment to video tracking data. Adv. Complex Syst. 10 (supp02), 271–288. https://doi.org/10.1142/s0219525907001355
Jolliffe, I.T., Cadima, J., 2016. Principal component analysis: a review and recent developments. Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci. 374 (2065), 20150202. https://doi.org/10.1098/rsta.2015.0202
Jones, D.R., Schonlau, M., Welch, W.J., 1998. Efficient global optimization of expensive black-box functions. J. Glob. Optim. 13 (4), 455–492. https://doi.org/10.1023/a:1008306431147
Ko, M., Kim, T., Sohn, K., 2012. Calibrating a social-force-based pedestrian walking model based on maximum likelihood estimation. Transportation 40 (1), 91–107. https://doi.org/10.1007/s11116-012-9411-z
Kretz, T., Lohmiller, J., Sukennik, P., 2018. Some indications on how to calibrate the social force model of pedestrian dynamics. Transp. Res. Rec.: J. Transp. Res. Board 2672 (20), 228–238. https://doi.org/10.1177/0361198118786641
Li, M., Zhao, Y., He, L., Chen, W., Xu, X., 2015. The parameter calibration and optimization of social force model for the real-life 2013 ya'an earthquake evacuation in china. Saf. Sci. 79, 243–253. https://doi.org/10.1016/j.ssci.2015.06.018
Li, T., Xu, B., Lu, W., Chen, Z., Zhang, S., Xia, F., 2024. The parameter calibration of social force model for pedestrian flow simulation based on YOLOv5. Sensors 24 (15), 5011. https://doi.org/10.3390/s24155011
Ma, L., Chen, B., Wang, X., Zhu, Z., Wang, R., Qiu, X., 2019. The analysis on the desired speed in social force model using a data driven approach. Phys. A: Stat. Mech. Appl. 525, 894–911. https://doi.org/10.1016/j.physa.2019.03.087
Moussaïd, M., Helbing, D., Theraulaz, G., 2011. How simple rules determine pedestrian behavior and crowd disasters. Proc. Natl. Acad. Sci. 108 (17), 6884–6888. https://doi.org/10.1073/pnas.1016507108
Moussaïd, M., Kapadia, M., Thrash, T., Sumner, R.W., Gross, M., Helbing, D., Hölscher, C., 2016. Crowd behaviour during high-stress evacuations in an immersive virtual environment. J. R. Soc. Interface 13 (122), 20160414. https://doi.org/10.1098/rsif.2016.0414
Niederreiter, H., 1992. Random Number Generation and Quasi-Monte Carlo Methods. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611970081
Owen, A.B., 1997. Scrambled net variance for integrals of smooth functions. Ann. Stat. 25 (4). https://doi.org/10.1214/aos/1031594731
Pastor, J.M., Garcimartín, A., Gago, P.A., Peralta, J.P., Martín-Gómez, C., Ferrer, L.M., Maza, D., Parisi, D.R., Pugnaloni, L.A., Zuriguel, I., 2015. Experimental proof of faster-is-slower in systems of frictional particles flowing through constrictions. Phys. Rev. E 92 (6), 062817. https://doi.org/10.1103/physreve.92.062817
Qi, Q., She, R., Liu, H., Zhang, J., Zhang, Y., 2022. Social force model calibration for preschool children evacuations based on multiscenario experiments. J. Adv. Transp. 2022, 1–8. https://doi.org/10.1155/2022/2934884
Rosenbrock, H.H., 1960. An automatic method for finding the greatest or least value of a function. Comput. J. 3 (3), 175–184.
Rudloff, C., Matyus, T., Seer, S., 2013. Comparison of Different Calibration Techniques on Simulated Data. Springer International Publishing. pp. 657–672. https://doi.org/10.1007/978-3-319-02447-9_55
Seer, S., Rudloff, C., Matyus, T., Brändle, N., 2014. Validating social force based models with comprehensive real world motion data. Transp. Res. Procedia 2, 724–732. https://doi.org/10.1016/j.trpro.2014.09.080
Shiwakoti, N., Shi, X., Ye, Z., 2019. A review on the performance of an obstacle near an exit on pedestrian crowd evacuation. Saf. Sci. 113, 54–67. https://doi.org/10.1016/j.ssci.2018.11.016
Singhee, A., Rutenbar, R.A., 2010. Why quasi-monte carlo is better than monte carlo or latin hypercube sampling for statistical circuit analysis. IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst. 29 (11), 1763–1776. https://doi.org/10.1109/tcad.2010.2062750
Sticco, I.M., Frank, G.A., Dorso, C.O., 2021. Social force model parameter testing and optimization using a high stress real-life situation. Phys. A: Stat. Mech. Appl. 561, 125299. https://doi.org/10.1016/j.physa.2020.125299
Sun, H., Zhu, K., Wang, G., Hu, H., Guo, P., Wu, K., Zhang, T., 2025. Experimental study on pedestrian evacuation characteristics through building bottleneck group structure. Dev. Built Environ. 23, 100734. https://doi.org/10.1016/j.dibe.2025.100734
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). IEEE, pp. 2026–2031. https://doi.org/10.1109/tmee.2011.6199614
Tao, Z., Li, Y., Huang, X., Wang, Y., Chen, M., Yang, R., 2024. Experimental study on the factors affecting pedestrian exit selection on the basis of the mixed reality evacuation LVC simulation system. Appl. Sci. 14 (13), 5741. https://doi.org/10.3390/app14135741
Weidmann, Ulrich, 1993. Transporttechnik der Fussgänger: Transporttechnische Eigenschaften des Fussgängerverkehrs, Literaturauswertung. Technical Report. https://doi.org/10.3929/ETHZ-B-000242008
Wu, W., Chen, M., Li, J., Liu, B., Zheng, X., 2022a. An extended social force model via pedestrian heterogeneity affecting the self-driven force. IEEE Trans. Intell. Transp. Syst. 23 (7), 7974–7986. https://doi.org/10.1109/tits.2021.3074914
Wu, W., Li, J., Yi, W., Zheng, X., 2022b. Modeling crowd evacuation via behavioral heterogeneity-based social force model. IEEE Trans. Intell. Transp. Syst. 23 (9), 15476–15486. https://doi.org/10.1109/tits.2022.3140823
Wu, W., Zheng, X., 2024. A systematic analysis of subgroup research in pedestrian and evacuation dynamics. IEEE Trans. Intell. Transp. Syst. 25 (2), 1225–1246. https://doi.org/10.1109/tits.2023.3318417
Xie, W., Lee, E. W.M., Cheng, Y., Shi, M., Cao, R., Zhang, Y., 2020. Evacuation performance of individuals and social groups under different visibility conditions: experiments and surveys. Int. J. Disaster Risk Reduct. 47, 101527. https://doi.org/10.1016/j.ijdrr.2020.101527
Xie, W., Lee, E. W.M., Li, T., Shi, M., Cao, R., Zhang, Y., 2021. A study of group effects in pedestrian crowd evacuation: experiments, modelling and simulation. Saf. Sci. 133, 105029. https://doi.org/10.1016/j.ssci.2020.105029
Xie, W., Wei, Y., Xuan, C., Ma, Y., Ming Lee, E.W., Li, T., Cheng, Y., 2024. Understanding pedestrian crowd dynamics on ramps: an empirical study triggered by seoul halloween crowd crush. Int. J. Disaster Risk Reduct. 109, 104601. https://doi.org/10.1016/j.ijdrr.2024.104601
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. https://doi.org/10.1016/j.sbspro.2014.07.233
Zhang, L., Wu, X., Lin, H., Zhang, M., Liu, Y., 2024. Experimental study on the synchronization mechanism and trigger characteristic density of vertical evacuation in crowds. Sci. Rep. 14 (1). https://doi.org/10.1038/s41598-024-77726-7
Zheng, X., Zhong, T., Liu, M., 2009. Modeling crowd evacuation of a building based on seven methodological approaches. Build. Environ. 44 (3), 437–445. https://doi.org/10.1016/j.buildenv.2008.04.002