

Benchmarking Pedestrian Dynamics Models for Common Scenarios: An Evaluation of Force-Based Models

Kanika Jain¹ · Shankar Prawesh² · Indranil Saha Dalal¹ · Anurag Tripathi¹

¹ Department of Chemical Engineering, Indian Institute of Technology Kanpur, Kanpur, India,
next line of long address,
E-mail: indrasd@iitk.ac.in, anuragt@iitk.ac.in

² Department of Management Sciences, Indian Institute of Technology Kanpur, Kanpur, India,
E-mail: sprawesh@iitk.ac.in

Received: 23 January 2025 / Last revision received: 9 October 2025 / Accepted: 27 November 2025
DOI: [10.17815/CD.2025.191](https://doi.org/10.17815/CD.2025.191)

Abstract Extensive research in pedestrian dynamics has primarily focused on crowded conditions and associated phenomena, such as lane formation, evacuation, etc. Several force-based models have been developed to predict the behavior in these situations. In contrast, there is a notable gap in terms of investigations of the moderate-to-low density situations. These scenarios are extremely commonplace across the world, including the highly populated nations like India. Additionally, the details of force-based models are expected to show significant effects at these densities, whereas the crowded, nearly packed, conditions may be expected to be governed largely by contact forces. In this study, we address this gap and comprehensively evaluate the performance of different force-based models in some common scenarios. Towards this, we perform controlled experiments in four situations: avoiding a stationary obstacle, position-swapping by walking toward each other, overtaking to reach a common goal, and navigating through a maze of obstacles. The performance evaluation consists of two stages and six evaluating metrics - successful trajectories, overlapping proportion, oscillation strength, path smoothness, speed deviation, and travel time. Firstly, models must meet an eligibility criterion of at least 80% successful trajectories and secondly, the models are scored based on the cutoff values established from the experimental data. We evaluated five force-based models where the best one scored 57.14%. Thus, our findings reveal significant shortcomings in the ability of these models to yield accurate predictions of pedestrian dynamics in these common situations.

Keywords Scoring system · simulation evaluation · low to moderate density scenarios · pedestrian models · controlled experiments

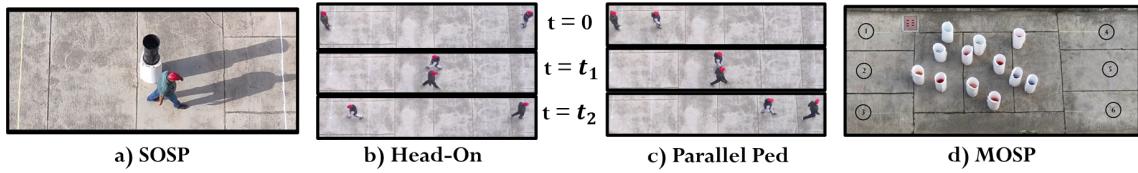


Figure 1 Controlled experiments illustrating four real-life scenarios. (a) Single Obstacle Single Pedestrian (SOSP): a volunteer passing through a non-living, human-sized stationary obstacle to reach their goal. (b) Head-On: two volunteers approaching from opposite directions, swapping their initial positions. (c) Parallel Ped: a faster volunteer overtaking a slower one while walking in the same direction aiming to a common goal. Snapshots at three different time intervals display the initial condition ($t = 0$), midway condition ($t = t_1$), and final condition ($t = t_2$) for head-on and parallel ped scenarios. d) Multiple Obstacles Single Pedestrian (MOSP): a volunteer navigating through a randomly placed maze of obstacles to reach their goal.

1 Introduction

The dynamics of pedestrians has been studied by various models proposed over past few decades by different researchers. These include the perspectives of the motion of individual pedestrians in any given scenario as well as the evolution of the state of groups of pedestrians. The types of models also vary widely and range from force-based, using classical laws of motion [1–3], to heuristic, rule-based approaches [4]. These studies have been accompanied by appropriate experimental ones as well, across various nations and situations [5–8]. The central problem, one of being able to accurately model pedestrian motion or the state of a group of pedestrians (or a crowd, which usually refers to a large number of pedestrians at high density), is of enormous significance. This is expected to enable the fundamental physics-based design of public infrastructure, where the aspects of safety and mitigation of stampede-like disasters. Such successful models will have a positive impact on public space designs across the world, more significantly in countries like India, with high population densities.

Table 1 Details of experiments (mimicking real-life scenarios) performed with the number of runs for each case.

Experiment	# runs
SOSP	58
Head-On	25
Parallel-Ped	27
MOSP	
Case A: 3.74%	239
Case B: 6.54%	188
Case C: 11.22%	184
Case D: 14.96%	276

As discussed, various models have been proposed by earlier researchers, showing great promise to mimic different pedestrian scenarios. These models must be systematically evaluated to determine which one is most appropriate for a given application, based on how well it captures the relevant properties of the scenario. We anticipate that the evaluation of the effectiveness of these models will involve various situations - ranging from those involving the motion of many pedestrians to a few, with or without other obstacles (stationary, may or may not be involving other human beings) around them. The former, or crowded scenarios, include lane formation [9–11], bottleneck [10, 12], evacuation [13–17], and crowd intersection [18–20], extensively documented in the literature. Evaluation metrics for these scenarios include fundamental diagrams [21–23], density measure [24], velocity fields [25], etc. There are empirical studies on different crowded scenarios in the literature [26, 27]. These studies provide benchmarks for assessing model performance in crowded conditions. However, note that all these situations involve a high density crowd, at closely packed conditions. Under such conditions, the close contact interactions are expected to govern pedestrian motion rather than the fundamental interactions between them, which might be active when they have some gaps around each other to move and manoeuvre. Thus, almost all reasonable force-based models (and maybe even rule based ones) may be expected to converge to similar answers for such highly packed scenarios. Interesting differences may be anticipated in moderate-to-low density situations, where the details of the interaction between pedestrians at some distance (i.e. not in close contact) are likely to influence the motion of individual pedestrians. Interestingly, in sharp contrast to crowded scenarios, much fewer studies focus on moderate-to-low density scenarios. Typically, such studies focus on the measurement of specific features like personal space [28, 29], gait features [30–32], etc., rather than the comparison of motion of pedestrians in situations from experiments and models. Thus, since such situations are not clearly established, the aspect of model evaluation with varying details of forces (and rules) remains incomplete. Hence, there exists a critical gap in terms of establishing benchmarks for evaluating pedestrian dynamics models, which we would address in this article. Note, a gap also exists for such comparisons involving Indian pedestrians, as most of the earlier investigations were performed in western or eastern nations. In this context, we note that such moderate-to-low density conditions are extremely common in various public spaces at all times, across India and other nations with large population densities. Our investigations here also extends to high density situations involving one moving pedestrian among many stationary obstacles (may or may not be humans). In this context, in a previous study, we have shown that the behavior of a single pedestrian around a non-living obstacle is similar to that around a similar-sized human being [33]. Ideally, we expect all the models to perform reasonably well at both high, moderate and low densities, but our study here proves the contrary (as shown later in the article).

From the perspective discussed above, this study addresses this gap for various force-based models in the literature. We focus on common situations at the moderate-to-low density regimes, where the details of interactions will govern the pedestrian motion. In the process, we have two broad objectives. Firstly, the benchmarking criteria for such situations are established, such that any future models can be evaluated similarly. Secondly, various force-based models present in the literature are evaluated within this criteria. To-

wards this, our study performs controlled experiments across four commonplace real-life scenarios (Figure 1), and respective data collection is given in Table 1. As explained later, these situations include one or two pedestrians, in the absence or presence of other stationary obstacles. These obstacles may represent other human beings (stationary) or inanimate objects (like pillars or similar objects). The number of obstacles is varied to traverse various regimes of density. Bottinelli and Silverburg observe instances where high-density crowds behave similarly to soft solids [34]. Here, the term “density” is used in the broader sense of the total number of objects within the experimental area, including the pedestrians. For evaluation across models, we use six evaluating metrics - successful trajectories, overlapping proportion, oscillation strength, path smoothness, speed deviation, and travel time, as explained in the article. There are many force-based models present in the literature [1–3, 35–37]. We evaluate five that are frequently used for applications - Universal Power Law (UPL) [1], Social Force Model - circular (SFMc) [35], Social Force Model - elliptical (SFMe) [2], Centrifugal Force Model - circular (CFMc) [36], and Centrifugal Force Model - elliptical (CFMe) [3]. These models were tested against the selected situations and evaluated using a scoring system based on a two-stage evaluation process, as explained later. In the first stage, models are assessed against specific eligibility criteria to determine their suitability. In the second stage, a finer comparison of the models is performed based on cutoff values established from experimental data. As highlighted later, our findings reveal significant deficiencies in the ability of force-based models to predict experimental pedestrian behavior. To regenerate the experimental scenarios and calculate evaluating metrics for scoring, initial position, final position, and desired speed for each volunteer and maximum travel time in each scenario are provided <https://github.com/kanika201293/Pedestrian-Experimental-Data>.

2 Methodology

2.1 Experimental Setup

In this study, we conducted controlled experiments for four different low to moderate density real-life-based scenarios. Note, all these scenarios involve 1-2 pedestrians, but some also have a high density of stationary obstacles. However, all of these represent commonly occurring everyday situations. Note, from the point of view of the moving pedestrian, an obstacle is equivalent to a stationary similar-sized human, as shown in our earlier study [33]. We selected obstacles that have similar diameter as a human being. Table 1 provides details of the experiments and data collection procedure. The experimental boundary was defined as a rectangular area with $x = 0\text{ m}$ to $x = 10\text{ m}$ and $y = -1.75\text{ m}$ to $y = 1.75\text{ m}$, with the measurement region spanning from $x = 2\text{ m}$ to $x = 8\text{ m}$. The experiments conducted are as follows:

1. Single Obstacle Single Pedestrian (SOSP): In this scenario, a single volunteer must reach a specified goal by avoiding a non-living, human-sized stationary obstacle placed in the middle of the path. Figure 2(a) shows the schematic of the setup.

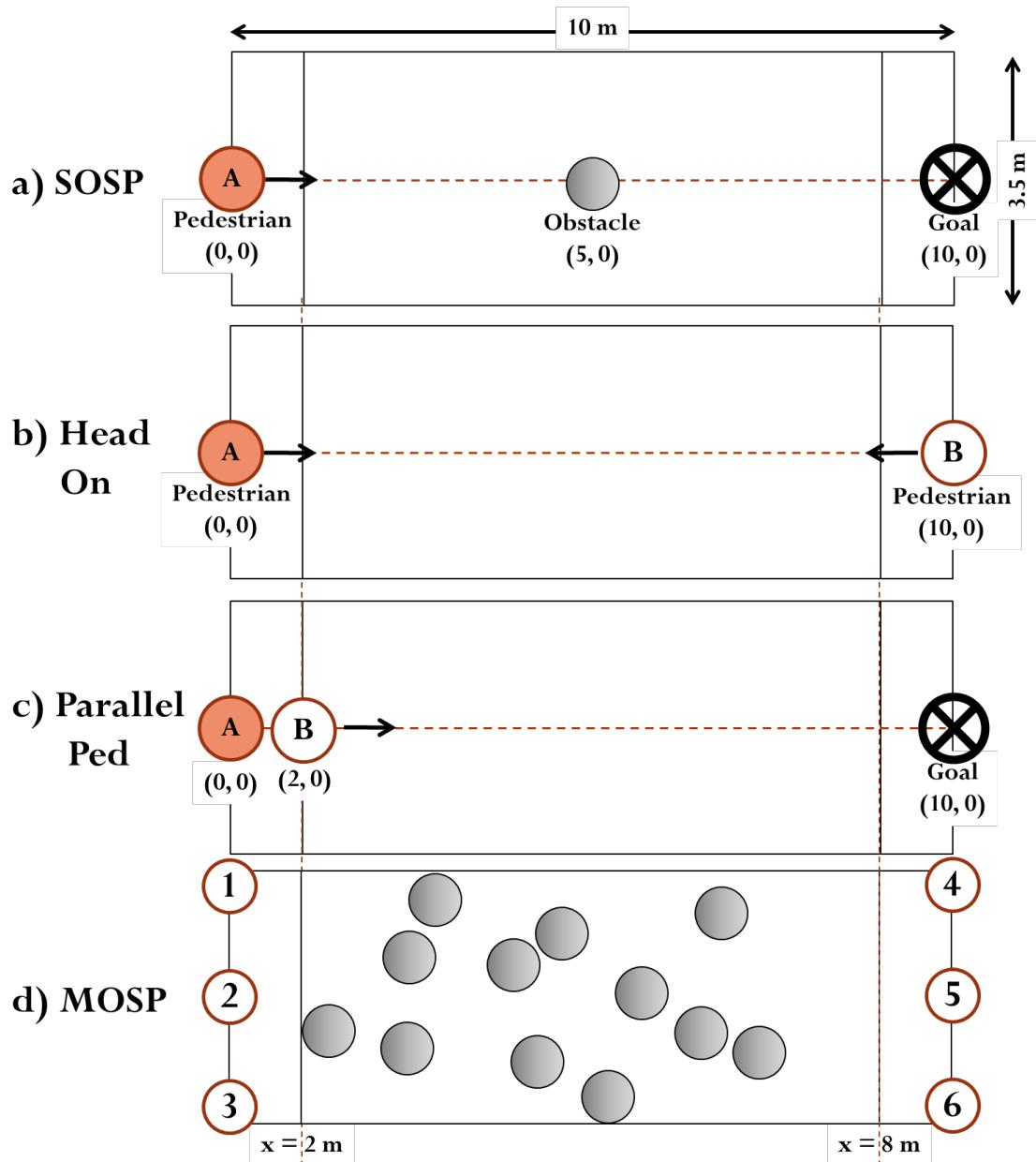


Figure 2 Schematic of all the experimental setups within a boundary defined as $x = 0\text{ m}$ to $x = 10\text{ m}$ and $y = -1.75\text{ m}$ to $y = 1.75\text{ m}$. The measurement region spans from $x = 2\text{ m}$ to $x = 8\text{ m}$. For future reference, volunteer/simulated pedestrian A is represented by a filled circle, and an empty circle represents volunteer/simulated pedestrian B.

2. Head-On: In the head-on scenario, two volunteers, denoted as A and B, walk towards each other at their desired speeds with the objective of exchanging their initial positions to reach their respective goals. A schematic illustrating this scenario is shown in Figure 2(b). Thus, the initial position of A is the goal of B and vice-versa.

3. Parallel-Ped: In the parallel-ped scenario, two volunteers, denoted as A and B, share a common destination, as illustrated in Figure 2(c). Volunteer A is initially positioned at a separation of 2 meters, behind volunteer B and has a tendency to walk at a faster pace than volunteer B. Consequently, it is expected that volunteer A will surpass B and both of them will walk towards the common goal.

4. Multiple Obstacles Single Pedestrian (MOSP): In this case, a single pedestrian passes through a maze of randomly placed obstacles to reach the goal. A series of experiments are performed by varying the area fraction occupied by the obstacles. Figure 2(d) gives a schematic of randomly placed obstacles covering 11.22% of the area. The experiment is repeated for four different area fractions: 3.74%, 6.54%, 11.22%, and 14.96%. The initial and final positions of a volunteer were selected randomly between position numbers 1 to 6, marked by circles (see Figure 2(d)).

2.2 Evaluating metrics

The following are the metrics used in this study to evaluate the performance of any model, for any given scenario.

Successful Trajectories: A trajectory is defined as *unsuccessful* when it fulfills any of the following conditions in the simulation.

- The time taken by the pedestrian to reach the goal exceeds twice the maximum travel time in the corresponding experimental setup, i.e., the pedestrian is stuck in between the path. Note that no stuck condition is observed during experiments.
- At any instant, the maximum overlap between a pedestrian and an obstacle or another pedestrian exceeds fifty percent.
- At any instant, more than fifty percent of the pedestrian's body extends outside the experimental setup boundary.

The number of successful trajectories for a particular scenario is determined by subtracting the number of unsuccessful trajectories from the total number of trajectories.

Overlapping Proportion: “Overlapping” between two pedestrians or between a pedestrian and an obstacle occurs when their geometrical form overlaps. In this study, all the pedestrians and obstacles are modeled as circles. A measure of the extent of overlap is defined earlier by Chraibi et. al. [3], as follows:

$$o^{(v)} = \frac{1}{nov} \sum_{t=0}^{t=t_{end}} \sum_{i=1}^{i=N} \sum_{j>i}^{j=N} o_{ij} \quad (1)$$

$$o_{ij} = \frac{A_{ij}}{\min(A_i, A_j)} \leq 1$$

Here, N represents the total count of simulated pedestrians, A_i and A_j denote the areas of the circular pedestrians i and j , respectively, and the variable A_{ij} represents the overlapping area between the two. In (1), the expression is normalized by the parameter nov , representing the number of simulation steps where overlapping appears. nov is 0 if there is no overlapping during the whole simulation. In order to evaluate the models thoroughly, the maximum overlap is also calculated in some cases.

Oscillation Strength: Backward movements of a pedestrian (opposite to the direction of the goal) during a simulation are called oscillations. Such movements are unrealistic in nature and need to be addressed in a model. To quantify the oscillations, we used the following measure introduced by Chraibi et al. [3]:

$$o^{(s)} = \frac{1}{nos} \sum_{t=0}^{t=t_{end}} \sum_{i=1}^{i=N} S_i \quad (2)$$

$$S_i = \frac{1}{2}(-s_i + |s_i|)$$

$$s_i = \frac{\vec{v}_i \cdot \vec{v}_i^0}{(\vec{v}_i^0)^2}$$

Here, N represents the total count of simulated pedestrians, \vec{v}_i is the current velocity, and \vec{v}_i^0 is the desired velocity of pedestrian i . In (2), the expression is normalized by the parameter nos , representing the number of simulation steps where oscillation appears; nos is set to 0 if there is no oscillation during the whole simulation. Note that the desired velocity is the desired speed of the pedestrian towards the goal.

Path Smoothness: This metric measures the smoothness of a path by identifying any sudden changes in path direction. It is calculated by measuring the maximum absolute change in the path angle which is calculated at each time instant. Let \vec{P}_t be the position vector of the pedestrian at time instant t . The displacement vector \vec{V}_t is defined as the change between position vectors \vec{P}_t and \vec{P}_{t+1} . The angular deviation of path, θ_t is the angle between the displacement vectors \vec{V}_t and \vec{V}_{t+1} . Path smoothness, θ_i , for pedestrian i , is then the maximum of the absolute value of angles θ_t . The lower value of θ_i denotes higher path smoothness for the pedestrian. The quantitative calculation of path smoothness is as follows:

$$PS = \frac{1}{N} \sum_{i=1}^{i=N} \theta_i \quad (3)$$

\vec{P}_t : position vector at t time instant

\vec{V}_t : displacement vector ($\vec{P}_{t+1} - \vec{P}_t$)

θ_t : angle between vectors ($\vec{V}_{t+1} - \vec{V}_t$)

$\theta_i : \max\{\theta_t\}$

Here, N refers to the total count of simulated pedestrians where pedestrians are denoted by i .

Speed Deviation: Speed deviation is defined as the average of the absolute change in speed relative to the desired speed at each time instant (t).

$$SD = \frac{1}{N} \sum_{i=1}^{i=N} \frac{1}{nsd} \sum_{t=0}^{t=t_{end}} \frac{|v_i^0 - v_i(t)|}{v_i^0} \quad (4)$$

Here, v_i^0 represents the desired speed, and $v_i(t)$ represents the speed at time instant t , respectively for pedestrian i . nsd refers to the number of simulation steps and N refers to the total count of simulated pedestrians.

Travel Time: This metric represents the total time taken by a pedestrian to traverse the measurement area. During simulation, a trajectory is deemed unsuccessful if the time taken by the simulator exceeds twice the maximum travel time recorded during the experiment.

Note the following for our calculations in this study:

- 1) Single values of evaluating metrics for all the scenarios are calculated (shown in Table 2), by averaging over all the trajectories.
- 2) All the evaluating metrics are calculated in the measurement region to avoid entry/exit effects.

2.3 Experimental Observations

In this section, we calculate the evaluating metrics defined in Section 2.2 for all experimental setups to establish a benchmark for assessing pedestrian dynamics models in these common scenarios. The values of these metrics are summarized in Table 2. Results show no unsuccessful trajectory in any scenario, which shows that the experimental setups, including all cases of MOSP, are such that: a) each pedestrian is able to find a path to their goal without fail, b) pedestrians are not squeezed or overlapped by more than 50% with other pedestrian or obstacles at any time, c) pedestrians remain within the defined exper-

Table 2 Evaluating metrics are calculated from the trajectories for all the experimental setups to define a benchmark to evaluate the functioning of pedestrian dynamics models for low pedestrian density scenarios. The experiments show no unsuccessful trajectories, oscillations, or overlaps. Path smoothness, measured by the change in angle between two consecutive motion vectors, ranges from 0 to 2 degrees. The speed deviation from the desired speed at any instant is not more than 23%. Additionally, the average travel time for volunteers is provided for each case.

Experimental Setup	Successful Trajectories	Avg. Overlap	Avg. Oscillation	Path Smoothness	Speed Deviation	Travel Time
SOSP	100%	0.08%	0	$0.74 \pm 0.33^\circ$	0.05 ± 0.03	$4.52 \pm 0.89s$
Head-On	100%	0%	0	$0.55 \pm 0.12^\circ$	0.02 ± 0.01	$3.97 \pm 0.29s$
Parallel Ped	100%	0%	0	$0.77 \pm 0.18^\circ$	0.05 ± 0.04	$5.21 \pm 1.4s$
MOSP						
Case A: 3.74%	100%	0%	0	$0.42 \pm 0.29^\circ$	0.05 ± 0.03	$4.16 \pm 0.55s$
Case B: 6.54%	100%	0%	0	$0.69 \pm 0.47^\circ$	0.11 ± 0.08	$4.31 \pm 0.55s$
Case C: 11.22%	100%	0.11%	1.1E-03	$1.08 \pm 0.51^\circ$	0.16 ± 0.07	$4.67 \pm 0.69s$
Case D: 14.96%	100%	0%	0	$0.93 \pm 0.36^\circ$	0.09 ± 0.05	$4.7 \pm 0.61s$

imental boundaries. Additionally, there are negligible instances of oscillation or overlap observed in any setup during the experiments. Path smoothness was measured by the change in angle between two consecutive motion vectors, and it remained within a low range of 0 to 2 degrees across all setups. This indicates that there were no sudden changes in the direction of motion, even when multiple obstacles were present. The maximum deviation from the desired speed observed among all scenarios was within the range of 24% of the value, demonstrating that pedestrians tend to maintain a consistent speed regardless of obstructions. Furthermore, the average travel time for volunteers was recorded for each case, providing additional insights into pedestrian dynamics. These evaluation metrics serve as a reliable benchmark for testing the effectiveness of pedestrian dynamics models in these situations.

3 Simulation results from force-based models

In this section, the experimental scenarios are reproduced using five force-based models, maintaining the exact initial and final positions observed during the experiments. Data for the same in each scenario are provided at <https://github.com/kanika201293/Pedestrian-Experimental-Data>. The values of the evaluating metrics are calculated from the simulations to assess the models' ability to predict the experimental outcomes. Additionally, a metric for maximum overlap is introduced to provide a more comprehensive evaluation of the models.

In our study, we considered five force-based models to replicate the aforementioned scenarios for the same initial, final positions and desired speeds, as provided by the volunteers. Force-based models represent pedestrians as inertial particles that experience

forces on the basis of their surroundings, allowing for the quantification of their behavior. The existing literature identifies three primary forces for motion commonly classified as self-driving force, interaction force, and contact force. The self-driving force enables the pedestrian to reach the goal at some desired speed. A pedestrian ideally should not collide with any other pedestrian or obstacle present in the path. This collision avoidance is fulfilled by a repulsive force, which is termed here as the interaction force. In the case of high pedestrian density, sometimes a pedestrian experiences physical contact with fellow pedestrians or obstacles. The force exerted on the pedestrian due to this physical contact is known as contact force. Contact force is primarily significant in high pedestrian density or panic situations where force experienced by the pedestrian due to physical contact is not under the control of the person. The situations reproduced in this paper are common but at low-to-moderate densities, thereby avoiding the need for contact force. The five force-based models are as follows:

Universal Power Law (UPL): The UPL is developed by Karamouzas et al. (2014) [1], which is formulated after an empirical analysis of two major crowd datasets. The fundamental concept is that the strength of the interaction force is determined only by a single factor, namely, the time-to-collision, denoted as τ . This term inherently influences the angle-dependent nature of interaction force. Note, interaction with walls utilizes the same formula provided that the pedestrian velocity is directed towards the nearest point to the wall, where the nearest point is treated as a static obstacle with a radius equal to 0 m.

Social Force Model - circular (SFMc): Helbing is one of the earliest researchers to develop a model to study pedestrian dynamics, i.e., Social Force Model [2]. SFM has been improved occasionally to bring the simulation results closer to reality. SFMc is a straightforward version of the social force model where the interaction force is only a function of surface-to-surface distance between the pedestrian and the entity. The interaction region of the pedestrian in this case is in the form of a circle [35].

Social Force Model - elliptical (SFMe): One of the earliest versions of social force models used the interaction region of a pedestrian in the form of an ellipse, which is oriented towards the direction of motion [2]. In this case, the force is dependent on the distance as well as the velocity of the pedestrian.

The derived formulae and optimized values of parameters for SFMc and SFMe are taken from the article published by Johansson et al. (2007) [38].

Centrifugal Force Model - circular (CFMc): The idea of the centrifugal force model was first proposed by Yu et al. (2005) [39] that has a resemblance to the concept of “centrifugal force”. Subsequently, Chraibi expanded upon the model, see [3, 36]. Similar to SFM, CFM also has a circular and elliptical interaction region distinguished by different definitions of the distance between the pedestrian and the obstacle. For CFMc, the radii of circular potential regions of pedestrians are dependent on the respective velocities. Note that in the case of a stationary pedestrian, the radius of the interaction region takes the minimum value of the actual radius of the entity.

Table 3 Comparison of the five force-based models with experimental data using different evaluating metrics for SOSP

	Successful Trajectories	Avg. Oscillation	Avg. Overlap	Max. Overlap	Path Smoothness	Speed Deviation	Travel Time
Experiment	100%	0	0.08%	0.2%	$0.74 \pm 0.33^\circ$	0.05 ± 0.03	4.52 ± 0.89s
UPL	100%	0	0%	0%	$3.33 \pm 0.88^\circ$	0.06 ± 0.04	4.53 ± 0.78s
SFMc	17.86%	0	35.87%	100%	$0.05 \pm 0.01^\circ$	0.02 ± 0	4.27 ± 0.75s
SFMe	98.21%	0	2.27%	52.75%	$27.33 \pm 64.12^\circ$	0.3 ± 0.08	6.27 ± 1.68s
CFMc	100%	0	1.29%	33.1%	$0.49 \pm 0.13^\circ$	0.31 ± 0.07	6.27 ± 1.13s
CFMe	100%	0	0%	0%	$0.45 \pm 0.2^\circ$	0.34 ± 0.08	6.53 ± 1.09s

Centrifugal Force Model - elliptical (CFMe): In this model, the potential region of a pedestrian is elliptical in nature where minor and major axes are dependent on the motion of the pedestrian.

Note that, for all the models, the simulation time step size Δt is set as 0.01 s, the maximum time step size that avoids numerical instability (refer to Supplementary material). This study assigns specific colors for each force-based model and the experimental data. The color coding remains consistent throughout the article.

Simulation Results

SOSP: As mentioned earlier, in this case, a pedestrian aims to reach a goal by avoiding a stationary obstacle placed in the middle of the path. Figure 2(a) illustrates the initial position of pedestrian A (taken as (0, 0)). In real-life situations, it is unlikely that a person stands exactly at the expected position without any offset. During the experiment, we observed that 58 participants had an average offset of 0.09 m with a standard deviation of 0.07 m from their indicated initial position (0, 0). The participants' average desired speed (speed without any obstacle) was 1.47 m/s with a standard deviation of 0.16 m/s.

Assuming a pedestrian starts from the exact initial position so that the pedestrian, the obstacle, and the goal are collinear, as shown in the schematic diagram, simulations using the UPL model predicts the pedestrian getting stuck after reaching near the obstacle positioned in the middle (refer to Figure 3(a)). The collinearity of the pedestrian, obstacle, and goal results in attractive and repulsive forces, also collinear but opposite in direction. As the pedestrian approaches the obstacle, the repulsive force increases, counterbalancing the self-driving force. Due to the absence of any transverse force, the pedestrian cannot change its direction and remains stuck. The same problem of collinearity is also observed in the case of all other force-based models considered in this study.

Figure 3(a) shows that the trajectories for all models overlap, indicating no transverse displacement due to collinear forces. The pedestrian gets stuck just before the obstacle

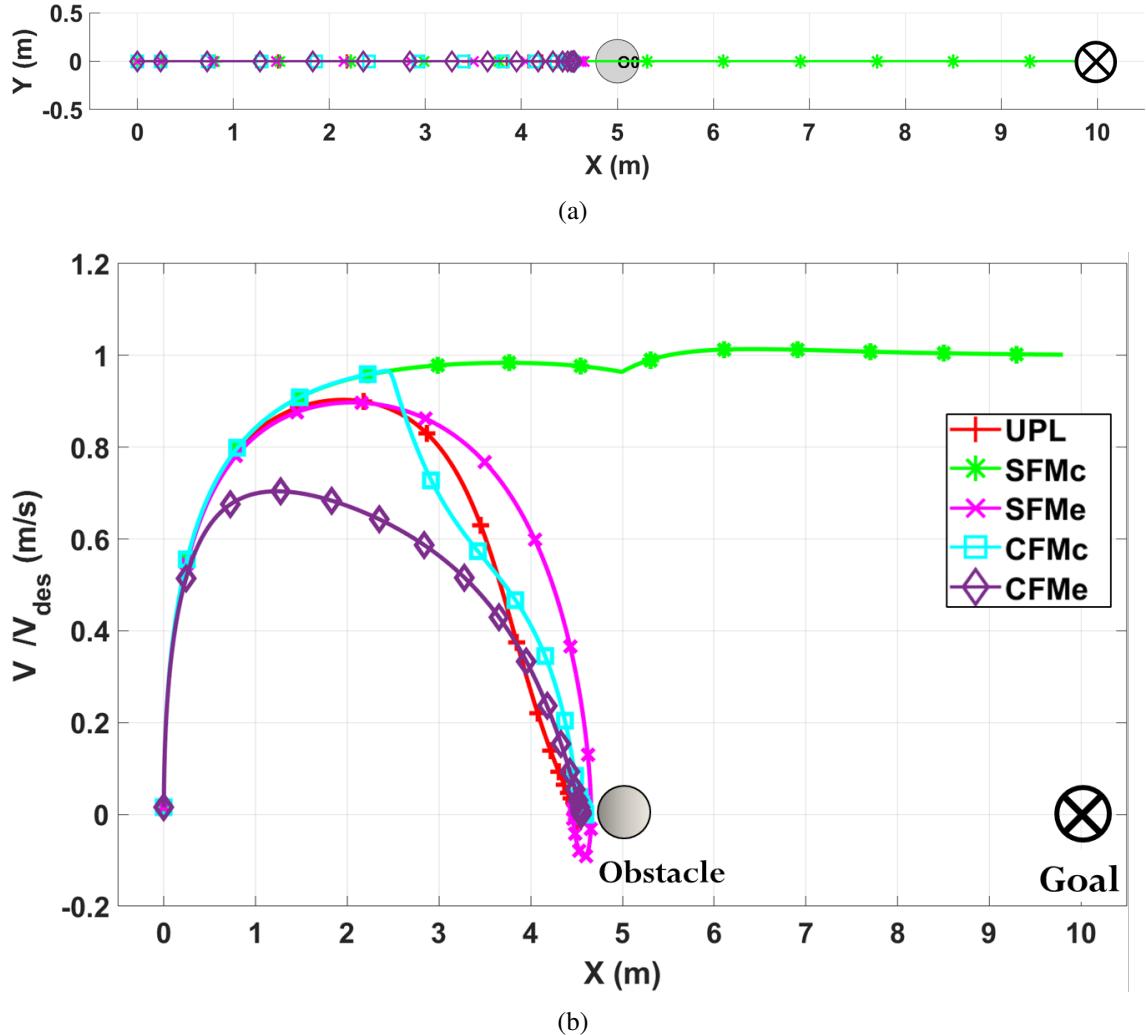
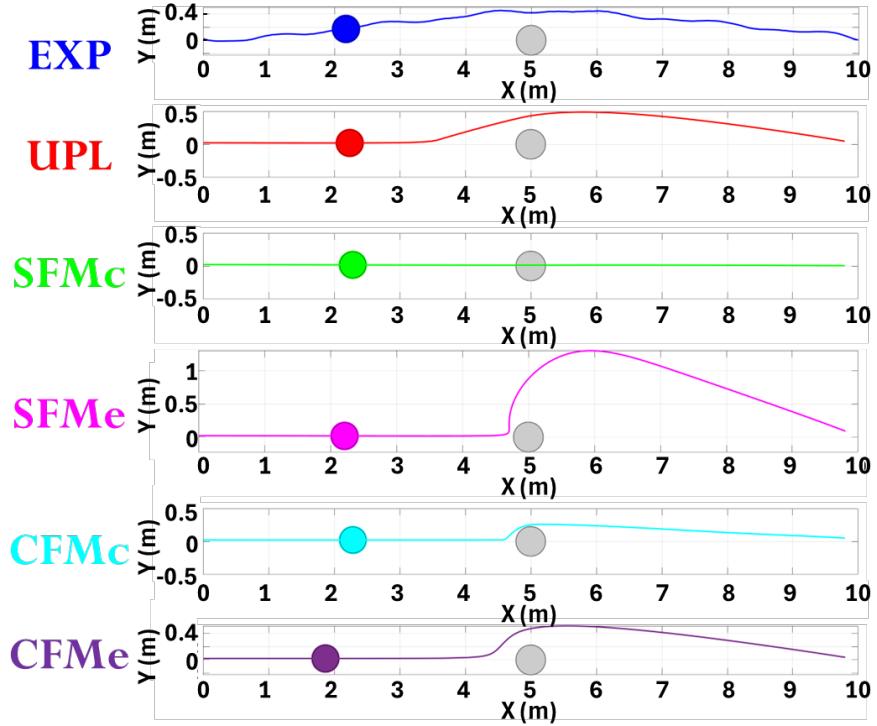
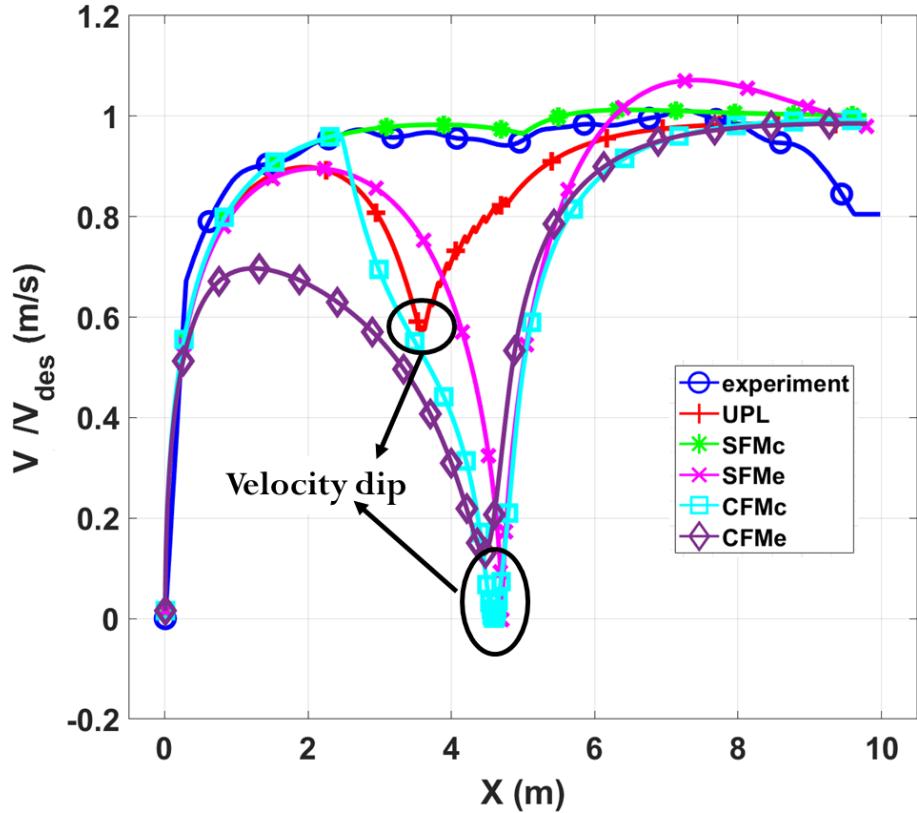


Figure 3 (a) Trajectories for all models show overlap, showing no transverse displacement due to collinear forces. Pedestrians get stuck before the obstacle in all models except for SFMc, where pedestrian can pass through the obstacle due to low repulsion. (b) Normalized speed plotted against the x-component of the pedestrian's position. Speeds for all models approach 0 m/s, indicating pedestrians get stuck, except for SFMc. SFMe exhibits oscillation before getting stuck, indicated by negative velocity.



(a) Trajectory graphs (Multimedia available online)



(b) Normalized speed graph

Figure 4 SOSP: (a) Trajectories and (b) Normalized speed for experimental and simulated trajectories. The experimental data shows anticipatory direction changes with minimal speed change to avoid the obstacle. In contrast, simulated models, except SFMc, exhibit sudden direction changes near the obstacle, resulting in velocity dips. Overlaps with the obstacle are noted for SFMc, SFMe, and CFMc, with no oscillation or backward motion ($V < 0$ m/s) observed in any model.

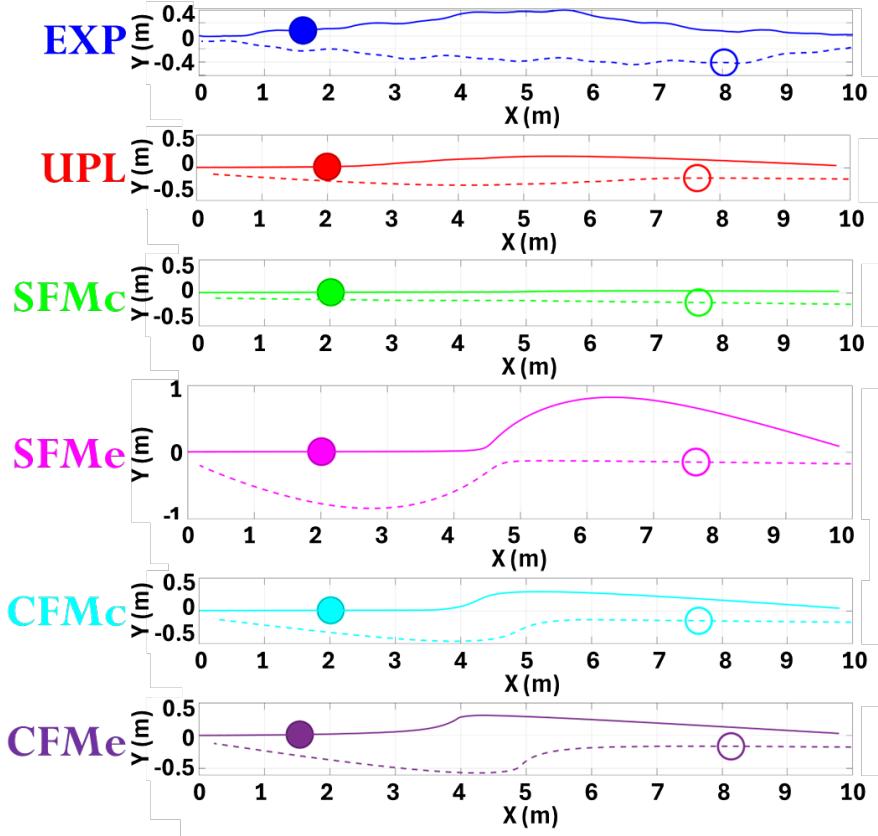
in all models except SFMc, where overlap with the obstacle is observed, allowing the pedestrian to reach the goal without deviation. In SFMc, the collinearity problem results in no transverse movement, and the low repulsive force of the model leads to overlap. The normalized speed is shown against the x-component of the pedestrian's position in Figure 3(b). We observe that the speeds for all models approach 0, indicating that pedestrians get stuck, except for SFMc. As discussed earlier, due to the low repulsive force magnitude, the pedestrian continues with overlap with no decrement in speed.

However, in real-life scenarios, it is unlikely that a person stands precisely in the expected position with no offset. Trajectories and normalized speed profiles for one of the data are plotted in Figure 4 (multimedia available online). In this case, the volunteer in the experiment started with an offset as low as 11 mm, yet there were no sudden changes in the trajectory or speed during the experiment, unlike simulation results. As explained next, all models showed significant slowdown of the pedestrian as it approached the obstacle in the simulations, which is in sharp contrast to the experimental observations. The possible reason behind such behavior is force-based models relying on forces to navigate around the obstacles, which naturally leads to a change in the magnitude of the velocity (speed), as can be seen in Figure 4(b).

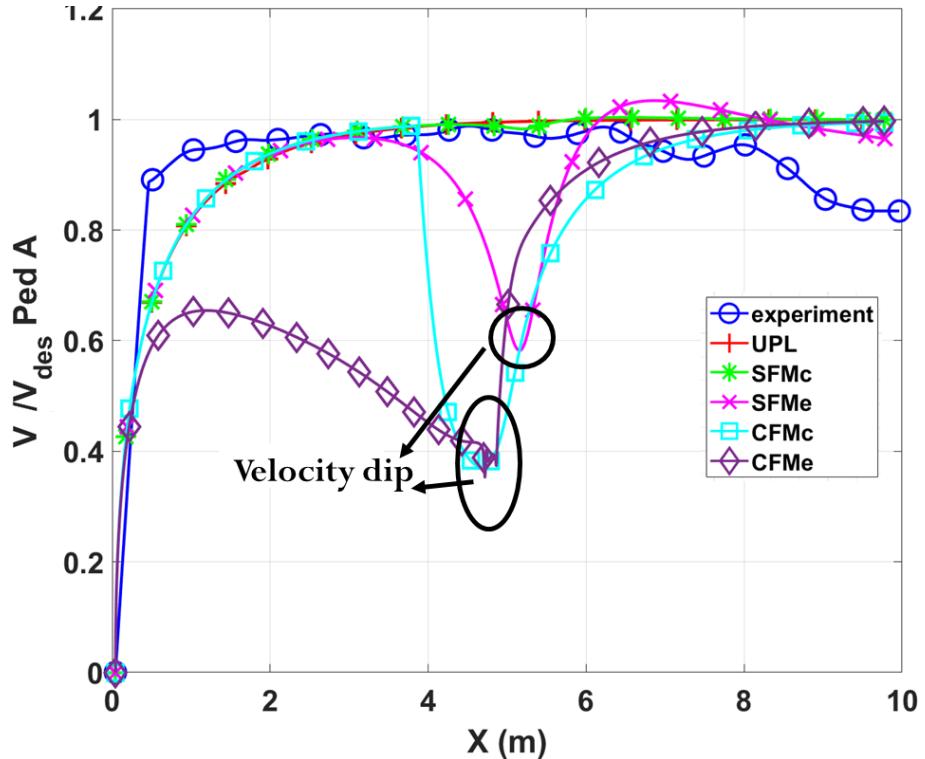
While simulating the experimental scenarios using force-based models using the same initial and final positions, we calculated and reported the corresponding evaluating metrics in Table 3. Model SFMc predicted over 80% unsuccessful trajectories, SFMe showed less than 2% unsuccessful trajectories, while all other models had all successful trajectories. The large number of unsuccessful trajectories for model SFMc is primarily due to high overlap, with a maximum overlap of 100%. This high failure rate stems from SFMc's weak interaction force, leading to an inability to avoid obstacles effectively, as depicted in Figure 4(a) (multimedia available online). Some overlap was observed in models SFMe and CFMc, but no overlap occurred with UPL and CFMe. No backward motion or oscillation was reported for any model. However, sudden changes in direction were noted under the 'path smoothness' metric, with model SFMe showing a maximum change exceeding 90 degrees, as illustrated in the trajectory. It is important to note that human walking is inherently less smooth than simulated motion, which can result in lower path smoothness values in simulations compared to experimental values in some cases. Large speed deviations from the desired speed were observed in models SFMe, CFMc, and CFMe, evident in the velocity dips highlighted in Figure 4(b). These speed slowdowns significantly increased travel time for these three models, and respective values are provided in Table 3.

An evaluation of force-based models for SOSP revealed that model SFMc had less than 20% successful trajectories due to weak interaction forces resulting in high overlap. In contrast, models SFMe had only a few unsuccessful trajectories, while UPL, CFMc, and CFMe had none. Additionally, models UPL and SFMe exhibited sudden changes in path direction, and models SFMe, CFMc, and CFMe experienced significant speed deviations, which led to increased travel times. Note, nearly all models failed when the pedestrian, obstacle, and goal are collinear with zero offset, as discussed earlier.

Head-On: One example from the head-on experiments is plotted in Figure 5 (multimedia available online). In trajectory plots, solid lines with filled circles represent pedestrian



(a) Trajectory graphs (Multimedia available online)



(b) Normalized speed (Ped A) graph

Figure 5 Head-On: (a) Trajectories and (b) Normalized speed (Ped A) graphs are presented for both experimental and simulated data. The experimental data indicates an anticipatory change in direction to avoid collisions, with minimal speed fluctuations. In the trajectory plots, solid lines with filled circles represent pedestrian A, while dashed lines with empty circles represent pedestrian B. The velocity profiles show a dip when avoiding collisions in all models except SFMc.

Table 4 Comparison of the five force-based models with experimental data using different evaluating metrics for Head-On.

	Successful Trajectories	Avg. Oscillation	Avg. Overlap	Max. Overlap	Path Smoothness	Speed Deviation	Travel Time
Experiment	100%	0	0%	0%	$0.55 \pm 0.12^\circ$	0.02 ± 0.01	3.97 ± 0.29s
UPL	100%	0	0%	0%	$2.81 \pm 0.44^\circ$	0.03 ± 0.03	3.99 ± 0.31s
SFMc	52.38%	0	24.68%	97.8%	$0.05 \pm 0.01^\circ$	0.01 ± 0.0	3.89 ± 0.26s
SFMe	76.19%	0	11.4%	81.66%	$10.03 \pm 38.29^\circ$	0.11 ± 0.03	4.41 ± 0.45s
CFMc	100%	0	0.02%	0.64%	$0.59 \pm 0.23^\circ$	0.28 ± 0.07	5.45 ± 0.69s
CFMe	80.95%	0	9.55%	74.2%	$8.62 \pm 4.73^\circ$	0.38 ± 0.06	6.08 ± 0.44s

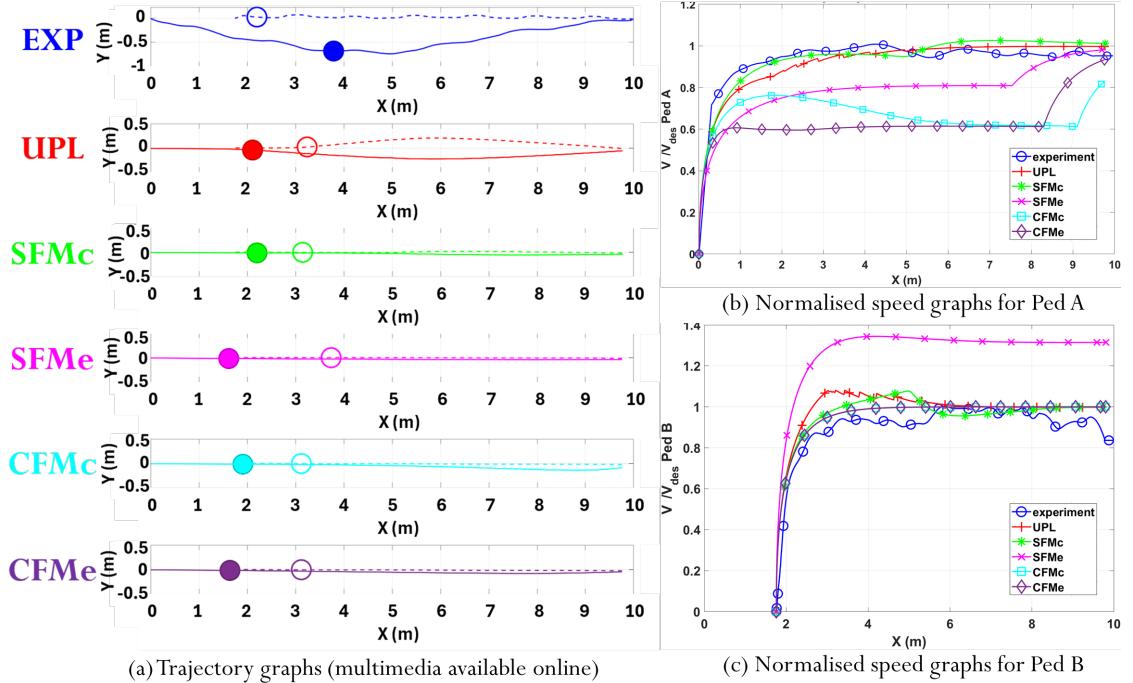


Figure 6 ParallelPed: (a) Trajectories (Multimedia available online) and (b) normalized speed graphs are presented for both experimental and simulated data. In the experiment, pedestrian A (faster, solid line with filled circles) avoids pedestrian B (slower, dashed line with empty circles) by changing direction, while B continues without interruption. Both show minimal speed changes. In simulations, A is unable to overtake B, resulting in straight-line trajectories in all models except UPL. Velocity profiles indicate A fails to reach its desired speed in SFMe, CFMc, and CFMe models, while B maintains consistent speed, with B experiencing a push effect in SFMe.

Table 5 Comparison of the five force-based models with experimental data using different evaluating metrics for Parallel-Ped.

	Successful Trajectories	Avg. Oscillation	Avg. Overlap	Max. Overlap	Path Smoothness	Speed Deviation	Travel Time
Experiment	100%	0	0%	0%	$0.77 \pm 0.18^\circ$	0.05 ± 0.04	$5.21 \pm 1.4s$
UPL	100%	0	0%	0%	$3.03 \pm 1.08^\circ$	0.02 ± 0.01	$5.0 \pm 1.2s$
SFMc	66.67%	0	15.61%	95.25%	$0.04 \pm 0.02^\circ$	0.04 ± 0.01	$5.01 \pm 1.2s$
SFMe	100%	0	0%	0%	$0.03 \pm 0.03^\circ$	0.22 ± 0.05	$4.71 \pm 0.26s$
CFMc	100%	0	0%	0%	$0.05 \pm 0.05^\circ$	0.19 ± 0.1	$5.56 \pm 0.83s$
CFMe	100%	0	0%	0%	$0.04 \pm 0.05^\circ$	0.18 ± 0.15	5.93 ± 0.57

A, and dashed lines with empty circles denote pedestrian B. Table 4 reports the evaluating metrics for head-on scenarios simulated using force-based models. Unsuccessful cases are predicted for models SFMc, SFMe, and CFMe, with high overlapping proportions. For SFMc and SFMe, the overlap is also evident from the trajectories shown in Figure 5(a) (multimedia available online). Models UPL and CFMc report 0% and less than 1% maximum overlap, respectively, and no oscillation is observed in any model. Large sudden deviations in the direction of motion are observed in models SFMe and CFMe, also shown in the trajectory. The metric of path smoothness measures the sudden change in the direction of motion rather than the overall change in the path. Hence, the value of path smoothness is lower for CFMc relative to CFMe. Although the overall deviation is large for CFMc, it is smoother compared to the abrupt changes observed in CFMe's trajectory at around $x = 4 m$ in Figure 5(a) (multimedia available online). Significant speed deviations are reported for models CFMc and CFMe, as can also be seen in the velocity profiles of Figure 5(b). Consequently, these models also exhibit longer travel times, as indicated in Table 4. For other models, the simulation predicts travel times similar to the experiments.

An evaluation of force-based models for the head-on simulations reveal that models SFMc and SFMe predict less than 80% successful trajectories, characterized by high overlap. Model CFMe has a few unsuccessful trajectories, while models UPL and CFMc had none. Furthermore, sudden changes in path direction were observed in models UPL, CFMe, and SFMe, and significant speed deviations were noted in models SFMe, CFMc, and CFMe, contributing to increased travel times. Note, similar to the previous case, models also fail for collinear pedestrians in Head-On simulations, in contrast to experiments.

Parallel-Ped: In this scenario, two pedestrians, A and B, share a common goal, as illustrated in Figure 2(c). One example of speed profiles and trajectories for both pedestrians generated using force-based models are provided in Figure 6 (multimedia available online). In trajectory plots, solid lines with filled circles represent pedestrian A, while dashed lines with empty circles represent pedestrian B. Evaluating metrics obtained from simulations are shown in Table 5. The model SFMc predicted more than 30% unsuccessful trajectories due to large overlap in most cases, with a maximum overlap of 95.25%.

Table 6 Comparison of the five force-based models with experimental data using different evaluating metrics for MOSP.

MOSP	Successful Trajectories	Avg. Oscillation	Avg. Overlap	Max. Overlap	Path Smoothness	Speed Deviation	Travel Time
------	-------------------------	------------------	--------------	--------------	-----------------	-----------------	-------------

Case A: 03.74%

Experiment	100%	0	0%	0%	$0.42 \pm 0.29^\circ$	0.05 ± 0.03	$4.16 \pm 0.55s$
UPL	100%	0.01	0%	0%	$2.58 \pm 1.5^\circ$	0.07 ± 0.06	$4.23 \pm 0.51s$
SFMc	61.51%	0	18.3%	100%	$0.04 \pm 0.01^\circ$	0.01 ± 0.0	$3.96 \pm 0.45s$
SFMe	83.26%	0.07	0.42%	27.4%	$29.04 \pm 64.65^\circ$	0.42 ± 0.18	$7.34 \pm 3.44s$
CFMc	100%	0	1.6%	39.5%	$0.56 \pm 0.51^\circ$	0.38 ± 0.14	$6.73 \pm 1.7s$
CFMe	92.05%	0	6.19%	100%	$1.27 \pm 10.83^\circ$	0.41 ± 0.12	7.07 ± 2.03

Case B: 06.54%

Experiment	100%	0	0%	0%	$0.69 \pm 0.47^\circ$	0.11 ± 0.08	$4.31 \pm 0.55s$
UPL	100%	0.02	0%	0%	$5.46 \pm 5.07^\circ$	0.18 ± 0.15	$4.9 \pm 1.41s$
SFMc	26.06%	0	27.35%	100%	$0.05 \pm 0.01^\circ$	0.03 ± 0.0	$3.86 \pm 0.52s$
SFMe	31.38%	0.07	7.31%	68.2%	$108.31 \pm 86.9^\circ$	0.75 ± 0.24	$13.74 \pm 6.99s$
CFMc	87.77%	0	2.29%	50.73%	$12.88 \pm 37.51^\circ$	0.57 ± 0.14	$9.71 \pm 4.08s$
CFMe	96.81%	0	1.27%	97.23%	$3.3 \pm 17.37^\circ$	0.58 ± 0.07	7.07 ± 2.03

Case C: 11.22%

Experiment	100%	1.1E-03	0.11%	0.18%	$1.08 \pm 0.51^\circ$	0.16 ± 0.07	$4.67 \pm 0.69s$
UPL	60.33%	0.01	0%	8.52%	$82.54 \pm 77.23^\circ$	0.58 ± 0.3	$12.15 \pm 6.8s$
SFMc	3.8%	0	31.53%	100%	$0.06 \pm 0.01^\circ$	0.03 ± 0.0	$3.92 \pm 0.52s$
SFMe	0%	0.01	1%	77.45%	NA	NA	NA
CFMc	84.24%	0	6.23%	100%	$7.56 \pm 27.4^\circ$	0.7 ± 0.08	$13.46 \pm 2.86s$
CFMe	11.95%	0	26.44%	100%	$73.58 \pm 65.29^\circ$	0.74 ± 0.15	14.89 ± 3.79

Case D: 14.96%

Experiment	100%	0	0%	0%	$0.93 \pm 0.36^\circ$	0.09 ± 0.05	$4.7 \pm 0.61s$
UPL	56.52%	0.01	0%	0.99%	$90.3 \pm 79.66^\circ$	0.62 ± 0.29	$13.74 \pm 7.35s$
SFMc	21.01%	0	28.92%	100%	$0.07 \pm 0.01^\circ$	0.04 ± 0.0	$4.3 \pm 0.44s$
SFMe	0%	0	0%	0%	NA	NA	NA
CFMc	72.46%	0	13.57%	100%	$13.06 \pm 37.16^\circ$	0.69 ± 0.15	$14.94 \pm 4.13s$
CFMe	6.88%	0	33.76%	100%	$47.83 \pm 62.57^\circ$	0.71 ± 0.17	14.18 ± 4.31

For scenarios SOSP, head-on, and parallel ped, involving only two entities, the main cause of unsuccessful trajectories is overlap rather than the pedestrian getting stuck. Oscillations are not predicted in any model for this scenario. The trajectories of pedestrian A in Figure 6(a) (multimedia available online) indicate that, for most models, the faster pedestrian A is unable to overtake the slower pedestrian B, resulting in mostly straight-line trajectories. However, in experiments, pedestrian A overtakes B, in contrast to these simulated predictions. Thus, path smoothness for most models is better than the experimental value, except for UPL. In the UPL model, pedestrian A successfully overtakes pedestrian B. However, pedestrian B also changes the path direction to allow passage for pedestrian A, unlike during the experiment where A finds its own path without affecting B's motion. This is logical as pedestrian B, positioned ahead of A, cannot see A and should not alter its path. For UPL, sudden direction change for both pedestrians ranges from 2 to 4 degrees. Models SFMe, CFMc, and CFMe exhibit significant speed deviations from the desired speed, with maximum deviations between 20% and 48%. Large deviation in velocities from the desired speed (V_{des}) is also observed in the normalized speed plots given in Figure 6(b)-(c). For these models, pedestrian A cannot overtake B, resulting in a slower speed of A (Figure 6(b)), while the speed of B is maintained for models CFMc and CFMe (Figure 6(c)). The SFMe model shows an unrealistic push effect on pedestrian B, increasing the speed by up to 1.4 times. The models predicted travel times similar to the experiment.

An evaluation of force-based models for parallel pedestrian scenarios revealed that SFMc had fewer than 80% successful trajectories and exhibited significant overlap, whereas all other models had 100% successful trajectories. Sudden changes in path direction were noted exclusively in model UPL, while significant speed deviations were observed in models SFMe, CFMc, and CFMe, leading to increased travel times. Apart from UPL, all models predicted A being behind B and getting slowed down, in contrast to experiments. However, as noted earlier, UPL predicts a path change of B, which contradicts the experiments.

MOSP: In the MOSP scenario, a single pedestrian passes through a maze of randomly placed obstacles to reach the goal, with experiments conducted across four area fractions: 3.74% (Case A), 6.54% (Case B), 11.22% (Case C), and 14.96% (Case D). Evaluating metrics calculated by reproducing each setup for the exact initial and final conditions are reported in Table 6. These show that as the area fraction covered by the obstacles increases, the number of unsuccessful trajectories also increases. While it is commonly reported in the literature that force-based models perform well in complex scenarios such as lane formation under moderate to high densities, our findings indicate a decline in model performance as density increases. However, one difference in our situations considered here is the aspect of navigation around a crowded zone with static obstacles, unlike the dynamics in a crowd where every pedestrian is mobile.

Note, such situations of pedestrians walking around several other stationary obstacles (or other standing pedestrians) are extremely common across crowded public spaces, like railway stations, bus stations, food joints etc. In MOSP, the large number of obstacles leads to unsuccessful trajectories due to a combination of three factors mentioned in Section 2.2. For the SFMe model, no successful trajectories were observed in Cases C and

D, as high repulsive forces prevented pedestrians from entering the measurement region, resulting in ‘Not Attempted (NA)’ for these cases, (refer to Appendix Figure B-5(b)). Surprisingly, the SFMc model showed an increase of successful trajectories from Case C to D due to the obstacle configuration favorable for the model, allowing the pedestrian to reach the goal with low overlap, even when there is a minimal deviation in path due to the weak repulsive forces of the model (refer to Supplementary material Figure B-5(a)). Both average and maximum overlaps increased with area fraction. However, this metric is influenced by obstacle configuration, pedestrian position, and the number of interacting obstacles. Thus, this trend should not be seen as a rule. Insignificant oscillation or backward motion was observed across all models. As the number of obstacles increased, the number of turns required to pass through the maze also increased, leading to higher sudden changes in motion direction, reflected in large path smoothness values. Due to frequent changes in path direction, speed deviation from the desired speed also increases with the area fraction. Normalized speed graphs for the four cases, averaged over all the trajectories, can be observed in Supplementary material Figure B-4. However, the SFMc model reported low path smoothness and speed deviation, resulting in increased travel time with area fraction for all models except SFMc. However, note that the proportion of successful trajectories with SFMc is pretty low (< 25%) for both area fractions.

In the MOSP scenario, the study reveals that as obstacle density increases, the frequency of unsuccessful trajectories rises, with no model achieving an 80% success rate at the density of 14.96%. The pedestrian models exhibit widely divergent responses to these challenges: the SFMe model completely fails in higher-density conditions, whereas the SFMc model demonstrates improved success rates, along with reduced path smoothness and speed deviation, due to favorable obstacle configurations for the model. The analysis also shows that with an increase in obstacles, there is a corresponding rise in path complexity, speed deviation, and travel time. However, these trends are not universally applicable, as they depend on the specific configurations of obstacles and the positioning of pedestrians. As discussed, this is a common scenario in any public gathering/space, where one person may need to navigate around obstacles and other people to reach a different location. For such an everyday scenario, simulation models predicting unsuccessful trajectories reveal serious shortcomings.

4 Evaluation of Models

The evaluation procedure involves two stages:

- Stage I: The models must meet an eligibility criterion of at least 80% successful trajectories (Table 7).
- Stage II: The shortlisted models are then evaluated using a scoring system, where a model’s score is determined by the number of times its evaluating metrics fall within the specified cutoff values (Table 8, 9).

Table 7 Stage I: Models meeting eligibility criteria are reported in this table. For each experimental setup, models achieving more than 80% successful cases are marked as \checkmark , otherwise \times . Models with the number of \checkmark greater than or equal to the number of \times are shortlisted for further evaluation in Stage II, given in Table 8.

	UPL	SFMc	SFMe	CFMc	CFMe
SOSP	\checkmark	\times	\checkmark	\checkmark	\checkmark
Head-On	\checkmark	\times	\times	\checkmark	\checkmark
Parallel Ped	\checkmark	\times	\checkmark	\checkmark	\checkmark
MOSP:					
Case A(03.74%)	\checkmark	\times	\checkmark	\checkmark	\checkmark
Case B(06.54%)	\checkmark	\times	\times	\checkmark	\checkmark
Case C(11.22%)	\times	\times	\times	\checkmark	\times
Case D(14.96%)	\times	\times	\times	\times	\times
	Elig.	NE	NE	Elig.	Elig.

In Table 7, the first stage of the models' eligibility criteria is reported. For each experimental scenario, models achieving more than 80% successful cases are marked as \checkmark ; otherwise, they are marked as \times . Models with a number of \checkmark greater than or equal to the number of \times are shortlisted for further evaluation in Stage II; Otherwise considered as 'Not Eligible (NE)'. This eligibility criterion primarily determines if a model can provide simulation results comparable to the experimental results for the situations considered. Table 7 clearly indicates that the UPL, CFMc, and CFMe models perform well ($> 80\%$ successful trajectories) for most situations considered. Note that this study does not imply that the remaining models are entirely ineffective, but rather that they are not suitable for these common situations with their current parametric values (refer to Section C in the Supplementary material for more details). A search across the parameter space of the models is beyond the scope of this article. Interestingly, almost all models are unsuccessful for the MOSP scenarios with higher obstacle densities.

In the second stage, \checkmark and \times marks are again assigned to the shortlisted models, based on whether their average metric values fall within the specified cutoff values (refer to Table 8). Cutoff values for path smoothness, speed deviation, and travel time are determined by using the 3σ rule, adding three standard deviations to the mean values observed during the experiments, while the cutoff values for overlap and oscillation metrics are set at 10% and 0.005, respectively, due to no observed values for these metrics during experiments.

Normalized success score for a model is derived by counting the number of \checkmark it receives (refer to Table 9). The results after the two-stage process show that UPL yields the best outcomes among the five force-based models, followed by CFMc and CFMe. However, the final scores of UPL, CFMc and CFMe are separated by $< 10\%$, suggesting that they yield comparable predictions for the situations considered here. It is important to recognize the significance of the first stage of shortlisting. If scoring had been di-

Table 8 Stage II: Labels ✓ and ✗ are assigned to the shortlisted models from Stage I, based on whether their average evaluating metric values fall within the specified cutoff values or not, respectively. The cutoff values are mentioned in the first column corresponding to each metric. The values for overlap and oscillation metrics are set as 10% and 0.005, respectively, for each experimental setup. The cutoff values are determined using the 3σ rule for the other three metrics, adding three standard deviations to the mean values observed during the experiments. The final scoring for the models is provided in Table 9

Metrics →	Avg. Overlap				Avg. Oscillation				Path Smoothness				Speed Deviation				Travel Time			
Exp. Setup ↓	Cutoff	UPL	CFMc	CFMe	Cutoff	UPL	CFMc	CFMe	Cutoff	UPL	CFMc	CFMe	Cutoff	UPL	CFMc	CFMe	Cutoff	UPL	CFMc	CFMe
SOSP	< 10%	✓	✓	✓	< 0.005	✓	✓	✓	< 1.73°	✗	✓	✓	< 0.14	✓	✗	✗	< 7.19s	✓	✓	✓
Head-On	< 10%	✓	✓	✓	< 0.005	✓	✓	✓	< 0.91°	✗	✓	✗	< 0.05	✓	✗	✗	< 4.84s	✓	✗	✗
Parallel-Ped	< 10%	✓	✓	✓	< 0.005	✓	✓	✓	< 1.31°	✗	✓	✓	< 0.17	✓	✗	✗	< 9.41s	✓	✓	✓
MOSP																				
Case A	< 10%	✓	✓	✓	< 0.005	✗	✓	✓	< 1.29°	✗	✓	✓	< 0.14	✓	✗	✗	< 5.81s	✓	✗	✗
Case B	< 10%	✓	✓	✓	< 0.005	✗	✓	✓	< 2.1°	✗	✗	✗	< 0.35	✓	✗	✗	< 5.96s	✓	✗	✗
Case C	< 10%	✓	✓	✗	< 0.005	✗	✓	✓	< 2.61°	✗	✗	✗	< 0.37	✗	✗	✗	< 6.74s	✗	✗	✗
Case D	< 10%	✓	✗	✗	< 0.005	✗	✓	✓	< 2.01°	✗	✗	✗	< 0.24	✗	✗	✗	< 6.53s	✗	✗	✗

Table 9 A model’s score for final evaluation is calculated by counting the number of ✓ it receives in Table 8. The evaluation process highlights that UPL shows the best results among the five force-based models, followed by CFMc and CFMe, while SFMc and SFMe are ‘Not Eligible (NE)’ from Stage I.

Models	✓	✗	SCORE
UPL	20	15	$20/35 = 57.14\%$
CFMc	19	16	$19/35 = 54.29\%$
CFMe	17	18	$17/35 = 48.57\%$
SFMc	-	-	NE
SFMe	-	-	NE

rectly applied to all models, SFMc would have received the highest score despite failing to achieve 80% successful trajectories for all the setups, including the simplest SOSP scenario. However, our analysis also indicates the failure of all models for the higher-density MOSP scenarios.

5 Conclusions

Thus in this study, we have provided a comprehensive scoring system designed to evaluate pedestrian dynamics models for commonly occurring situations involving moderate-to-low pedestrian density (with large obstacle density for some cases). Towards this, we have selected four common situations and performed experiments with volunteers spanning age groups and gender. The experimental results serve as a baseline to develop a two-stage evaluation process including eligibility criteria and scoring. We provide a detailed evaluation of five force-based models by mimicking experimental scenarios using exact initial and final conditions. Our overall observations from the experiments are summarized below.

- SOSP: SFMc fails to produce 80% successful trajectories and shows high overlap. SFMe passes the eligibility but demonstrates abrupt path changes, significant speed deviations, and longer travel times. UPL results in sudden path changes but minimal speed deviation, while CFMc and CFMe exhibit high speed deviations but low path direction changes. Nearly all models show a significant decrease in speed as the pedestrian approaches the obstacle.
- Head-On: Both SFMc and SFMe fail to achieve 80% success rates. CFMe shows some unsuccessful trajectories, whereas UPL and CFMc have none. CFMe also shows abrupt path changes, significant speed deviations, and increased travel times. UPL resulted in sudden path change but minimal speed deviation, whereas CFMc shows high speed deviation but low change in path direction.

- Parallel-Ped: SFMc fails to reach an 80% success rate and has high overlap, while all other models show 100% success. UPL causes sudden path changes with minimal speed deviations, whereas the remaining models display high speed deviations with low path direction changes. In experiments, the faster pedestrian overtakes the slower one, without any significant change in the dynamics of the slower one. For all models except UPL, the faster pedestrian couldn't overtake the slower one. UPL predicts an overtake but with a significant path change of B which is not observed in experiments.
- MOSP: With increasing obstacle density, all models showed higher proportions of unsuccessful trajectories, greater path complexity, speed deviations, and longer travel times, although these effects depend on specific obstacle configurations and pedestrian positions. Both SFMc and SFMe fail to achieve 80% success in nearly all cases, while other models perform relatively better. UPL exhibits backward motion and abrupt path changes, whereas CFMc and CFMe have sudden path changes and high speed deviations.

Here, we note that nearly all models fail for MOSP situations with a large number of obstacles (14.96%). This is a highly common scenario in India, where one needs to navigate through moderately crowded public spaces (even simple corridors with a moderate density of standing people). While it is true that parameter tuning (e.g., adjusting k and τ_0 in the case of UPL) can influence model behavior, our additional analysis (see Figure E-8 in Supplementary material) demonstrates that even significant variations (up to tenfold changes) in these parameters do not substantially improve the model's performance in the tested scenarios. This supports the article's focus on addressing deeper structural issues in the models, rather than parameter sensitivity alone.

Next, we finalize our evaluation by a two-stage process explained in Tables 7, 8, and 9. Overall, the UPL [1] emerged as the most successful in our scoring system, which might provide decent predictions in the most common situations, closely followed by CFMc and CFMe.

Note that, while we state that UPL performs well relative to other models, notable limitations are visible within this study. In particular, UPL consistently demonstrates sudden changes in the pedestrian path across all scenarios, which indicates the unrealistic feature for collision avoidance. In SOSP, UPL predicts a slowdown of the pedestrian as it approaches the obstacle. For Parallel-Ped scenarios, UPL predicts a path change of B (and subsequent overtake of A), which contrasts experiments. Furthermore, in the multiple obstacle scenario (MOSP), UPL exhibits noticeable oscillations, suggesting an unrealistic fluctuation in pedestrian movement when navigating around obstacles. Additionally, UPL predicts increasingly unsuccessful trajectories for higher obstacle densities in MOSP situations. These features imply difficulties in accurately replicating the decision-making process and adaptive behaviors observed in real-life pedestrian dynamics. Thus even though UPL provides the most reasonable predictions in most scenarios, it shows significant limitations that need to be corrected. To the best of our knowledge, no other study has compared all these models for such commonly occurring situations. Thus, this study

will help future researchers in the selection of appropriate pedestrian models for a given scenario.

6 Supplementary Material

The supplementary material offers further insights to enhance the findings discussed in the main manuscript. While the main text sets a benchmark for pedestrian dynamics models in common scenarios, the supplementary section addresses several key aspects: the determination of the time step size used in simulations, the unrealistic outcomes produced by models in the MOSP scenario, parameter selection for the Social Force Model (SFM), and additional trajectory data for the SOSP scenario.

7 Acknowledgement

The authors would like to extend their appreciation to all the students of IIT Kanpur who participated as volunteers in this study. We are also deeply grateful to the security personnel and administration of IIT Kanpur for their invaluable assistance in conducting the experiments. Additionally, we acknowledge the foundational contributions made by former dual-degree students of IIT Kanpur—Ishan Prashant, Amulya Kale, and Satyendra Pandey—whose work significantly informed this study. This research was supported by the CRG project (Sanction number: IME/SERB/2023137) from the Science and Engineering Research Board (SERB), Department of Science and Technology (DST), India.

References

- [1] I. Karamouzas, B. Skinner, and S. J. Guy, “Universal power law governing pedestrian interactions,” *Physical review letters*, vol. 113, no. 23, p. 238701, 2014.
- [2] D. Helbing and P. Molnar, “Social force model for pedestrian dynamics,” *Physical review E*, vol. 51, no. 5, p. 4282, 1995.
- [3] M. Chraibi, A. Seyfried, and A. Schadschneider, “Generalized centrifugal-force model for pedestrian dynamics,” *Physical Review E*, vol. 82, no. 4, p. 046111, 2010.
- [4] M. Moussaïd, D. Helbing, and G. Theraulaz, “How simple rules determine pedestrian behavior and crowd disasters,” *Proceedings of the National Academy of Sciences*, vol. 108, no. 17, pp. 6884–6888, 2011.
- [5] D. Helbing, A. Johansson, and H. Z. Al-Abideen, “Dynamics of crowd disasters: An empirical study,” *Physical review E*, vol. 75, no. 4, p. 046109, 2007.
- [6] M. Haghani and M. Sarvi, “Following the crowd or avoiding it? empirical investigation of imitative behaviour in emergency escape of human crowds,” *Animal behaviour*, vol. 124, pp. 47–56, 2017.

- [7] C. Appert-Rolland, J. Pettré, A.-H. Olivier, W. Warren, A. Duigou-Majumdar, E. Pinsard, and A. Nicolas, “Experimental study of collective pedestrian dynamics,” *arXiv preprint arXiv:1809.06817*, 2018.
- [8] M. Moussaïd, D. Helbing, S. Garnier, A. Johansson, M. Combe, and G. Theraulaz, “Experimental study of the behavioural mechanisms underlying self-organization in human crowds,” *Proceedings of the Royal Society B: Biological Sciences*, vol. 276, no. 1668, pp. 2755–2762, 2009.
- [9] C. Feliciani and K. Nishinari, “Empirical analysis of the lane formation process in bidirectional pedestrian flow,” *Physical Review E*, vol. 94, no. 3, p. 032304, 2016.
- [10] A. Schadschneider and A. Seyfried, “Empirical results for pedestrian dynamics and their implications for modeling,” *Networks and Heterogeneous media*, vol. 6, no. 3, pp. 545–560, 2011.
- [11] K. B. Kramer and G. J. Wang, “Social distancing slows down steady dynamics in pedestrian flows,” *Physics of Fluids*, vol. 33, no. 10, 2021.
- [12] D. Helbing, A. Johansson, J. Mathiesen, M. H. Jensen, and A. Hansen, “Analytical approach to continuous and intermittent bottleneck flows,” *Physical review letters*, vol. 97, no. 16, p. 168001, 2006.
- [13] C. Von Krüchten and A. Schadschneider, “Empirical study on social groups in pedestrian evacuation dynamics,” *Physica A: Statistical Mechanics and its Applications*, vol. 475, pp. 129–141, 2017.
- [14] N. Shiwakoti, X. Shi, and Z. Ye, “A review on the performance of an obstacle near an exit on pedestrian crowd evacuation,” *Safety science*, vol. 113, pp. 54–67, 2019.
- [15] S. Heliövaara, J.-M. Kuusinen, T. Rinne, T. Korhonen, and H. Ehtamo, “Pedestrian behavior and exit selection in evacuation of a corridor—an experimental study,” *Safety science*, vol. 50, no. 2, pp. 221–227, 2012.
- [16] C. Lui, N. K. Fong, S. Lorente, A. Bejan, and W. K. Chow, “Constructal design for pedestrian movement in living spaces: Evacuation configurations,” *Journal of Applied Physics*, vol. 111, no. 5, 2012.
- [17] C. Lui, N. K. Fong, S. Lorente, A. Bejan, and W. K. Chow, “Constructal design of pedestrian evacuation from an area,” *Journal of Applied Physics*, vol. 113, no. 3, 2013.
- [18] A. Gorrini, S. Bandini, M. Sarvi, C. Dias, and N. Shiwakoti, “An empirical study of crowd and pedestrian dynamics: the impact of different angle paths and grouping,” *Transportation Research Record*, vol. 41, p. 42, 2013.

- [19] N. Shiwakoti, X. Shi, Z. Ye, and W. Wang, “Empirical study on pedestrian crowd behaviour in right angled junction,” in *37th Australasian Transport Research Forum (ATRF)*, vol. 1, 2015.
- [20] L. Lian, X. Mai, W. Song, Y. K. K. Richard, X. Wei, and J. Ma, “An experimental study on four-directional intersecting pedestrian flows,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2015, no. 8, p. P08024, 2015.
- [21] A. Seyfried, B. Steffen, W. Klingsch, and M. Boltes, “The fundamental diagram of pedestrian movement revisited,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2005, no. 10, p. P10002, 2005.
- [22] A. Seyfried, B. Steffen, W. Klingsch, T. Lippert, and M. Boltes, “The fundamental diagram of pedestrian movement revisited—empirical results and modelling,” in *Traffic and Granular Flow'05*, pp. 305–314, Springer, 2007.
- [23] A. Seyfried, M. Boltes, J. Kähler, W. Klingsch, A. Portz, T. Rupprecht, A. Schadschneider, B. Steffen, and A. Winkens, “Enhanced empirical data for the fundamental diagram and the flow through bottlenecks,” *Pedestrian and evacuation dynamics 2008*, pp. 145–156, 2010.
- [24] A. Lerner, Y. Chrysanthou, A. Shamir, and D. Cohen-Or, “Data driven evaluation of crowds,” in *Motion in Games: Second International Workshop, MIG 2009, Zeist, The Netherlands, November 21-24, 2009. Proceedings 2*, pp. 75–83, Springer, 2009.
- [25] J. Liddle, A. Seyfried, B. Steffen, W. Klingsch, T. Rupprecht, A. Winkens, and M. Boltes, “Microscopic insights into pedestrian motion through a bottleneck, resolving spatial and temporal variations,” *Collective dynamics*, vol. 7, pp. 1–23, 2022.
- [26] M. Haghani and M. Sarvi, “Crowd behaviour and motion: Empirical methods,” *Transportation research part B: methodological*, vol. 107, pp. 253–294, 2018.
- [27] M. Haghani, “Empirical methods in pedestrian, crowd and evacuation dynamics: Part i. experimental methods and emerging topics,” *Safety science*, vol. 129, p. 104743, 2020.
- [28] X. Shan, J. Ye, and X. Chen, “Critical walking space requirement for collision avoidance of pedestrians: an experimental study,” in *CICTP 2014: Safe, Smart, and Sustainable Multimodal Transportation Systems*, pp. 2369–2380, 2014.
- [29] D. R. Parisi, P. A. Negri, and L. Bruno, “Experimental characterization of collision avoidance in pedestrian dynamics,” *Physical Review E*, vol. 94, no. 2, p. 022318, 2016.
- [30] S. Cao, J. Zhang, W. Song, R. Zhang, *et al.*, “The stepping behavior analysis of pedestrians from different age groups via a single-file experiment,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2018, no. 3, p. 033402, 2018.

- [31] V. Renaudin, M. Susi, and G. Lachapelle, “Step length estimation using handheld inertial sensors,” *Sensors*, vol. 12, no. 7, pp. 8507–8525, 2012.
- [32] V. Racic, A. Pavic, and J. Brownjohn, “Experimental identification and analytical modelling of human walking forces: Literature review,” *Journal of Sound and Vibration*, vol. 326, no. 1-2, pp. 1–49, 2009.
- [33] K. Jain, A. Gupta, I. S. Dalal, A. Tripathi, and S. Prawesh, “Automatic estimation of pedestrian gait features using a single camera recording: Algorithm and statistical analysis for gender difference and obstacle interactions,” *Under Review*.
- [34] A. Bottinelli and J. L. Silverberg, “When dense crowds act like soft solids,” *Physics Today*, vol. 72, no. 9, pp. 70–71, 2019.
- [35] D. Helbing, I. Farkas, and T. Vicsek, “Simulating dynamical features of escape panic,” *Nature*, vol. 407, no. 6803, pp. 487–490, 2000.
- [36] M. Chraibi, A. Seyfried, A. Schadschneider, and W. Mackens, “Quantitative description of pedestrian dynamics with a force-based model,” in *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, vol. 3, pp. 583–586, IEEE, 2009.
- [37] X. Hu, T. Chen, and Y. Song, “Anticipation dynamics of pedestrians based on the elliptical social force model,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 33, no. 7, 2023.
- [38] A. Johansson, D. Helbing, and P. K. Shukla, “Specification of the social force pedestrian model by evolutionary adjustment to video tracking data,” *Advances in complex systems*, vol. 10, no. supp02, pp. 271–288, 2007.
- [39] W. Yu, R. Chen, L.-Y. Dong, and S. Dai, “Centrifugal force model for pedestrian dynamics,” *Physical Review E*, vol. 72, no. 2, p. 026112, 2005.