



OPEN

DATA DESCRIPTOR

Dense Crowd Dynamics and Pedestrian Trajectories: A Multiscale Field Dataset from the Festival of Lights in Lyon

Oscar Dufour¹✉, Huu-Tu Dang², Jakob Cordes^{3,4}, Raphael Korbacher⁵, Mohcine Chraïbi³✉, Benoit Gaudou²✉, Alexandre Nicolas¹✉ & Antoine Tordeux⁵✉

The dynamics of dense crowds have received considerable attention from researchers seeking fundamental understanding or aiming to develop data-driven algorithms to predict pedestrian trajectories. However, current research mainly relies on data collected in controlled settings. We present one of the first comprehensive field datasets describing dense pedestrian dynamics at different scales, from contextualized macroscopic crowd flows over hundreds of meters to microscopic trajectories (around 7000 individual trajectories). In addition, a sample of GPS traces, some statistics of contacts and pushes, and a list of non-standard crowd phenomena observed in the video recordings are provided. The data were collected during the 2022 Festival of Lights in Lyon, France, as part of the French-German MADRAS project and cover densities up to 4 pedestrians per square meter. We suggest using this extensive dataset, acquired in complex real-world settings, to benchmark models of pedestrian dynamics.

Background & Summary

Large gatherings raise challenges regarding public safety and flow management; these challenges tend to be all the more acute as the crowd is large and dense. Religious festivals, music concerts, and significant outdoor events are therefore of particular concern, with a record of poor crowd management and, in the worst cases, deadly crowd crushes^{1–4}. Beyond the rules of thumb that have been refined over the years, a deeper fundamental understanding of the dynamics of dense crowds will be instrumental for more efficient event planning and crowd management. Here, a crowd will be described as dense if its density exceeds the arbitrary threshold of 1.5 or 2 ped/m², thus falling in Fruin's Level of Service⁵; critical conditions with extreme densities (above 8 ped/m²), which must be avoided in practice, are left out of our scope.

The present theoretical grasp of dense crowd dynamics mainly stems from controlled experiments⁶, conducted in idealized settings. This protocol enables researchers to finely control variables and settings to observe a whole gamut of none-too-common scenarios, such as emergency evacuations⁷ or high-density flows and crossings. The Research Center of Jülich, in particular, has collected numerous pedestrian trajectories in a wide array of such experiments^{8,9}. Let us also mention, among others, the datasets collected by the groups of Haghani¹⁰ and Zuriguel¹¹ to explore crowd dynamics during emergency evacuation drills, Murakami *et al.*¹² to probe the emergence of unidirectional and bidirectional pedestrian flows. An open-access data archive can be found here¹³. However, the controlled conditions may substantially differ from reality, and the participants are aware of their involvement in the research. Another drawback of controlled experiments is that they do not afford a comprehensive picture of the crowd; the broader context in which the dynamics occur is missing.

¹Université Claude Bernard Lyon 1, CNRS, Institut Lumière Matière, UMR5306, F-69100, Villeurbanne, France. ²UMR 5505 IRIT, Université Toulouse Capitole, Toulouse, France. ³Institute of Advanced Simulation, Forschungszentrum Jülich GmbH, Jülich, Germany. ⁴Institut für Theoretische Physik, Universität zu Köln, Köln, Germany. ⁵Fakultät für Maschinenbau und Sicherheitstechnik, Bergische Universität Wuppertal, Wuppertal, Germany. ✉e-mail: oscar.dufour@univ-lyon1.fr; m.chraibi@fz-juelich.de; benoit.gaudou@ut-capitole.fr; alexandre.nicolas@cnrs.fr; tordeux@uni-wuppertal.de

The thirst for *field data* to train data-based methods, such as machine learning algorithms, remains unquenched for dense crowds: field studies typically involve situations of low density. For instance, the widely used ETH¹⁴ and UCY¹⁵ datasets, which originate from surveillance videos, capture pedestrian scenes at density 0.1–0.5 ped/m², with many avoidance situations. In this regime, the pedestrian dynamics are believed to be governed by different mechanics than at higher density^{5,16,17}. Empirical datasets often encompass a heterogeneous mix of road users, as in the Stanford Drone Dataset (SDD)¹⁸, including pedestrians, cyclists, skateboarders, cars, and buses. While these datasets capture small scenes, the Grand Central Station Dataset (GS)¹⁹, collected in New York, covers a vast area. Although one scene can contain hundreds of pedestrians, the average density is below 0.2 ped/m² due to the region's size. The Edinburgh Informatics Forum Dataset (EIF)²⁰ provides trajectories of 92 000 people in a university playground from an overhead camera, with minute average densities (most of the time just a few pedestrians at a time). Also worth mentioning are mass-gathering studies relying on a sparse sample of smartphone signals²¹. Besides, the group of Toschi and Corbetta at the Technical University of Eindhoven have collected extensive datasets of pedestrian trajectories in the Eindhoven train station and on the university campus in the last decade^{22,23}. Several reviews on existing field studies of pedestrian trajectories have recently appeared^{24–26}.

The present contribution aims to make up for the lack of field data on dense crowds by providing a comprehensive picture of pedestrian flows at a large gathering around a hot spot of a major cultural and entertainment event, the 2022 edition of the Festival of Lights, which took place in Lyon, France. For this purpose, we collected various data relevant to pedestrian dynamics and crowd management. These data cover an extensive range of length scales, by all standards in the field, from the global flow picture and contextual elements down to individual pedestrian trajectories and some statistics on physical contacts. In the following, an emphasis shall be placed on all observations that depart from what is typically prescribed or observed in controlled experiments, thus further highlighting the added value of actual field data.

Lyon's yearly Festival of Lights is a four-evening event (from December 7 to 11 in 2022, mostly from 7 pm to 11 pm) wherein the city is lit up remarkably. Originally a religious tribute to the Virgin Mary, it has become a massive international festival renowned for its innovative light shows and artistic projections on historic buildings. The event attracts millions of local and international visitors (more than 2 million officially in 2022²⁷). Key attractions include *Place des Terreaux* and *Place Saint-Jean*, reportedly attended by 150 000 and 80 000 spectators every night, respectively, in 2022²⁷. Quite interestingly, for our purposes, managing the associated crowd flows is one of the most prickly issues for the event organizers²⁸. They aim to ensure smooth flows and reasonable delays for a pleasant experience. Still, above all, to ward off crowd accidents, after a difficult situation witnessed in the 2000s (information obtained from a private communication with the organizers of the event) and the tragedies that have occurred in massive entertainment events around the world, e.g. at the Love Parade in Duisburg, Germany, in 2010² or during Halloween on the streets Seoul, Korea, in 2022²⁹. This is achieved by regulating flows at different scales: macroscopically, by suggesting routes through the city to visit the multiple light installations, for instance, starting at *Place Bellecour*, moving to *Place des Terreaux*, and then exploring *Vieux Lyon*, especially near *Saint-Jean's Cathedral*; mesoscopically, by installing barriers and safety agents, particularly near *Place des Terreaux* and imposing unidirectional flows in many streets to ease congestion microscopically by continuously monitoring the event with CCTV. Our methodology explores these three scales for crowd flows, focusing on the microscopic one at the central location of *Place des Terreaux*. The crowd's movement was notably monitored with strategically placed cameras (Fig. 1) installed by us or by the city of Lyon.

Methods

General organization of the data collection campaign. Approximately ten staff members planned and carried out the data collection campaign. The aim was to capture a complete picture of pedestrian motion during the Festival of Lights. For that purpose, we combined various methods targeting different scales. Macroscopically, we inspected the broad patterns of macroscopic crowd flows on the ground around *Place des Terreaux* and recorded the scene with a broad overview. We surveyed passers-by close to the entrance of the square, asking them the following questions:

- 'How many people were with you?'
- 'How many children were with you?'
- 'What was the last screening you attended before the one at Place des Terreaux?'
- 'What screening do you plan to attend next? (You can answer that you don't know)'

We strove to limit the selection bias on the 79 respondents by making no distinction about general or physical appearance and by ensuring that we interviewed a sample from around the square and not limited to a single street.

Mesoscopically, we recruited participants willing to share their GPS data and record how often they bumped into other people.

Microscopically, we installed several cameras filming specific zones from the top (with all required authorizations to ensure privacy preservation). Below, we distinguish between *TopView* cameras (labeled 1, 2, 3, 5, 6, 7, 8), filming from a zenithal position, and *LargeView* cameras (labeled 4), recording an even broader view that encompasses the whole square, but with a more tilted (less vertical) angle. An overview of the data collected in this study is presented in Fig. 2.

GPS data of recruited participants and collision counts. A group of 24 undergraduate and doctoral students from *Institut Lumière Matière* in Lyon were invited to participate in the field study. More precisely, they

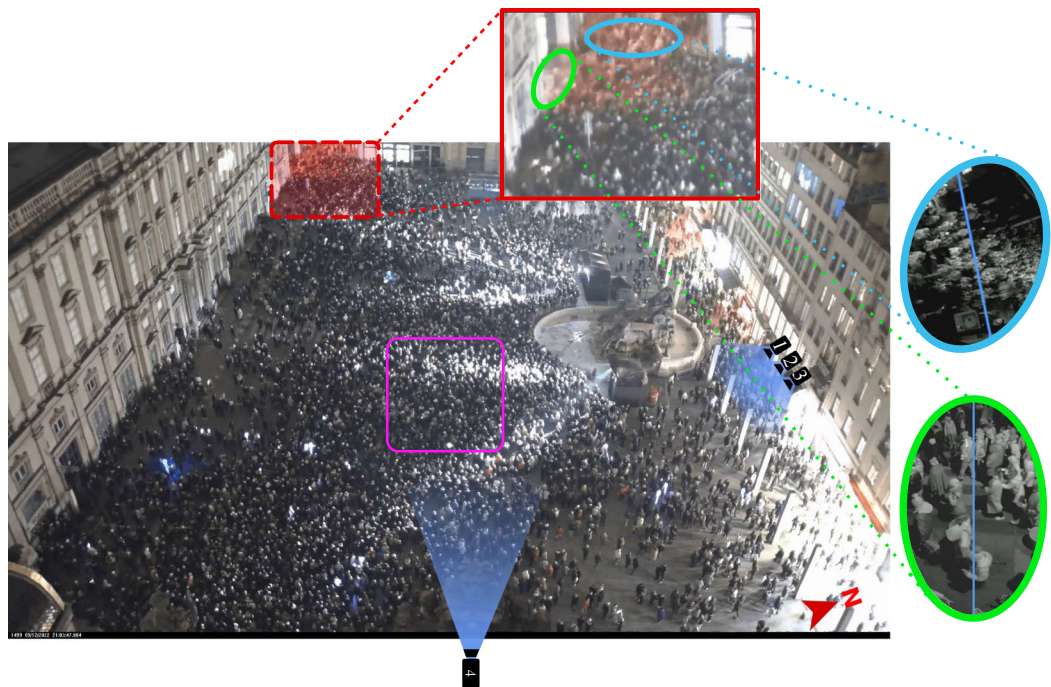


Fig. 1 The thick crowd gathered in and around *Place des Terreaux* at the end of a light show on December 12, 2022, at 9:41 PM. The cameras' fields of view are highlighted in blue; the pink rounded square delimits the area where pedestrians were exhaustively tracked in the *LargeView* datasets³⁰, *LargeView* Trajectories³⁷. Inset: Zoom on two exit streets where we monitored the outflow.

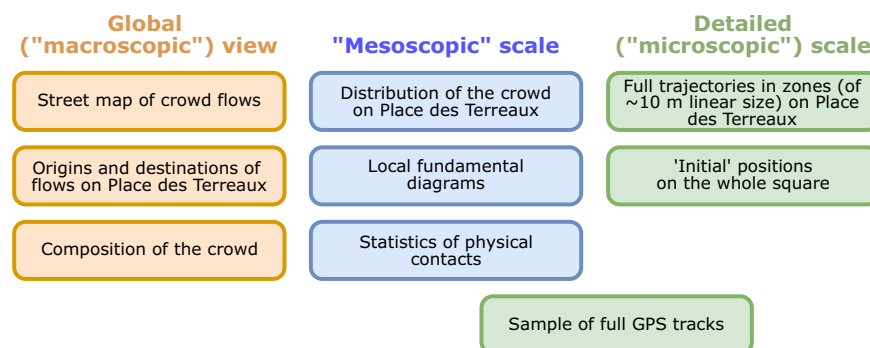


Fig. 2 Scales probed in this study and synthesis of the data collected at each scale.

were asked to follow a general route (from the queue at the entrance to the square, on *Rue du Président Edouard Herriot* to the exit on *Rue Lanterne*, following the flow of the crowd), behave as a standard spectator. Meanwhile, their GPS positions were recorded using the *GeoTracker* application on their smartphones. Although the accuracy of the measurements varied from phone to phone, the trajectories obtained were generally reliable, with a margin of error of around 10 meters on absolute positions. Besides, each participant used a stopwatch on their smartphone to record each time they collided with another pedestrian. Upon synchronization with the GPS data, this method allowed us to precisely determine the time and location of the collisions. Each participant was taught what to consider a collision, which was demonstrated by jostling them hard. Therefore, minor rubbing of the clothing was not recorded. Of the 24 students involved in that experiment, 16 were able to provide us with a detailed list of physical contacts, among which eight could be coupled to GPS data (Dataset³⁰, *GPS Data*). Our efforts to involve on-site spectators in the campaign proved fruitless; none managed to send us GPS data.

Video processing and pedestrian tracking for the *TopView* cameras. To obtain a finer view of pedestrian trajectories, we deployed lightweight cameras *SJCAM A10* in strategic locations around *Place des Terreaux*. These cameras filmed the scene from a zenith perspective. These cameras were selected for their night vision capabilities and long battery life; three of them, labeled 1, 2, and 3 in Fig. S1 in the Supplementary Material, were placed on the north side of *Place des Terreaux*, protruding from the windows of an apartment *Airbnb* and a restaurant on the second floor. They captured the bidirectional movement of pedestrians below. Another camera, numbered 8, was mounted atop an existing post in the square's South-East corner to monitor incoming flows;

however, its nighttime footage was unusable due to lighting problems. Two additional cameras, numbered 5 and 6, were placed on *Rue Constantine*, one of the main exits after the light show, hanging from the balconies of an *Airbnb* apartment to film the pedestrian egress from above. A final camera numbered 7 was temporarily held at the end of a stick in the South-West corner to provide a closer view of the pedestrian outflow. Overall, the portable *SJCAM A10* cameras recorded nearly 200 GB of video at 30 frames per second.

From the collected footage, pedestrian trajectories were extracted from 10 excerpts (see Table 3 in the Supplementary Material and Dataset³⁰, *TopView Trajectories*), using the *PeTrack* software^{31,32}. This software is commonly used to automatically detect and track pedestrian heads in controlled experimental settings. The process notably involves calibrating the cameras to match pixel coordinates with real-world coordinates, which is split into two distinct phases: internal and external calibration. Internal calibration corrects the specific optical distortions of each camera lens by determining the optimum parameters to transform a reference pattern (such as a checkerboard) into its recorded image. Then, external calibration yields the conversion between real-world coordinates and pixel coordinates using three successive operations: a rotation and a translation of the reference frame, followed by a projection to move from the camera frame to the screen frame. The parameters of these operations minimize the differences between the known real-world coordinates of objects (here, staff members) positioned at predefined positions (here, at regularly spaced positions on a virtual ‘grid’) and the associated pixel coordinates. Such calibration is only an approximation if the recorded pedestrians are of unequal heights, especially without a stereoscopic camera to reconstruct the scene in three dimensions. Still, the inaccuracy is even lower as the camera is positioned higher and films from a zenithal viewpoint. Here, we expect the maximum uncertainty due to unequal heights to be of order³³ $\delta h \cdot \frac{D}{H}$, where $H \approx 12$ m is the altitude of the camera concerning the heads, $D \approx 5$ – 10 m is the maximal horizontal stretch between the camera axis and people on the periphery of the scene, and $\delta h \lesssim 20$ cm is the height difference between filmed people and the staff member who served as reference for the calibration; hence below 15 cm in most cases for deployed *SJCAM A10* cameras. The uncertainty in detecting the central point on the head ($\lesssim 5$ cm) and the error should be added to this because not all people stand perfectly upright. Furthermore, occlusion phenomena due to the perspective can increase the measurement uncertainty for small pedestrians.

After the calibration step, the *PeTrack* software semi-automatically detects and tracks pedestrians on the videos. Compared to controlled experimental conditions, some difficulties arose. In our case, this task was mainly complicated by the lighting conditions. Thus, pedestrian heads were first detected manually and then tracked from frame to frame using the extended pyramidal iterative Lucas Kanade feature tracker integrated into *PeTrack*. Although this method is surprisingly robust, manual corrections were often required, mainly when the illumination suddenly changed from dark to bright areas.

Video processing and pedestrian tracking for the *LargeView* camera. In addition to the previously mentioned *TopView* cameras, we have gathered extensive *LargeView* video footage from cameras that offer a bird’s-eye view of the entire square. Two of these cameras are located at *Place Saint-Jean*. Another camera (numbered 4 in Fig. S1 in the Supplementary Material) captures *Place des Terreaux* from the City Hall tower, approximately 48 meters above ground level. Although all recordings are accessible, only the videos from *Place des Terreaux* have been thoroughly analyzed and discussed here.

The expansive view provided by the camera, numbered 4, along with variations in lighting, precludes automated tracking using *PeTrack*, in favor of a largely manual approach. Internal calibration was deemed unnecessary because the camera lens exhibited minimal optical distortion: straight lines in reality appear as straight lines on the video. Complete external calibration was performed by positioning a staff member at predefined evenly spaced points in the square 32 and applying the geometric transformations outlined in Sect. 2. The resulting precision for *absolute* positions, assessed using independently collected positions (either of the same staff member or others), ranged from 10 cm to approximately 2 meters at the farthest end of the square and over 80 meters from the camera horizontally. Given that this inaccuracy is primarily geometric, it is expected that the positions *relative* between pedestrians and their neighbours are significantly more precise.

Initially, we manually identified the positions of all individuals throughout the square at a specific time point, occasionally reviewing video frames to locate pedestrians temporarily obscured from view. This was done after a show cycle, as people began to exit, on Thursday 8 December 2022. Subsequently, we tracked a random sample of approximately 270 individuals over several seconds, with around 100 of them being tracked for a duration of 20 seconds (Dataset³⁰, *LargeView Trajectories*), file “*LargeView_tracers.txt*”). Finally, we focus on a particular area of interest, where opposing flows of people who do not necessarily follow traffic directions meet. The area has a square shape (before the correction of the geometric distortion) and is located near the fountain, a crucial convergence point of high density. Using homemade Python-based software, we manually tracked the trajectories of all observable people in this area over around 30 seconds, at a rate of typically two frames per second. Trajectories were upsampled to 10 Hz by linear interpolation and exported into CSV files (dataset³⁰, *LargeView Trajectories*) files “*LargeView_zoom_A.txt*” and “*LargeView_zoom_O.txt*”).

It should be noted that, compared to the *TopView* videos, these ones (addressing a more complex flow scenario) are of lower resolution and quality. Combined with the varying illumination, this hindered the detection of some individuals, especially children and shorter people, in specific frames. The datasets may thus exhibit imperfections, such as some individuals escaping detection and occasional swaps between trajectories. However, our subsequent tests have shown that they are accurate to a very large degree and nearly comprehensive (see *Technical validation*).

Conversion into global coordinates. All trajectories were mapped to global coordinates using the RGF-93 Lambert-93 coordinate reference system (EPSG:2154) for precise global positioning. This transformation

involved adjusting the positions of several landmarks visible in the videos and satellite imagery. The geometric shapes and locations of all obstacles in the square, including the fountain, temporary crowd barriers surrounding it, bollards, and Buren columns, were determined via direct measurements, together with satellite imagery and photographs taken on-site; they are provided as an external dataset³⁰, *Geometry*].

Quantitative indicators. Various relevant static and dynamic indicators can be computed using the detailed pedestrian trajectories.

Flow rate. The outflow rate during an egress from *Place des Terreaux* was measured by drawing virtual cross-section lines at the two main exits, as shown in Fig. 1. People crossing these lines were counted manually by watching the recorded videos.

Density field. To be useful, the extracted microscopic trajectories often need to be smoothed into continuous fields. Specifically, the local density field³⁴, denoted as $\rho(\mathbf{r}, t)$, is derived by computing the convolution of the microscopic particle density $\rho_\mu(\mathbf{r}, t) = \frac{1}{A} \sum_j \delta(\mathbf{r} - \mathbf{r}_j(t))$ - where $\delta(\cdot)$ is the Dirac delta function and A is the surface area - with a Gaussian kernel $\phi_\xi(\mathbf{r}) \propto \exp(-\frac{\mathbf{r}^2}{2\xi^2})$ whose integral is normalized to 1, for a chosen half-width ξ , viz.:

$$\rho(\mathbf{r}, t) = \int_A \rho_\mu(\mathbf{r} - \mathbf{r}') \phi_\xi(\mathbf{r}') d^2\mathbf{r}' \quad (1)$$

For further smoothing, the time dependence can also be coarse-grained by averaging the field over a short time window Δt :

$$\bar{\rho}(\mathbf{r}, t) = \frac{1}{\Delta t} \int_{t-\frac{\Delta t}{2}}^{t+\frac{\Delta t}{2}} \rho(\mathbf{r}, t') dt' \quad (2)$$

Velocity field. Similarly, the microscopic velocities $\mathbf{v}_j(t) = \frac{1}{\delta t} \cdot [\tilde{\mathbf{r}}_j(t + \delta t) - \tilde{\mathbf{r}}_j(t)]$, estimated from the vector difference between two positions of pedestrian j over a small time interval δt (with $\delta t = 0.5$ or 1 second in this context), can be coarse-grained. The trajectories are initially smoothed using a second-order Butterworth low-pass filter. Trajectories that are too short for effective filtering remain unfiltered. Near the start (t_s) and end (t_e) of each trajectory, linear interpolation is applied between the raw trajectory \mathbf{r} and the filtered trajectory $\tilde{\mathbf{r}}$ to address the Butterworth filter's limitations when past or future data points are absent. This interpolation is expressed as $\tilde{\mathbf{r}}_j(t) \leftarrow \alpha(t)\mathbf{r}_j(t) + [1 - \alpha(t)]\tilde{\mathbf{r}}_j(t)$, where $\alpha(t) = \max\{e^{-(t-t_s)}, e^{-(t_e-t)}\}$. Note that at the start ($t = t_s$) and end ($t = t_e$) of the trajectory, α is close to 1, giving more weight to the original data. This helps preserve the trajectory's endpoints. In the middle, α decreases, giving more weight to the smoothed data. The beginning and end of the raw trajectories are thus preserved, reducing artefacts from filtering. Subsequently, the trajectories are transformed into a velocity field through Gaussian convolution:

$$\mathbf{v}(\mathbf{r}, t) = \frac{\sum_j \mathbf{v}_j(t) \phi_\xi(\mathbf{r} - \tilde{\mathbf{r}}_j(t))}{\sum_j \phi_\xi(\mathbf{r} - \tilde{\mathbf{r}}_j(t))} \quad (3)$$

The resulting field exhibits abrupt variations due to many individuals deviating from, or walking counter to, the primary local flow direction. Like the density field, the velocity field can be averaged over a finite time interval Δt to smooth out these variations:

$$\bar{\mathbf{v}}(\mathbf{r}, t) = \frac{1}{\Delta t} \int_{t-\frac{\Delta t}{2}}^{t+\frac{\Delta t}{2}} \mathbf{v}(\mathbf{r}, t') dt' \quad (4)$$

The coarse-grained picture given by the smooth velocity field masks possible counterflows and fluctuations, whose presence can be ascertained by computing a *variance* field:

$$\text{Var}_{\mathbf{v}}(\mathbf{r}, t) = \frac{\int_{t-\frac{\Delta t}{2}}^{t+\frac{\Delta t}{2}} \sum_j \phi_\xi(\mathbf{r} - \mathbf{r}_j(t')) \|\mathbf{v}_j(t') - \bar{\mathbf{v}}(\mathbf{r}, t')\|^2 dt'}{\int_{t-\frac{\Delta t}{2}}^{t+\frac{\Delta t}{2}} \sum_j \phi_\xi(\mathbf{r} - \mathbf{r}_j(t')) dt'} \quad (5)$$

Trajectories of counter-walking pedestrians j significantly deviate from the continuous flow, thus exhibiting a large variance relative to the velocity field, denoted as $\text{Var}_{\mathbf{v}}^j$. More precisely, the variance is computed by averaging the squared difference between pedestrian j 's velocity $\mathbf{v}_j(t)$ and the coarse-grained velocity $\bar{\mathbf{v}}(\mathbf{r}_j(t), t)$ at that position, over the entire duration of pedestrian j 's trajectory, i.e. $\text{Var}_{\mathbf{v}}^j = \langle \|\mathbf{v}_j(t) - \bar{\mathbf{v}}(\mathbf{r}_j(t), t)\|^2 \rangle$.

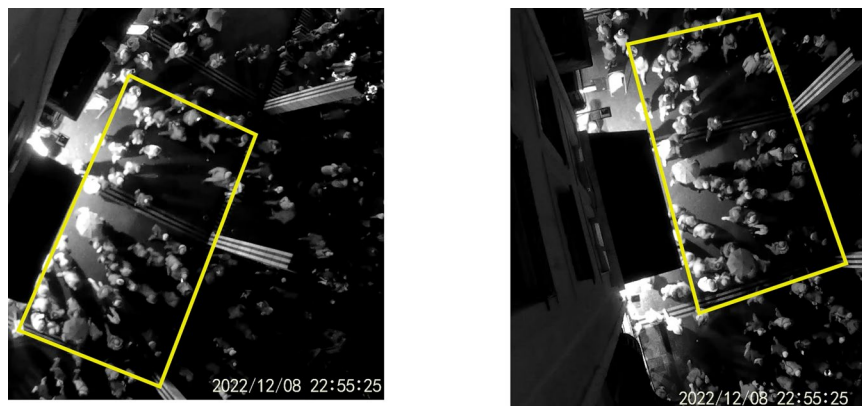


Fig. 3 Point of view for the *TopView_1B* (left panel) and *TopView_2C* (right panel) camera videos, which simultaneously captured the same scene from different locations and angles.

Data Records

The datasets referenced in this study in relation with dense pedestrian dynamics during the 2022 Festival of Lights in Lyon, France, are publicly available and can be accessed via Zenodo³⁰. The datasets include macroscopic crowd flows, microscopic individual trajectories, GPS traces, contact data, and statistics on crowd interactions, providing a valuable resource for further research in pedestrian dynamics. In addition, we have released an open online platform that allows users to visualize and plot most of the collected data.

- We have released an open online platform³⁵ that allows users to visualize and plot most of the collected data; it can be accessed at <https://go.fzj.de/madras-app>.
- The survey results are provided in a CSV file, with each line representing a respondent. A README file is included to help interpret the columns.
- The geometry of the square and the (temporary) obstacles are provided as CSV files using the WKT format for geometries.
- Contact data are compiled in a CSV file, detailing the duration, end time, and instances of contact. A README file is included to assist with column interpretation. GPS data for 10 individuals are linked with the contact data. Note that some individuals in the contact data did not provide GPS trajectories.
- Each file is identified by a number, which corresponds to the camera and recording location (see Fig. S1 in the Supplementary Material), and a letter, which indicates the sequence. The videos from which the trajectories were extracted are also available. Trajectories are given in real-world coordinates at a frequency of 10 Hz after all geometric corrections and interpolation have been applied, but without any smoothing or filtering. They are available in both local coordinates (suffix 'loc') and absolute coordinates (EPSG:2154) for global location. The time reference is set to Friday 9 December 2022, at 8 pm, with time given in seconds before or after this reference.
- Many more videos were collected than we could analyze. We welcome contributions from volunteers to assist with the tracking efforts. We have made our tracking and camera calibration scripts available to support this.

Descriptive information about the data are available in the Supplementary Material and in [please add the reference <https://doi.org/10.48550/arXiv.2410.05288>].

Technical Validation

Trajectory datasets for the *TopView* recordings. All trajectory datasets were visually inspected and manually corrected where necessary. In addition, comparing the overlapping time series of density and mean speed obtained from the *TopView_1B* and *TopView_2C* video recordings, on the one hand, and the *TopView_1C* and *TopView_2D* videos, on the other hand, further validated the results. The fields of view of these cameras are similar (see Fig. 3), and their recordings overlap in time (see Table 3 in the Supplementary Material). We find a relatively small root-mean-square differences (from 5 to 10%) between overlapping sequences, both for the mean speed and for the density (see Table 1 and Fig. 4). Nonetheless, a systematic bias is observed in opposite directions for *TopView_1B/TopView_2C* and *TopView_1C/TopView_2D*. These biases may be explained by perspective effects and partly biased correction of optical and geometric distortions.

Trajectory datasets for the *LargeView* recordings. As mentioned above, owing to the lower resolution of the *LargeView* recordings, the quality of the extracted data is not quite as good. Occasionally, we may have failed to detect shorter individuals or swapped intersecting trajectories. Despite these challenges, two staff members (called 'coders') independently extracted trajectories from different recordings and then analyzed each other's work. The coders largely agreed on the extracted data, although there were occasional disagreements or uncertainties regarding some data points. The common trends observed in the density and velocity fields, as shown in Figs. S4–S5 in the Supplementary Material, which correspond to different days of recording, further support the robustness of the presented data.

File	RMSD Mean Speed [m/s]	RMSD Density [ped/m ²]
TopView_1B / TopView_2C	0.02	0.05
TopView_1C / TopView_2D	0.05	0.07

Table 1. Root mean square differences (RMSD) between the pedestrian mean speed and density time series for the *TopView_1B* and *TopView_2C* video recordings and the *TopView_1C* and *TopView_2D* video recordings, as shown in Fig. 4.

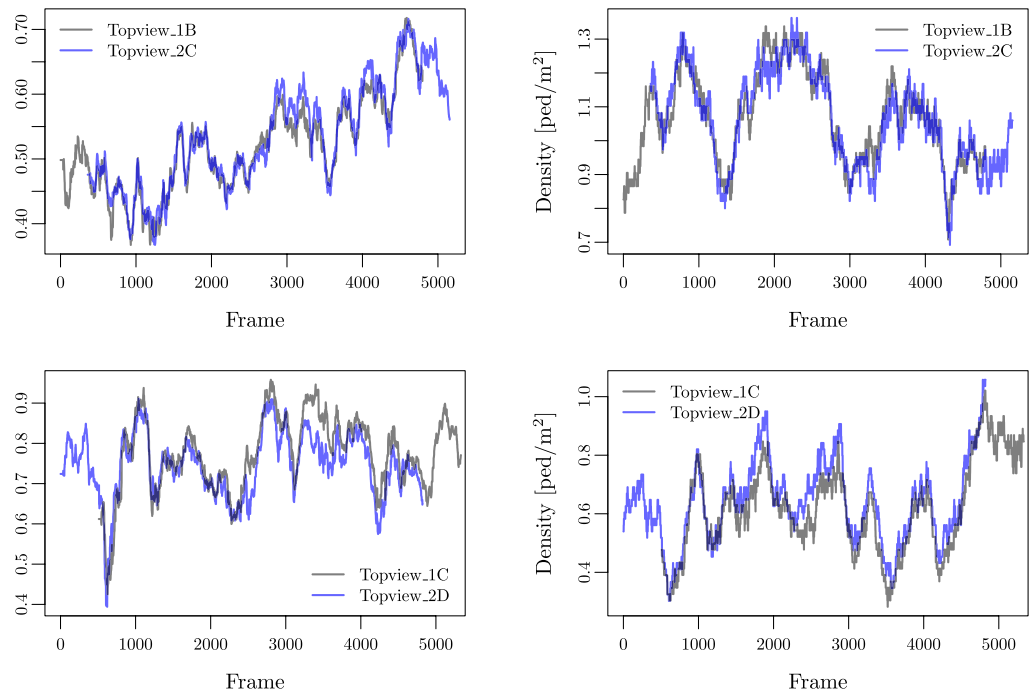


Fig. 4 Superposition of the pedestrian mean speed (left panels) and density (right panels) time series for the *TopView_1B* and *TopView_2C* video recordings (upper panels) and for the *TopView_1C* and *TopView_2D* video recordings (lower panels), which capture the same scene from different locations and angles of view, and which partly overlap in time.

To further validate the *LargeView* trajectory dataset, possible detection or tracking errors were analyzed jointly by the two coders in a second stage. Errors were categorized as either major or minor. Major errors included omissions of clearly visible pedestrians and the creation of non-existent ones. Minor errors involved misidentifying or possibly swapping pedestrians as well as slightly inaccurate clicks on a pedestrian's head. The joint analysis led to the detection of 10 major errors (1.3%) out of ~740 trajectories and 6 minor errors in the *LargeView Zoom_O* dataset, and 3 major errors (0.4%) out of ~740 trajectories and 5 minor errors in the *LargeView Zoom_A* dataset, over 10 seconds. (All these errors were corrected in the final dataset.)

In order to assess errors on the local densities, two sub-regions of rectangular shape were defined at distinct locations in time and space, each measuring $4 \times 6 \text{ m}^2$, and the two coders separately counted all people in these regions, including flickering appearances who were *likely* to be people, even if this was not certain, in order to get an upper bound. Their respective counts typically differed by less than 10%, and exceeded the number of actually tracked pedestrians by 20% to 30% in the *LargeView Zoom_O* dataset and 9% to 12% in the *LargeView Zoom_A* dataset, depending on the location of the rectangle. This leads to the conclusion that the local densities given by our dataset underestimate the actual densities by *at most* 9% to 30%.

Mapping to real-world coordinates. To map the pedestrian positions in pixel coordinates to real-world coordinates, calibration using people standing at predefined positions was performed; the distances between the predefined positions were carefully measured on the ground. In the most distant part of the square, the calibration error on the real-world coordinates (but not the *relative* positions) may reach a couple of meters. Then, after conversion, we successfully checked the compatibility of the crowd positions with the geometry of the premises obtained from *Google Earth* data and our independent positioning of obstacles.

Surveys. Six distinct staff members gave out oral surveys about origins, destinations, and group sizes. In addition to the collected statements, group sizes were also passively observed on the field.

Respect of privacy. The data that we released on a public repository³⁰ do not contain personal or sensitive information. Besides, the video recordings that we collected comply with the GDPR insofar as the recording conditions (from an elevated zenithal viewpoint, at night, without colors) do not allow the identification of passers-by. Finally, the information obtained by means of surveys was aggregated and the results that are presented do not disclose personal details about any of the respondents.

Code availability

Custom code was developed to generate, process, and analyze the datasets presented in this study. Additionally, a Streamlit application enables interactive data exploration and access under specific conditions. The source code is available on GitHub³⁵ under the MIT license distribution, and modification with proper attribution. The repository includes:

- The scripts and notebooks used for data collection, preprocessing, analysis, and visualization.
- The Streamlit application code, along with instructions for installation and deployment³⁵.
- A requirements.txt file specifying the versions of the software and libraries used to ensure compatibility. Essential software versions include PedPy 1.0.2³⁶ as a backend for speed, density and flow calculations.

The repository also includes comprehensive documentation that details the specific variables, parameters, and settings used to facilitate the reproduction of the results and analyses presented in this study. This encompasses, but is not limited to, parameters for data filtering and the statistical analysis methods applied.

Received: 15 October 2024; Accepted: 28 February 2025;

Published online: 30 April 2025

References

1. Helbing, D., Johansson, A. & Al-Abideen, H. Z. Dynamics of crowd disasters: An empirical study. *Physical review E* **75**, 046109, <https://doi.org/10.1103/PhysRevE.75.046109> (2007).
2. Sieben, A. & Seyfried, A. Inside a life-threatening crowd: Analysis of the love parade disaster from the perspective of eyewitnesses. *Safety science* **166**, 106229, <https://doi.org/10.1016/j.ssci.2023.106229> (2023).
3. Feliciani, C., Corbetta, A., Haghighi, M. & Nishinari, K. Trends in crowd accidents based on an analysis of press reports. *Safety science* **164**, 106174, <https://doi.org/10.1016/j.ssci.2023.106174> (2023).
4. Sharma, A. *et al.* Global mass gathering events and deaths due to crowd surge, stampedes, crush and physical injuries-lessons from the seoul halloween and other disasters. *Travel Medicine and Infectious Disease* **52**, <https://doi.org/10.1016/j.tmaid.2022.102524> (2023).
5. Fruin, J. J. Designing for pedestrians: A level-of-service concept. *Highway Research Record* **355**, 1–15 (1971).
6. Haghighi, M. Empirical methods in pedestrian, crowd and evacuation dynamics: Part i. experimental methods and emerging topics. *Safety science* **129**, 104743, <https://doi.org/10.1016/j.ssci.2020.104760> (2020).
7. Garcimartin, A. *et al.* Redefining the role of obstacles in pedestrian evacuation. *New Journal of Physics* **20**, 123025, <https://doi.org/10.1088/1367-2630/aaf4ca> (2018).
8. Zhang, J. & Seyfried, A. Comparison of intersecting pedestrian flows based on experiments. *Physica A: Statistical Mechanics and its Applications* **405**, 316–325, <https://doi.org/10.1016/j.physa.2014.03.004> (2014).
9. Cao, S., Seyfried, A., Zhang, J., Holl, S. & Song, W. Fundamental diagrams for multidirectional pedestrian flows. *Journal of Statistical Mechanics: Theory and Experiment* **2017**, 033404, <https://doi.org/10.1088/1742-5468/aa620d> (2017).
10. Haghighi, M., Sarvi, M. & Shahhoseini, Z. Evacuation behaviour of crowds under high and low levels of urgency: Experiments of reaction time, exit choice and exit-choice adaptation. *Safety science* **126**, 104679, <https://doi.org/10.1016/j.ssci.2020.104679> (2020).
11. Pastor, J. M. *et al.* Experimental proof of faster-is-slower in systems of frictional particles flowing through constrictions. *Physical Review E* **92**, 062817, <https://doi.org/10.1103/PhysRevE.92.062817> (2015).
12. Murakami, H., Feliciani, C., Nishiyama, Y. & Nishinari, K. Mutual anticipation can contribute to self-organization in human crowds. *Science Advances* **7**, eabe7758, <https://doi.org/10.1126/sciadv.abe7758> (2021).
13. Forschungszentrum Jülich, Jülich Supercomputing Centre. Data archive of experiments on pedestrian dynamic, <http://ped.fz-juelich.de/dataarchive>. Accessed: 2021-08-01.
14. Pellegrini, S., Ess, A., Schindler, K. & Van Gool, L. You'll never walk alone: Modeling social behavior for multi-target tracking. In *2009 IEEE 12th International Conference on Computer Vision*, 261–268, <https://doi.org/10.1109/ICCV.2009.5459260> (IEEE, 2009).
15. Lerner, A., Chrysanthou, Y. & Lischinski, D. Crowds by example. In *Computer Graphics Forum*, vol. 26, 655–664, <https://doi.org/10.1111/j.1467-8659.2007.01089.x> (Wiley Online Library, 2007).
16. Best, A., Narang, S., Curtis, S. & Manocha, D. Densesense: Interactive crowd simulation using density-dependent filters. In *Symposium on Computer Animation*, 97–102, <https://doi.org/10.2312/sca.20141127> (2014).
17. Cordes, J., Schadschneider, A. & Nicolas, A. Dimensionless numbers reveal distinct regimes in the structure and dynamics of pedestrian crowds. *PNAS nexus* **3**, pgae120, <https://doi.org/10.1093/pnasnexus/pgae120> (2024).
18. Robicquet, A., Sadeghian, A., Alahi, A. & Savarese, S. Learning social etiquette: Human trajectory understanding in crowded scenes. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VIII* **14**, 549–565, https://doi.org/10.1007/978-3-319-46484-8_33 (Springer, 2016).
19. Zhou, B., Wang, X. & Tang, X. Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2871–2878, <https://doi.org/10.1109/CVPR.2012.6248013> (IEEE, 2012).
20. Majecka, B. Statistical models of pedestrian behaviour in the forum. *Master's thesis, School of Informatics, University of Edinburgh* (2009). weblink.
21. Wirz, M. *et al.* Probing crowd density through smartphones in city-scale mass gatherings. *EPI Data Science* **2**, 1–24, <https://doi.org/10.1140/epjds17> (2013).
22. Corbetta, A., Bruno, L., Muntean, A. & Toschi, F. High statistics measurements of pedestrian dynamics. *Transportation Research Procedia* **2**, 96–104, <https://doi.org/10.1016/j.trpro.2014.09.013> (2014).
23. Corbetta, A., Schilders, L. & Toschi, F. High-statistics modeling of complex pedestrian avoidance scenarios. *Crowd Dynamics, Volume 2: Theory, Models, and Applications* 33–53, https://doi.org/10.1007/978-3-030-50450-2_3 (2020).
24. Amirian, J. *et al.* Opentraj: Assessing prediction complexity in human trajectories datasets. In *Asian Conference on Computer Vision (ACCV), CONF*, https://doi.org/10.1007/978-3-030-69544-6_34 (Springer, 2020).

25. Haghani, M. Empirical methods in pedestrian, crowd and evacuation dynamics: Part ii. field methods and controversial topics. *Safety science* **129**, 104760, <https://doi.org/10.1016/j.ssci.2020.104760> (2020).
26. Korbmacher, R. & Tordeux, A. Review of pedestrian trajectory prediction methods: Comparing deep learning and knowledge-based approaches. *IEEE Transactions on Intelligent Transportation Systems* **23**, 24126–24144, <https://doi.org/10.1109/TITS.2022.3205676> (2022).
27. Le Progrès. Fête des lumières 2022: Plus de 2 millions de visiteurs cette année. *Le Progrès weblink* (2022).
28. Zuili, T. Fête des lumières: Colis suspects, foule, pannes... comment se prépare TCL à lyon. *ActuLyon weblink* (2023).
29. Liang, H., Lee, S., Sun, J. & Wong, S. Unraveling the causes of the seoul halloween crowd-crush disaster. *PLoS one* **19**, e0306764, <https://doi.org/10.1371/journal.pone.0306764> (2024).
30. Dufour, O. et al. Dense Crowd Dynamics and Pedestrian Trajectories: A Multiscale Field Study at the Fête des Lumières in Lyon, <https://doi.org/10.5281/zenodo.13830435> (2024).
31. Boltes, M. & Seyfried, A. Collecting pedestrian trajectories. *Neurocomputing* **100**, 127–133, <https://doi.org/10.1016/j.neucom.2012.01.036> (2013).
32. Boltes, M. et al. PeTrack. *Zenodo*, <https://doi.org/10.5281/zenodo.6320753> (2022).
33. Boltes, M., Seyfried, A., Steffen, B. & Schadschneider, A. Automatic extraction of pedestrian trajectories from video recordings. In *Pedestrian and evacuation dynamics 2008*, 43–54, https://doi.org/10.1007/978-3-642-04504-2_3 (Springer, 2010).
34. Johansson, A., Helbing, D., Al-Abideen, H. Z. & Al-Bosta, S. From crowd dynamics to crowd safety: a video-based analysis. *Advances in Complex Systems* **11**, 497–527, <https://doi.org/10.1142/S0219525908001854> (2008).
35. Chraïbi, M. & Dufour, O. MADRAS-data-app, <https://doi.org/10.5281/zenodo.10694867>, <https://go.fzj.de/madras-app> (2024).
36. Schrödter, T. & The PedPy Development Team. Pedestriandynamics/pedpy: v1.0.2, <https://doi.org/10.5281/zenodo.10016938> (2023).
37. Largeview 08 12 2022 20h26, <https://www.youtube.com/watch?v=jIC3dNOZsk0>.

Acknowledgements

We acknowledge the precious help of Clément Albin, David Rodney, and Nicolas Verstaev, as well as that of the technical municipal services of Lyon and organizers of the Fête des Lumières (Christophe Doucet, Thierry Courtot, Loïc Malacher, Antony Sabater). We are also grateful to Maik Boltes and Alessandro Corbetta for their advice on the choice of hardware, to Tobias Schrödter for his support in using *PeTrack* and to Alexandre Martel and Anthony Boueri for helping us collect the videos and surveys. The authors acknowledge the Franco-German research project MADRAS, funded in France by the Agence Nationale de la Recherche (ANR, French National Research Agency), grant number ANR-20-CE92-0033, and in Germany by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), grant number 446168800.

Author contributions

Each team member played a significant role in various aspects of the project, including data collection, analysis, writing, and tool development. All authors reviewed the manuscript, participated in data acquisition, and ensured the accuracy of the collected data. Below is a summary of the contributions: O.D.: Organized data collection in Lyon, tracked *LargeView* trajectories manually, collected and analyzed contact numbers, GPS tracks, and *LargeView* trajectories, contributed to the Festival of Lights description and gathering methodology, calibrated *LargeView* cameras and contributed to building an interactive app for data exploration. H.D.: Calibrated the *TopView* cameras, extracted *TopView* trajectories using *PeTrack*, and contributed to the gathering methodology. J.C.: Selected time sequences and zones for analysis, extracted trajectories using *PeTrack*, and contributed to the gathering methodology. R.K.: Participated in data acquisition, extracted *TopView* trajectories using *PeTrack* and contributed to the literature survey and gathering methodology. M.C.: Developed an interactive app for data exploration, extracted *TopView* trajectories using *PeTrack*, and contributed to the analysis of quantitative trajectory data. B.G.: Participated in the collection of field data, monitored pedestrian flows, identified singular phenomena, wrote part of the description of the context. A.N.: Organized data collection in Lyon, participated in data acquisition, tracked and analyzed *LargeView* trajectories, manually tracked heads in the square scene, and contributed to identifying statistics, and qualitative phenomena. A.T.: Contributed to the literature survey, extracted *TopView* trajectories using *PeTrack*, quantitative data analysis, technical validation, and indicators/measures formulae development.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors have interacted with the organizers of the Festival of Lights, but have not received any funding from them. They are hierarchically and financially independent from them and are not aware of any other competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-025-04732-3>.

Correspondence and requests for materials should be addressed to O.D., M.C., B.G., A.N. or A.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025