



State-space models reveal bursty movement behaviour of dance event visitors

Philip Rutten^{1*}, Michael H. Lees¹, Sander Klous¹ and Peter M.A. Sloot^{1,2} 

*Correspondence: p.rutten@uva.nl

¹Informatics Institute, University of Amsterdam, Science Park 904, 1098XH Amsterdam, The Netherlands

Full list of author information is available at the end of the article

Abstract

Pedestrian movements during large crowded events naturally consist of different modes of movement behaviour. Despite its importance for understanding crowd dynamics, intermittent movement behaviour is an aspect missing in the existing crowd behaviour literature. Here we analyse movement data generated from nearly 600 Wi-Fi sensors during large entertainment events in the Johan Cruijff ArenA football stadium in Amsterdam. We use the state-space modeling framework to investigate intermittent motion patterns. Movement models from the field of movement ecology are used to analyse individual pedestrian movement. Joint estimation of multiple movement tracks allows us to investigate statistical properties of measured movement metrics. We show that behavioural switching is not independent of external events, and the probability of being in one of the behavioural states changes over time. In addition, we show that the distribution of waiting times deviates from the exponential and is best fit by a heavy-tailed distribution. The heavy-tailed waiting times are indicative of bursty movement dynamics, which are here for the first time shown to characterise pedestrian movements in dense crowds. Bursty crowd behaviour has important implications for various diffusion-related processes, such as the spreading of infectious diseases.

Keywords: Human mobility; Wi-Fi data; State-space models; Movement ecology; Bursty dynamics

1 Introduction

Modeling the dynamics of pedestrians in large crowds is important for a number of reasons, from understanding how dangerous situations arise from individual behaviours [1–4], to predicting the spread of epidemic diseases [5–7]. So far, existing models of crowd dynamics are typically used in scenarios in which individual driving forces such as ‘desired velocity’ are considered constant [8, 9]. In reality however, desired speed and direction are conditional on an internal behavioural state. Humans generally switch between different modes of movement behaviour [10]. People stay in one place for some time, and then decide to change location, usually in one continuous movement bout. This kind of intermittent movement behaviour also typically occurs during large crowded events that span long time periods [11]. Despite the fact that crowd management is most critical for these events, intermittent movement behaviour is an aspect that is still missing in the ex-

© The Author(s) 2021. 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/>.

isting crowd dynamics literature. While the research literature on human mobility has grown extensively in the last two decades, there are few empirical results that pertain to the relative timing of behavioural events in motion patterns [12–15]. However, without exception these studies focus on mobility scales ranging from intra-urban to inter-urban, involving various transportation modalities other than walking (see [15] for an overview). This leaves open the question of how to correctly characterise the natural intermittency in pedestrian movement patterns.

Intermittent movement behaviour is something humans have in common with many other animals. Intermittent motion is a general characteristic of animal locomotion and animal foraging strategies [16, 17]. Intermittency has been the focus of many studies on animal search behaviour, and has invited various mathematical descriptions, mostly within the random walk framework [18, 19]. Meanwhile, there has been a growing interest of ecologists to go beyond the phenomenological level of description inherent to using simple random walk models [20]. Ecologists have looked for models that can capture more of the complexity and heterogeneity of animal movement patterns. As a result, there has been a growing number of studies assuming that animal movement consists of multiple (usually two) behavioural modes [21]. These studies start from the assumption that model parameters vary over time and are conditional on behavioural state. These studies then explicitly infer behavioural states from movement data by incorporating switching behaviour in the statistical models. These studies are based on using hidden Markov models (HMMs) or more general state-space models (SSMs) (see [21] for an overview). SSMs are the most natural candidate in the presence of non-negligible measurement error, which is the case for the research presented in this paper. Therefore, we focus on SSMs in what follows. SSMs allow the combination of a mechanistic movement model with an observation model involving measurement error, in one statistical data fitting procedure. Moreover, SSMs can be used to simultaneously estimate probabilities of movement modes, spatial locations, model parameters and measurement errors [22]. Existing methods in the context of pedestrian mobility are tailored on GPS data, which are usually very precise [10, 13]. SSMs provide flexibility in handling measurement errors, which allows us to exploit other types of data, such as Wi-Fi detections in indoor settings.

Here we use the state-space modeling framework to analyse temporal aspects of pedestrian movement in the Johan Cruijff ArenA football stadium in Amsterdam. We use localization of smart phones based on Wi-Fi detections to reconstruct individual trajectories, and use this as a proxy for human movement. We use movement models from the field of movement ecology to describe individual pedestrian movement. The joint estimation of multiple movement tracks allows us to investigate statistical properties of estimated behavioural state sequences. More specifically, we investigate whether waiting times are exponentially distributed in time, or whether they follow heavy-tailed distributions. These heavy-tailed waiting times are indicative of ‘bursty’ behaviour. Although bursty dynamics have been shown to characterise several human and animal activities, including movement behaviour (see [23] for an overview), there is no previous empirical evidence of burstiness in pedestrian movements in dense crowds.

To explain bursty dynamics from a common behavioral point of view, a priority list model was proposed by Barabási (2005) [24]. The underlying assumption of this model is that human behaviour is the result of a decision-making process based on individual priorities. To see whether the observed waiting times could be possibly explained by such

a model, we need to establish whether the observed switching behaviour is the result of individual decisions, or whether the switching behaviour is driven by external factors. To do so, we analyse two events, namely a major league football match, and a large dance event with DJ show. These two events represent different degrees of constrained movement. We consider a football match as a highly constrained system: during the episodes of the match people sit and watch, and after the match they leave the building. We use a simple method to quantify this observation and use the result for comparison to the dance event, on which we will focus our attention. Tickets for this event did not allocate seats, and people were free to walk around the stadium, including the pitch, where the DJ stage was located. The event lasted more than 6 hours, and the only important external drivers were the start and end of the DJ show, which lasted approximately 4 hours. Therefore, we consider this event to a much lesser extent a constrained system. We expect that this movement better corresponds to individual decisions about when (and where) to move.

We quantitatively assess these assumptions, and investigate the statistical properties of the observed intermittent movement behaviour. In doing so, we uncover some surprising aspects of pedestrian crowd motion, such as the resemblance with the motion patterns of various monkeys.

The rest of the paper is organized as follows. In Sect. 2 we describe how we translate Wi-Fi measurements into a collection of movement tracks. We introduce the methods we use to quantitatively assess results. We present the movement models used in the state-space framework, and aspects of the model fitting procedures. In Sect. 3 we describe the statistical results of our analysis, and discuss the findings in Sect. 4.

2 Methods

2.1 Data

We analyse data collected by the Wi-Fi network in the Johan Cruijff Arena stadium in Amsterdam. The wireless network consists of 591 access points (APs) with known spatial coordinates. The network is designed for complete coverage in the stadium. Though APs are not distributed homogeneously, AP locations are chosen to maximise coverage given the unique structure of the building and its accessible areas.

At a constant frequency (1 Hz) the APs switch to ‘monitor mode’ and capture all wireless traffic, regardless of destination addresses. The APs send reports of the monitoring results to a server where data are extracted, anonymised, and stored. The data contain the following information relevant for our research: the identity of the AP, the anonymised identity of source devices, received signal strengths (RSS) values, and timestamp indicating time of measurement.

The data used in this paper were collected during two large events. The first was the Armin van Buuren dance event with DJ show, in May 2017. The second is the major league football match Ajax–Feyenoord, in October 2019. The data sets typically contain detections of tens of thousands unique MAC addresses. However, many devices have only a few detections that are spread out in time. When someone is not using his/her smart phone, the device eventually pauses all wireless communication. This drastically reduces the number of devices that allow tracking. To prepare the data for movement track reconstruction, we search for devices with detection periods without large gaps. Detection gaps range from small (seconds, minutes) to large (e.g. >2 hours). We select data from devices with detection periods of minimum length `minperiod`, containing time gaps that

do not exceed a maximum length maxgap . Here, we use ($\text{minperiod} = 5$, $\text{maxgap} = 1$) minutes. After applying this criterion we have, for each selected device, one or several detection periods ranging from 5 minutes (minperiod) to several hours, that are separated in time at least 1 minute (maxgap).

We estimate locations of smart phones using proximity detection, which determines the position of a device based on its closeness to an AP with known spatial coordinates [25]. First, we discretize all the data in time intervals $\Delta t = 10$ seconds. Then we group, for each device, and per time interval, the set of APs at which the device was detected. We select the AP with the strongest received signal strength (RSS) value in the time interval as the one being closest. The proximity detection method produces temporal sequences of APs which are estimated to have been near a device. We ignore the z -coordinate and simplify the analysis to two-dimensions.

During crowded events, noise can reach levels that lead to considerable distortion of the distribution of RSS values over the APs. This is problematic for the proximity detection method, which simply selects the AP with the single highest RSS value in the time interval. As a result, AP sequences contain random fluctuation: jumps back and forth between APs that do not reflect real movement. This can also happen when a device is completely stationary, leading to oscillating behaviour between two or more APs. These oscillations may occur between APs which are not nearest neighbours, but are at distances that would require unrealistic velocities. To deal with this problem we use a simple moving average to smooth the movement tracks (see Appendix A for more details).

Drawing a line between successive smoothed location estimates produces a trajectory, or movement track. In Fig. 1 we show typical examples of movement tracks of both the football match Ajax–Feyenoord (a), (c), and the Armin van Buuren dance event (b), (d). We see that for the football match, the movement tracks mostly stay in one location, and have a larger displacement only at ingress/egress. These movement tracks have a rather predictable character. For the dance event on the other hand, we see a pattern of multiple waiting times at different locations, and with variable duration. The examples demonstrate why we focus our attention on the dance event, and use the football match only for comparison.

The accuracy of the proximity detection method is low, and we do not pretend to track individuals with high accuracy. We use the positioning results to study temporal aspects of the motion only, that are therefore not strictly dependent on the exact locations (see Appendix A for more details).

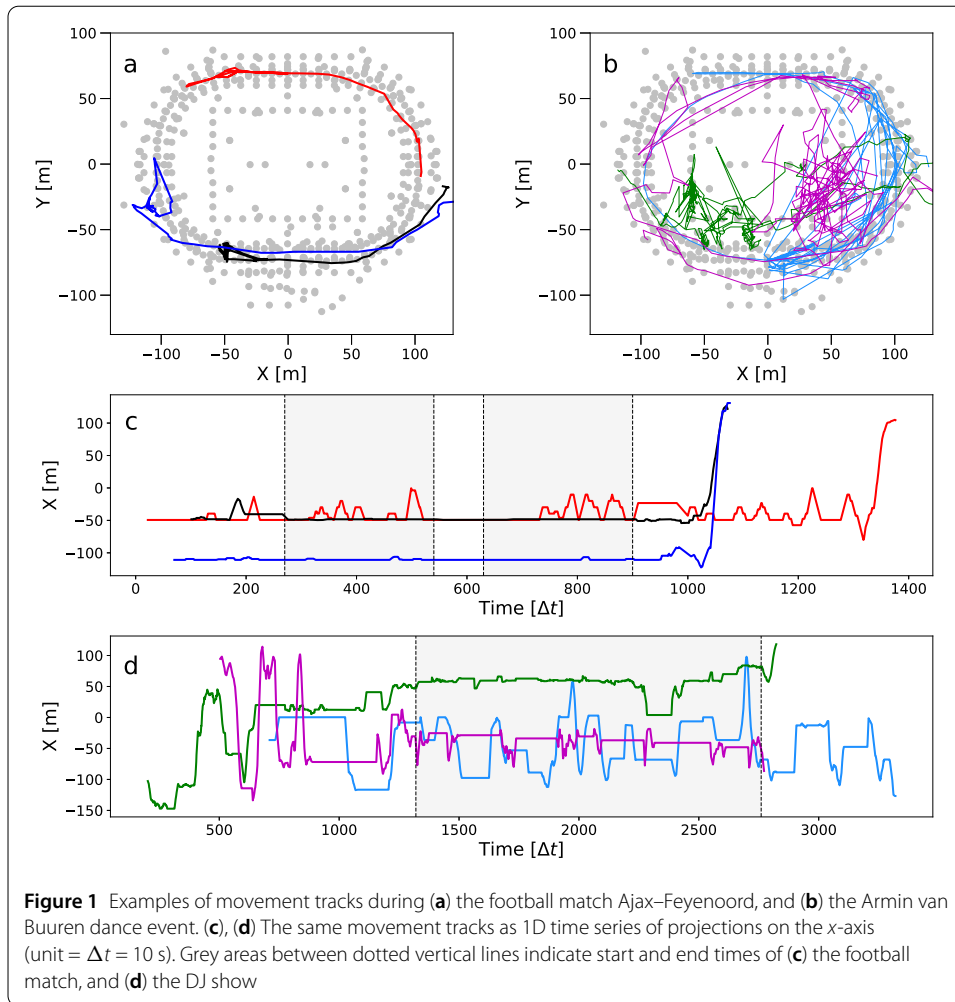
2.2 Measuring movement

We wish to establish when people are moving during the events. As a simple measure for quantifying movement, we use the radius of gyration (as in [26]). The radius of gyration is defined as

$$r_{g,t} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\mathbf{x}_t - \mathbf{x}_{\text{com}})^2}, \quad (1)$$

where the center-of-mass of the trajectory, \mathbf{x}_{com} is

$$\mathbf{x}_{\text{com}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i. \quad (2)$$



To make the radius of gyration sensitive to changes in speed (due to relatively short movement episodes), we apply it within a two-sided moving time window, called moving radius of gyration (MRG) hereafter. In Equation (2) the window size $n = 2q + 1$, where q a nonnegative integer.

In Fig. 2 we show an example of a movement track, decomposed in 1D projections onto the x - and y -axes, together with the MRG. We see that the MRG responds well to changes in speed.

As an indicator of the total amount of movement in the stadium at a given time, we average the MRGs over multiple movement tracks. We can use the averaged MRGs to assess the results of the SSM analysis *indirectly*. Unfortunately no objective truth data are available to validate the SSM results. However, for the football match we have a rather strong intuition of what the total amount of movement in the stadium must have been (people sitting during the match, and moving before and after). Thus we can ‘calibrate’ the averaged MRGs using the football match as a benchmark. Using the calibrated averaged MRGs we can gauge the amount of movement during the dance event. In addition, we can assess the SSM results by comparing it to the average MRGs.

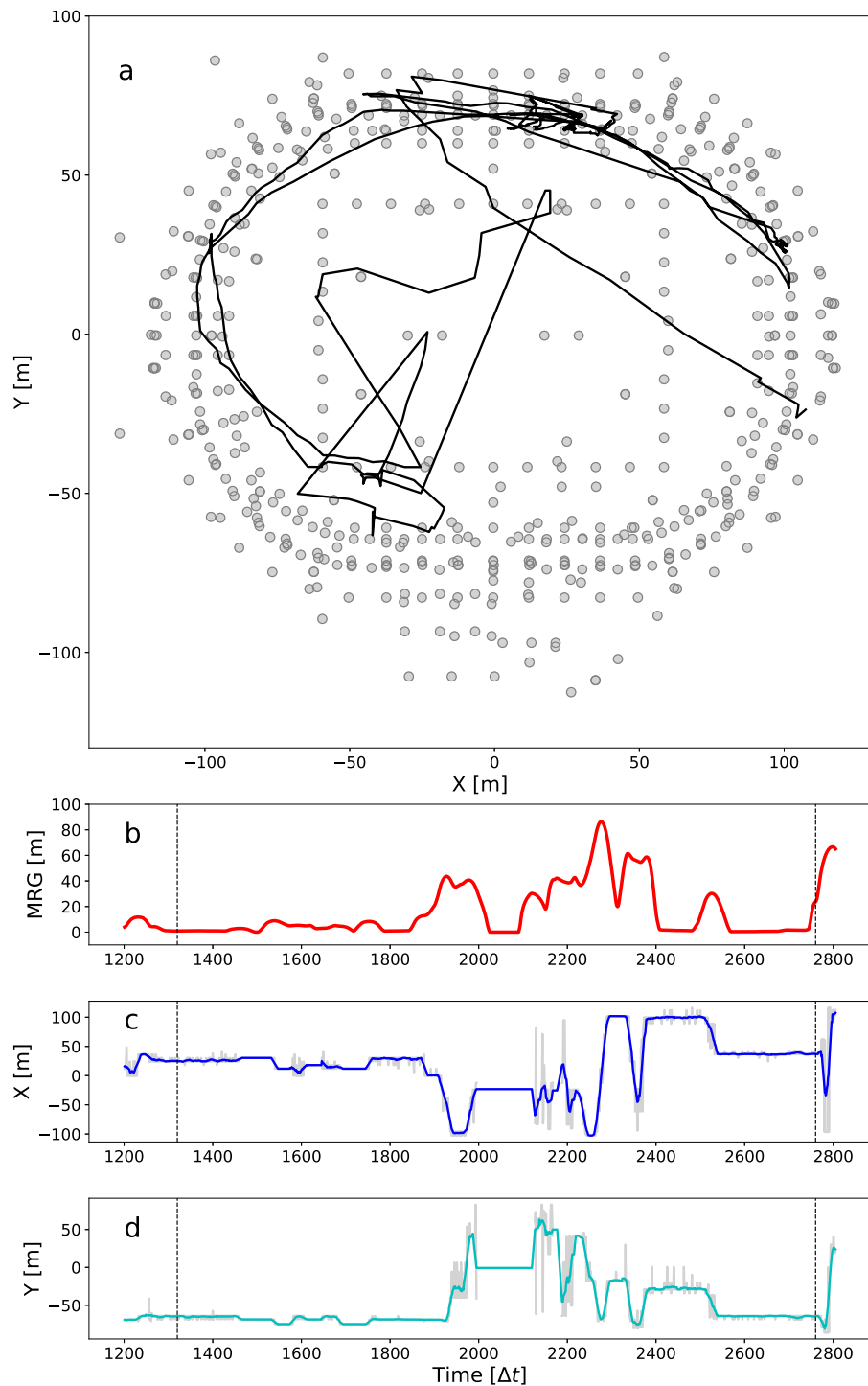


Figure 2 (a) Example of a movement track during the Armin van Buuren dance event (black line). (b) Moving radius of gyration (MRG) of the same movement track (red line) (unit = $\Delta t = 10$ s). (c), (d) Decomposition in x- and y-coordinate time series (blue and cyan lines), together with the original AP sequences produced by the proximity detection method (grey lines). The movement track starts at time bin 1201, which corresponds to 8:40 PM. Dotted vertical lines indicate start and end times of the DJ show

2.3 Device selection

We restrict our analyses to movement tracks that fulfill certain requirements. First we select devices whose detection periods span the duration of the event. This allows us to track the movement behaviour in relation to the various parts of the event programs (e.g. start, break, etc.). To this end, we select devices that were detected in both the time periods before the start, and after the end of the main show (see Appendix A for more details). For the football match the main show is the match, while for the dance event this was the DJ show. After this selection criterion we have devices with movement tracks with a total time length in the range 1.5 to 4 hours for the football match, and a total time length in the range 4 hours to 9 h 40 min for the dance event.

The movement track of each device typically consists of multiple detection periods of minimum length $\text{minperiod} = 5$ minutes, that are separated in time at least $\text{maxgap} = 1$ minute (see Sect. 2.1). To restrict our attention to movement tracks with minimum gap length, we select devices that were detected at least in 1/3 of the time intervals between their first and last detection (see Appendix A for more details).

To filter out devices that did not move at all during the whole event, we select movement tracks for which the MRG exceeds a threshold value at least once. We use an intuitively chosen threshold value ($r_{g,t} > 10$ m) which is large enough to include only tracks for which movement is unambiguous.

After filtering according to these criteria we have movement tracks of 361 devices of the football match, and 1048 devices of the dance event. We sample 320 devices from the 1048, which we use for joint estimation (see Sect. 2.4.2). The devices are sampled in proportion to 5-quantiles in the frequency distribution of the observed maximum of $r_{g,t}$ over the devices.

To prepare the movement tracks for fitting the state-space models using MCMC inference, we merge the separate detection periods into one movement track. For each track, we simply close detection gaps and concatenate the detection periods, but keep appropriate time indices in a parallel sequence. The SSM will interpret the data as one continuous movement track.

2.4 Movement models

Movement tracks of selected devices are characterised by an intermittent movement pattern. Periods of rest alternate with episodes of movement that form larger displacements. This aspect of the movement data shows up clearly in the 1D projections of the movement tracks onto the x - and y -axes (see Figs. 1 and 2 for examples). Therefore, it seems reasonable to assume a movement process consisting of two discrete behavioural states. One state consists of either no movement or small-scale movements with frequent reversals, due to an individual ‘staying in one place’, and a second state consists of faster and more directionally persistent movement, due to an individual ‘going somewhere’.

We apply two different SSMs. For an initial exploration, we are primarily interested in exploring the state-switching behaviour. Therefore, we first apply a SSM based on the most simple movement model, which is a 2D random walk. We use the first model for joint estimation of multiple movement tracks and study statistical properties of the results. The second model is more detailed and takes into account some characteristic properties of pedestrian movements. We use the second model to study specific aspects of the motion in more detail.

2.4.1 Model 1: conditionally Gaussian linear SSM

The first model is one of the simplest state-space models, which is obtained by adding measurement noise to a random walk [27]. It has however the added complication that the hidden state is composed of a finite-valued discrete variable, and a continuous (vector-valued) variable. In the literature, this model is called a conditionally Gaussian linear state-space model (CGLSSM) [28]. The model is defined as

$$\mathbf{z}_t = \mathbf{z}_{t-1} + \eta_t, \quad \eta_t \sim N(0, \mathbf{R}(S_t)), \quad (3)$$

$$\mathbf{y}_t = \mathbf{z}_t + \epsilon_t, \quad \epsilon_t \sim N(0, \mathbf{Q}), \quad (4)$$

where \mathbf{z}_t is the true location, \mathbf{y}_t the observed location, η_t and ϵ_t are both drawn from bivariate Gaussian distributions, \mathbf{Q} is the measurement error variance, and the process variance $\mathbf{R}(S_t)$ is conditional on the discrete state variable S_t at time t . We assume that the state $S_t \in \{1, 2\}$ follows a Markov chain, and introduce the probabilities $\alpha_1 = P(S_t = 1 | S_{t-1} = 1)$ and $\alpha_2 = P(S_t = 1 | S_{t-1} = 2)$. The full transition probability matrix is defined as

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 & 1 - \alpha_1 \\ \alpha_2 & 1 - \alpha_2 \end{bmatrix}. \quad (5)$$

The complete hidden state, to be inferred from the data, is the compound $\mathbf{x}_t = (S_t, \mathbf{z}_t)$. Note that we use the model here for ‘change point detection’, i.e. to find the time-points where the switches occur [28]. The state-space analysis effectively divides the movement tracks into segments based on behaviour, and the SSM functions as a path segmentation algorithm.

2.4.2 Joint estimation

We use the CGLSSM for the joint estimation of multiple movement tracks. This usually requires defining a hierarchical model including common prior distributions from which individual parameters are drawn [29, 30]. Here we are primarily interested in estimating the behavioural states and not in estimating movement parameters. Note that the movement model is chosen for its simplicity and not for its accurate description of the movement process itself. As shown by Jonsen (2016) [31], if the analysis is not focused on estimating movement parameters, it is useful to assume that individuals share identical parameters. In fact, a simple hierarchical model without common prior distributions, used for joint estimation, improves inference of behavioural states. Therefore we simply assume that parameters are identical among individuals and fit the CGLSSM (Equations (3) and (4)) as a joint estimation model.

We do not fit a single model to all movement tracks at once. We divided 320 tracks into 40 independent sets, each containing 8 tracks. Each set is then fit separately (as in [32]). However, statistical quantities (introduced in the next sections) are derived from aggregating results from all the sets. Our strategy thus chooses a trade-off between the benefits of joint estimation, and fitting individual movement tracks separately, in a non-hierarchical manner.

2.4.3 Analysis of state sequences

The SSM produces sequences of estimated behavioural states. We re-allocate state sequences to their original locations on the timeline of the event.

The joint estimation of multiple movement tracks allows us to investigate statistical properties of the estimated behavioural state sequences. We first count for each time step t_i the number of individuals N_m estimated to be in the movement state, and calculate the fraction of movers out of the total number of estimated individuals $f_i = N_m/N_{\text{tot}}$. We then compare this result with the averaged MRGs, using the Pearson correlation coefficient between the two time series. If the behavioural states are estimated correctly, we expect the fraction of movers to evolve similarly as the averaged MRGs.

The joint estimation of multiple movement tracks allows us to measure waiting times. Waiting times t are defined as the time an individual stays in the same location, and are measured as the number of consecutive time intervals a device is in the non-moving behavioural state. In the context of mobility studies this time is sometimes referred to as an ‘inter-event time’, where an event is a displacement [23]. This approach divides each detection period in observed inter-event times, and so-called residual times, which are the waiting times truncated by the finite observation time window (e.g. see [33]). The underlying assumption of this approach is that we measure the timing of events in a stationary stochastic process, on which we randomly place an observation window. This is not the case here, as the movement process (in the stadium) naturally starts and ends within the observation window. The size and position of the detection periods are determined by other mechanisms (e.g. someone using his/her smart phone). Therefore, in our case, all detected waiting times have equal statistical importance. To avoid confusion, we refer to waiting times, as is common in the random walk framework [14].

Now several possibilities arise for measuring the waiting times in relation to the detection gaps. A conservative approach (strategy A) considers only waiting times estimated from the observed detection periods. However, it is interesting to explore a less conservative approach (strategy B). If a device is at the same location before and after a detection gap, we might assume that the device has not moved during the gap. Therefore, if the device is in a waiting time before and after the detection gap we can assume it spent the time in the detection gap also as a waiting time. Thus, if the SSM estimates a waiting time which extends across the detection gap, we assume it didn’t move during the gap. We measure the length of the whole episode which includes the waiting times at both ends of the gap, and the gap itself. On the other hand, if a device has moved during the detection gap, the SSM automatically inserts behavioural states of the movement class (to make the jump). The waiting times at both ends of the detection gap (but at different locations) are measured separately. We explore a third option (strategy C), which takes the level of speculation further and fully restores the movement tracks in the detection gaps. To do so, we deploy the interpolation method used by Rhee et al. (2011) [13], which works as follows. From the last location where a device was detected we assume that the individual walks to the next location (i.e. where the device is detected again) at a walking speed of $1 \text{ m}\cdot\text{s}^{-1}$. The time this takes is subtracted from the detection gap, and the remaining time is added as a waiting time after the last detection (i.e. the time interval where the device ‘disappeared’). Clearly, the insertion of waiting times is expected to impact the distribution of waiting times. We compare results of the strategies A, B, and C (see Appendix A for more details).

The state-space modeling approach assumes that the switching behaviour between states is a Markov process. This implies that at each time interval the probability of switching is a Bernoulli random variable with probability α_{ij} . The expected time spent in state i is distributed as a geometric random variable with mean $1/\alpha_{ij}$. We therefore expect the pe-

riods of time spent in one state (before switching), to follow a Poisson distribution. To test this assumption we fit statistical distributions to the measured waiting times, using maximum likelihood estimation (MLE) methods [34–36] (see Appendix C for more details). Specifically, we determine whether the data are better fit by:

- The exponential distribution, with probability density function defined as:

$$p(t) = \lambda \exp(-\lambda(t - a)). \quad (6)$$

- The truncated power law distribution distribution, with probability density function defined as:

$$p(t) = (t + a)^{-\alpha} e^{-t/t_c}. \quad (7)$$

- The stretched exponential distribution, with probability density function defined as:

$$p(t) = \beta \lambda t^{\beta-1} \exp(-\lambda(t^\beta - a^\beta)). \quad (8)$$

- The log-normal distribution, with probability density function defined as:

$$p(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log t - \mu)^2}{2\sigma^2}\right), \quad (9)$$

where t_c is the exponential cutoff value, and a is the lower bound of the fitting range. While Poisson processes are characterised by exponentially distributed inter-event times, there has been a number of reports showing evidence that waiting times in the movement patterns of various animals (e.g. [37, 38]), and also humans [13], follow power law distributions. In our case, the waiting times are limited by the duration of the event and can only be reasonably identified as a truncated power law. The stretched exponential and log-normal distributions are models commonly used to describe heavy-tailed phenomena in complex systems [23]. We select the most appropriate model using the model selection method based on Akaike's information criterion (AIC) [39] (see Appendix C).

2.4.4 Model 2: switching first-difference CRW

Although the first model is computationally convenient, which allows us to analyse multiple movement tracks, it is interesting to add more realism to the movement model. Animals usually tend to keep on moving in the same direction [40]. Visual inspection of the movement tracks reveals that during movement episodes individuals are indeed moving with persistence [11]. In random walk models, persistence in direction is expressed through autocorrelation in the relative turning angles between successive movement steps. Such behaviour can be modeled using a correlated random walk (CRW) [41]. We adopt a two-state switching CRW model (DCRWS) which was introduced by Jonsen et al. (2005) [42]. The model is a random walk on the differences in consecutive locations $\mathbf{d}_t = \mathbf{z}_t - \mathbf{z}_{t-1}$. The movement model is defined as

$$\mathbf{d}_t = \gamma \mathbf{T} \mathbf{d}_{t-1} + \eta_t, \quad \eta_t \sim N(0, \mathbf{R}), \quad (10)$$

$$\mathbf{y}_t = \mathbf{z}_t + \epsilon_t, \quad \epsilon_t \sim N(0, \mathbf{Q}), \quad (11)$$

where \mathbf{z}_t , \mathbf{y}_t , η_t , ϵ_t , and \mathbf{Q} are similarly defined as for the CGLSSM (Equations (3) and (4)), but where we have removed the dependency of \mathbf{R} on state S_t . The parameter γ defines the persistence ($0 < \gamma < 1$), and \mathbf{T} is a transition matrix defining the rotational component of the correlated random walk:

$$\mathbf{T} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}. \quad (12)$$

In this case, the parameters for persistence γ , and mean turning angle θ , are conditional on the behavioural state S_t (and actually are $\gamma(S_t)$ and $\theta(S_t)$). For the state variable S_t we follow the same assumptions as the CGLSSM defined above (Equations (3) and (4)). The DCRWS is in fact a first-order autoregressive (AR(1)) process of the displacements and a discrete version of Langevin-like equations such as those defining the *social force* model [43, 44].

In both CGLSSM and DCRWS we assume measurement errors are normally distributed, and use a fixed covariance matrix

$$\mathbf{Q} = \begin{bmatrix} 6 & 0 \\ 0 & 6 \end{bmatrix}. \quad (13)$$

The proximity sensing method based on Wi-Fi measurements produces a heterogeneous error structure, presumably correlated to device location within the building. We assume that the smoothing of the movement tracks alleviates this problem (see Sect. 2.1), but the estimated locations still contain error. Therefore, we would like to further smooth the movement tracks, and avoid ‘overfitting’ the movement tracks to random fluctuations that are still present in the smoothed data. This overfitting occurs when we let the MCMC method estimate the values of both \mathbf{R} and \mathbf{Q} simultaneously. The MCMC method is free to converge to any arbitrary set of values for \mathbf{R} and \mathbf{Q} that optimize the fit of the posterior distribution. Using fixed values for the measurement error effectuates further smoothing of the movement tracks, and also reduces the number of parameters to be estimated. In absence of knowledge about the true distribution of the errors (after applying the moving average), we use standard, zero-mean, Gaussian noise. To estimate error covariance values σ_x^2 and σ_y^2 , we run a second pass of the moving average filter over the tracks. Error values are then estimated by the residuals after subtracting the second smoothed result from the first (see Appendix A for more details).

2.5 Implementation

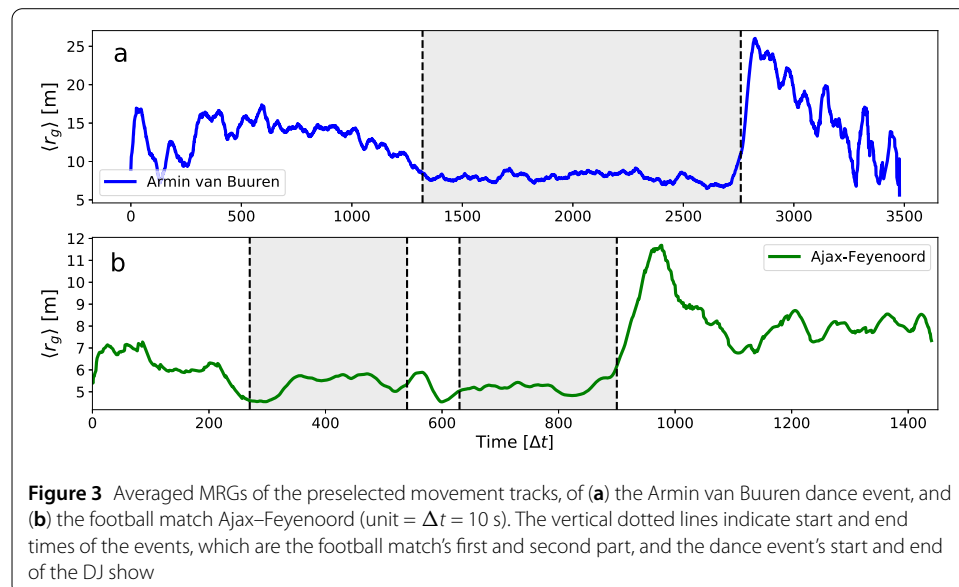
The models are fit using a fully Bayesian approach to the inference of hidden states and model parameters, based on Markov chain Monte Carlo (MCMC) sampling. For the implementation of the MCMC sampling we use the JAGS software [45] and its R interface `rjags` [46]. For both models, two Markov chains are run, from which the first 1000 samples of each chain are discarded as a burn-in. For the CGLSSM, posterior inference of the unobserved variables, S_t and \mathbf{z}_t , as well as model parameters, α and \mathbf{R} , is performed from 1000 samples per chain. For the DCRWS, the posterior inference of variables, S_t and \mathbf{z}_t , and parameters γ , θ , α , and \mathbf{R} , is performed from 5000 samples per chain after thinning by a factor of 5 to reduce within-chain sample autocorrelation, yielding a final 1000 samples from the joint posterior. For more details on prior distributions and convergence diagnostics see Appendix B.

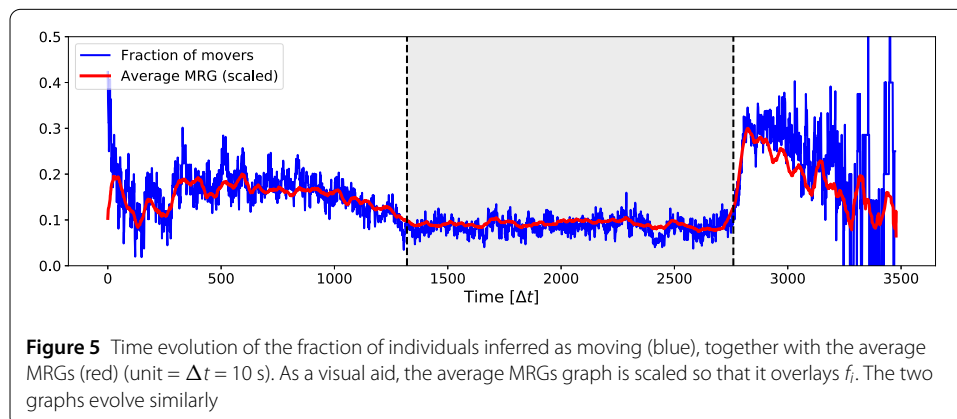
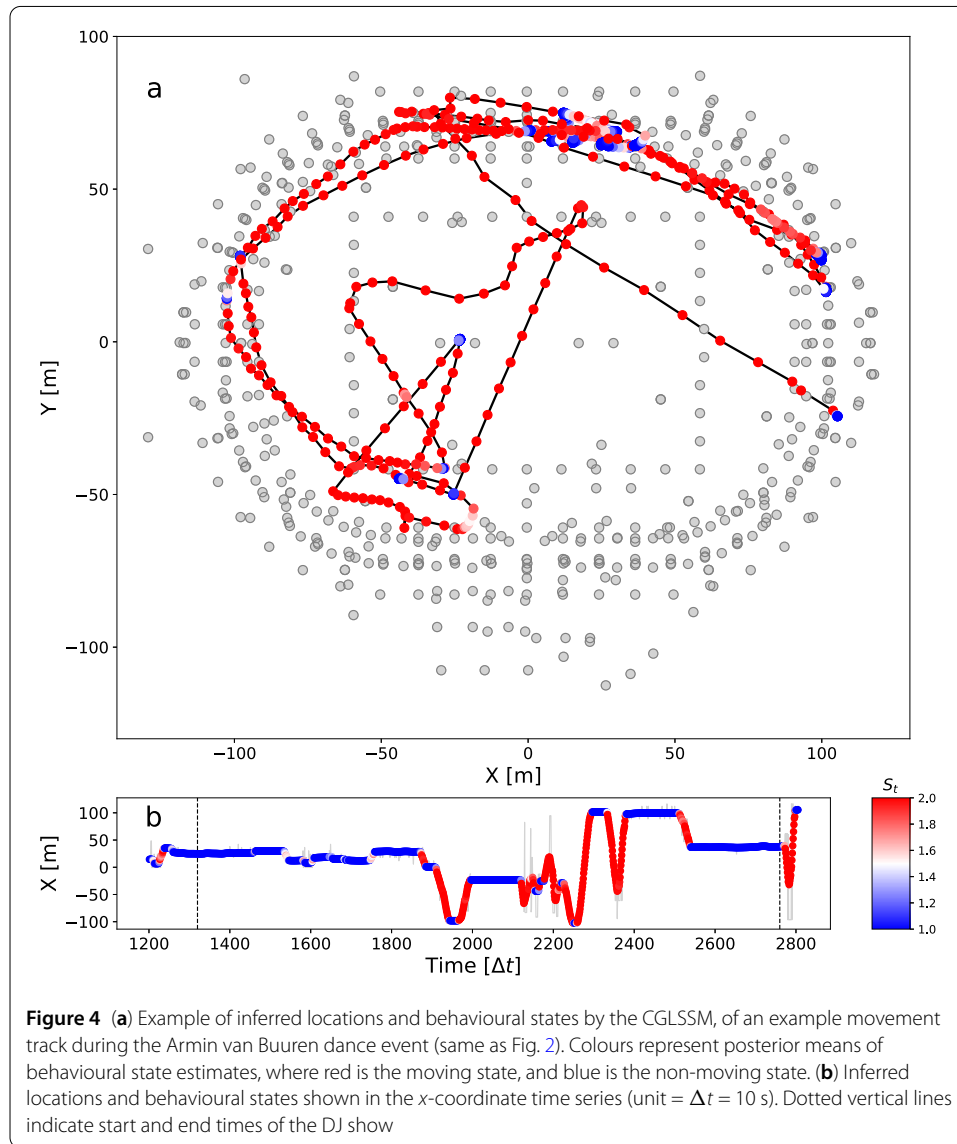
3 Results

In Fig. 3 we show the MRGs averaged over the preselected movement tracks for both the football match (Ajax–Feyenoord) and the dance event (Armin van Buuren). The vertical dotted lines indicate start and end times of the events, which are the football match’s first and second part, and the dance event’s start and end of the DJ show. We see that in both cases the average MRGs clearly decrease during the events, and then peak afterwards, indicating people collectively leaving the stadium after the events. Although the overall shapes of the two graphs are similar, in the case of the dance event it is stretched over a much longer period of time. For the dance event we show data from 17:20 to 03:00 next morning. Doors were open from 18:00, and the DJ show started at 21:00 and ended at 01:00. For the football match we show data from 16:00 to 20:00. The match started at 16:45 and consists of two 45 minutes halves, with a 15 minute break in between. The pre-show time period of the dance event lasts 3 hours, while the time period before the start of the football match, shown in Fig. 3, is 45 minutes. Thus, the average MRG maintains high levels over much longer periods of time in case of the dance event.

To illustrate the utility of the state–space framework, we show results of fitting the CGLSSM to an example track from the Armin van Buuren data set in Fig. 4. Posterior means of behavioural states take values between 1 and 2 (moving and waiting respectively) and summarise behavioural state estimates. Locations are inferred as being in state 1 if the probability to be in the moving state is most likely, i.e. $S_t < 1.5$, and state 2 otherwise. In Fig. 4 we observe that the result is visually agreeable, i.e. the movement states seem correctly inferred, though we cannot objectively validate the result.

To explore the probability of being in the movement state, we look at the fraction of individuals inferred to be in the movement state $f_i = N_m/N_{\text{tot}}$. In Fig. 5 we show the time evolution of f_i together with the average MRGs. Note that the relative heights of the two graphs do not convey any meaningful information, and we should only compare the relative shapes. Therefore, as a visual aid, we have scaled the average MRGs so that it overlays f_i . We see that the two graphs evolve similarly; the Pearson correlation coefficient is 0.77. However, at the start and end of the time series of f_i the number of individuals N_{tot} be-





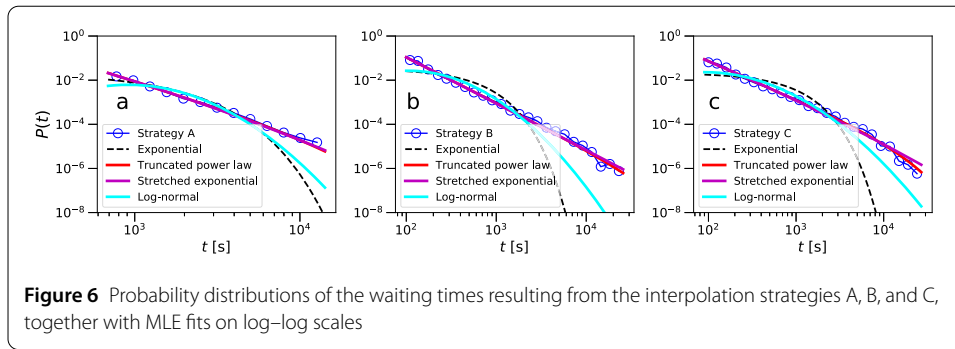


Table 1 Overview of MLE values for the 3 different strategies. Shown are the MLEs for the truncated power law, which provided the best model for the data, according to model selection based on Akaike weights

	t_{\min} (s)	α	t_c (s)
Strategy A	690	2.32	11,540
Strategy B	100	2.07	43,630
Strategy C	90	1.74	14,610

comes very small (e.g. $N_{\text{tot}} < 10$ in the last 200 time intervals), and the fluctuation in f_i becomes large. Therefore, if we exclude the first and last 200 time intervals from both time series the Pearson correlation coefficient increases to 0.9. These results suggest that the joint estimation of behavioural states agrees well with the average MRG during the event.

In Fig. 6 we show the empirical distributions of waiting times sampled according to the three different strategies on log–log scales. We also show the maximum likelihood estimates (MLE) of the exponential, truncated power law, stretched exponential, and log-normal distributions. For fitting the probability density functions we use lower cutoff values determined using the methods in [34]. We see that, contrary to expectation, the exponential distribution does not provide good descriptions of the waiting times. We select the most appropriate models using the model selection method based on Akaike’s information criterion (AIC) [39] (see Appendix C for details). According to the Akaike weights the truncated power law provides the best model for the resulting waiting times of all three strategies. For the waiting times of strategies B and C, the Akaike weight for the truncated power law $w_{\text{tpl}} = 1$. For strategy A, $w_{\text{tpl}} = 0.88$ and stretched exponential $w_{\text{se}} = 0.12$. In Table 1 we give an overview of MLEs of the parameters of the truncated power law. The estimated exponential cutoff values for strategy A, $t_c = 11,540$ s (~ 3 h 12 min) and particularly of strategy C, $t_c = 14,610$ s (~ 4 h 3 min) correspond to the duration of the DJ show (4 hours). In these cases, the power law scaling of the waiting times in the movement patterns does not extend beyond the duration of the show.

We fit the DCRWS (Equations (10) and (11)) to the example movement track from the Armin van Buuren dance event (same as Fig. 4). In Fig. 7 we show inferred locations and behavioural states by the DCRWS model.

We see that the DCRWS and CGLSMM produce similar behavioural state estimates for the example movement track. In Table 2 we show parameter estimates of fitting the DCRWS. For each parameter we show 95% intervals (quantiles) of the posterior probability. The values support the idea that the movement process consists of two discrete states.

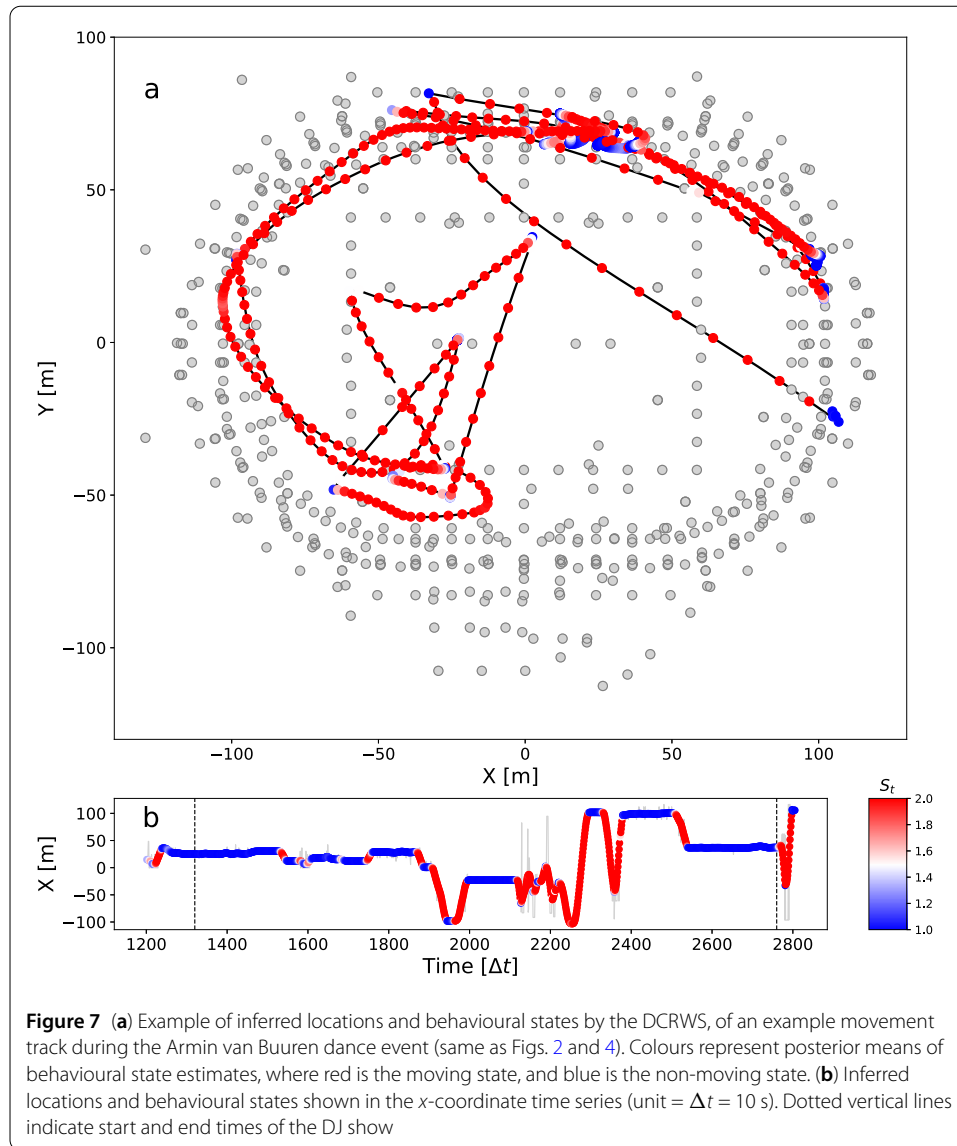


Table 2 Overview of 95% posterior intervals of the DCRWS parameter estimates

Parameter	0.025	0.5	0.975
γ_1	0.99	1.0	1.0
γ_2	0.24	0.36	0.48
θ_1	-0.02	0	0.01
θ_2	-2.93	-2.67	-2.42
α_1	0.90	0.93	0.96
α_2	0.01	0.02	0.04
σ_x	0.84	0.74	0.65
σ_y	0.71	0.63	0.57

The posterior intervals of γ and θ have no overlap and the medians are well separated. The median values $\gamma_1 = 1.0$ and $\gamma_2 = 0.36$ suggest the movement track can be described by two modes, where the first has perfect persistence, and the second is characterised by frequent changes in speed. The median value $\theta_1 = 0$ supports the turning angle is centered on zero in the first behavioural state, and $\theta_2 = -2.67$ suggest the turning angle is near $-\pi$ in

the second state, which indicates complete reversals. Estimates of the transition probabilities indicate a high probability of remaining in either the movement mode ($\alpha_1 = 0.93$), or 'staying in one place' ($1 - \alpha_2 = 0.98$). The small difference in values for process variance in x - and y -direction, $\sigma_x = 0.74$ and $\sigma_y = 0.63$, agrees with the elongated plan of the stadium in the x -direction.

4 Discussion

We present temporal aspects of the movement behaviour of visitors of large entertainment events in the Johan Cruijff ArenA football stadium in Amsterdam, using the state-space modeling framework. The fitting results of both the CGLSSM and the DCRWS suggest the movement tracks can be appropriately described as correlated random walks consisting of two discrete states. This supports the idea that the intermittent movement process is driven by switches between behavioural modes.

The time evolution of the fraction of individuals inferred as being in the movement state agrees with the time evolution of the averaged MRGs. This suggests the movement states inferred by the SSM analysis agree with our indirect measurement of the amount of movement during the event. Although individual movement tracks may appear idiosyncratic and unrelated to the event times (such as the example track in Fig. 2), collectively they do expose a pattern. The movement behaviour of individuals during the event remains unpredictable, and we cannot say exactly when a pedestrian decides to change its behaviour. This raises the possibility that the observed behaviour can be explained by a priority-based mechanism to decide between 'competing tasks' [24], such as enjoying the show, going to the bar, exploring the venue, i.e. different activities that translate to either movement or staying in one place. However, we have also shown that the switching behaviour is not independent of external events, and the probability of being in one of the behavioural states changes over time. During the DJ show people were more likely to stay in one place, while in the time before the show there was more movement. This time period may have been used to (casually) explore the venue or meet up with friends, etc. The end of the DJ show clearly marks the onset of a collective burst of movement, when visitors start leaving the building.

In addition, we investigated switching probabilities by looking at the observed waiting times. We have shown that empirical frequency distributions of waiting times deviate from the exponential and are best fit by heavy-tailed distributions. The estimated power law exponents decrease from interpolation strategy A to C, indicating an increase of large waiting times in the resulting distributions. This is not surprising as for strategies B and C we allow more speculative waiting times to interpolate the detection gaps. We conclude that real waiting times must be somewhere in between results of strategies A and C.

Bursty dynamics, as indicated by heavy-tailed inter-event times, have been found in various human activities [23, 24]. Heavy-tailed waiting times have been reported in human mobility captured by mobile phone traces [14], and the movement behaviour of pedestrians that make occasional use of other transport modalities such as bus or subway train [13]. Heavy-tailed waiting times characterise the motion patterns of various animals (see [47]), and have been explicitly reported for, notably, various monkeys [37, 38]. The estimated power law exponent $\alpha = 1.74$ of strategy C is closest to the result ($\mu = 1.7$) found by Ramos-Fernández et al. [37] for spider monkeys in the forest of the Yucatan Peninsula, Mexico. It is also close to the estimated $\mu = 1.8$ in the paper by Boyer et al. (2012) [38] on

capuchin monkeys in Barro Colorado Island, Panama. Boyer et al. compare the motion patterns of monkeys to those of humans as reported in the paper by Song et al. (2010) [14], who find the same exponent ($\mu = 1.8$) for mobile-phone traces. Rhee et al. (2011) [13] report approximately similar values for waiting times in GPS traces of human walks performed in several outdoor settings of tens of kilometers. However, in the case of the North Carolina state fair, which is the case most similar to our research as it (seemingly) involves pedestrian movement only, the exponent is $\mu = 2.68$ and the truncated Pareto does not appear to be the best fit [13].

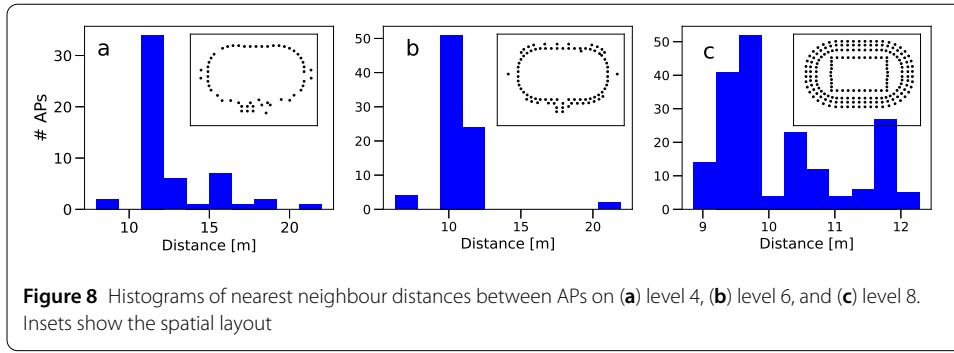
Heavy-tailed waiting times in movement behaviour affect the scaling of the mean-square displacement [19], and have important implications for diffusion-related processes, such as the spread of an epidemic disease [48], or success of random search strategies [49, 50].

The SSM framework matches the Wi-Fi tracking use case, as Wi-Fi based movement data involve non-negligible positioning error. SSMs can deal with various kinds of measurement error, such as heavy-tailed noise [28]. Here we have not fully exploited this potential as we have prefiltered the movement data using a simple moving average. The Wi-Fi measurement errors maintain a complex relationship with device position in the building. The distribution of positioning errors seems to have heterogeneous characteristics in different segments of the movement tracks, indicative of at least heteroscedasticity, if not regime switching. It is not expected that this complex error process holds (accessible) insights that can be generalized to other situations. In this paper, we have chosen to sidestep the nontrivial endeavour of modeling this process, and to simply alleviate the problem by smoothing the movement data. The smoothed movement tracks still contain positioning error however, and SSMs are to be preferred over hidden Markov models.

Previous studies on human mobility analysed movement patterns within the continuous-time random walk (CTRW) modeling framework (e.g. [13, 14]). The CTRW framework is especially relevant for motion patterns characterized by heavy-tailed distributions of step lengths and waiting times. Translating movement data into discrete random walks for the identification of Lévy walks depends on procedures involving *ad hoc* choices of parameters [41]. For example, Rhee et al. (2011) use a threshold radius to determine whether two consecutive locations are a ‘flight’ or a pause [13]. Though methods have been proposed that remediate some of these issues [51, 52], additional methods and choices are still required to assign behavioural states to movement tracks. It has been shown that such conventional approaches to time-varying random walks have shortcomings, and may lead to incorrect interpretations [53, 54]. The advantage of the SSM approach is that it allows to infer movement model parameters and behavioural states simultaneously, from which movement metrics (e.g. mean turning angles, or waiting times) can be directly derived. In addition, SSMs can model behavioural switches as functions of time, internal state, or environmental characteristics [22]. It would be interesting for example, to model switching behaviour in relation to local crowd density, and to see how people re-distribute themselves in the presence of overcrowding. Exploring aspects such as these merits further research.

Appendix A: Data processing pipeline

In this Appendix section we reiterate the steps in our data processing pipeline, and illustrate effects of some of the transformational steps and corresponding parameter choices.



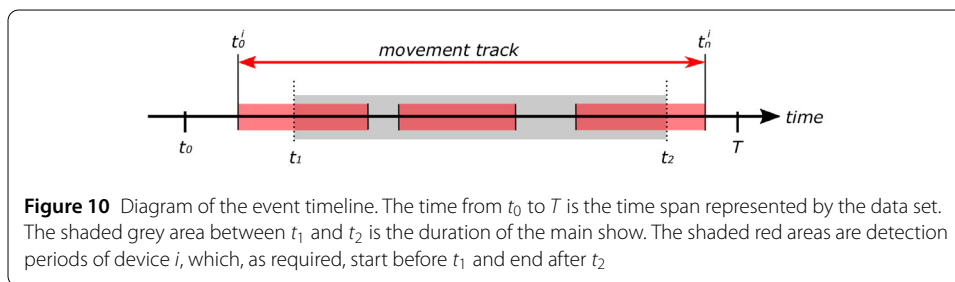
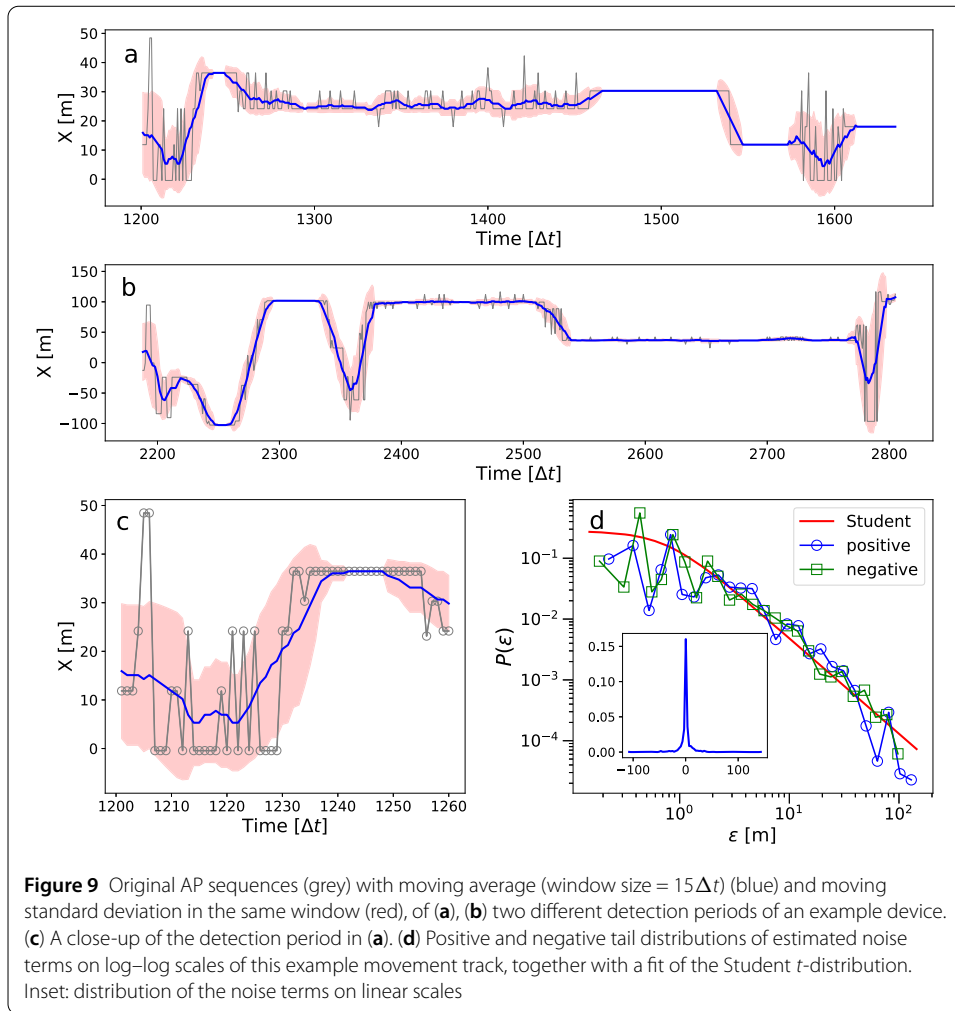
Wireless network The maximum accuracy of the proximity sensing method (explained below) is limited by the spatial resolution of the access points (APs) grid. To give an indication of this accuracy, we show here a characterisation of the distances in the grid. The wireless network consists of 591 APs with known spatial coordinates. The APs are distributed over 8 levels in the stadium. In Fig. 8 we show the distribution of nearest neighbour distances between APs on levels 4, 6, and 8. The APs on these levels cover most of the areas accessible to visitors of the entertainment events under study. Level 4 includes the main corridor connecting all parts of the building, while levels 6 and 8 cover the stands and pitch. The average nearest neighbour distance is 12.7 m on level 4, 10.7 m on level 6, and 10.2 m on level 8. The maximum accuracy of positions is thus on the order of ~ 10 meters.

Detection periods We select devices with detection periods of minimum length $\text{minperiod} = 5$ minutes, containing time gaps that do not exceed a maximum length $\text{maxgap} = 1$ minute. After applying this criterion we have, for each selected device, one or several detection periods ranging from 5 minutes (minperiod) to several hours, that are separated in time at least 1 minute (maxgap).

Proximity detection We estimate locations of smart phones using proximity detection, which determines the position of a device based on its closeness to an AP with known spatial coordinates [25]. First, we discretize all the data in time intervals $\Delta t = 10$ seconds. Then we group, for each device, and per time interval, the set of APs at which the device was detected. We select the AP with the highest RSS value in the time interval. The highest RSS value can occur at more than one AP. In that case we randomly select one of the duplicates. The proximity detection method produces temporal sequences of APs which are estimated to have been near a device. As a result of the initial method of selecting detection periods, the temporal AP sequences contain gaps up to 1 minute (maxgap) or equivalently 6 time bins. These gaps are closed with a simple forward filling (of the empty time bins) using the last selected AP before the gap. We ignore the z -coordinate and simplify the analysis to two-dimensions.

Smoothing To deal with fluctuations due to measurement noise we use a simple moving average to smooth the movement tracks. We apply a moving average filter

$$\hat{m}_{t,i} = (2q + 1)^{-1} \sum_{j=-q}^q X_{t-j,i} \quad (14)$$



to each coordinate i of the two-dimensional movement track time series $\mathbf{X}_t = (X_{t,1}, X_{t,2})$. In Fig. 9 we show examples of the result of applying the moving average with window size $2q + 1 = 15$, together with the moving standard deviation in the same window. We see that the Wi-Fi measurements have a heterogeneous error structure, and have different characteristics in different segments of the tracks. In Fig. 9(d) we show distributions of estimated noise terms $\epsilon_i = X_{t,i} - \hat{m}_{t,i}$. We show negative and positive tail distributions $P(|\epsilon_i|)$, together with a fit of the Student t -distribution. The t -distribution is shown to illustrate that (at least in some parts of the tracks) the estimated noise terms can be described by a heavy-tailed distribution. In this research, we have chosen to sidestep modelling the

complex error process, and to simply alleviate the problem by smoothing the movement data.

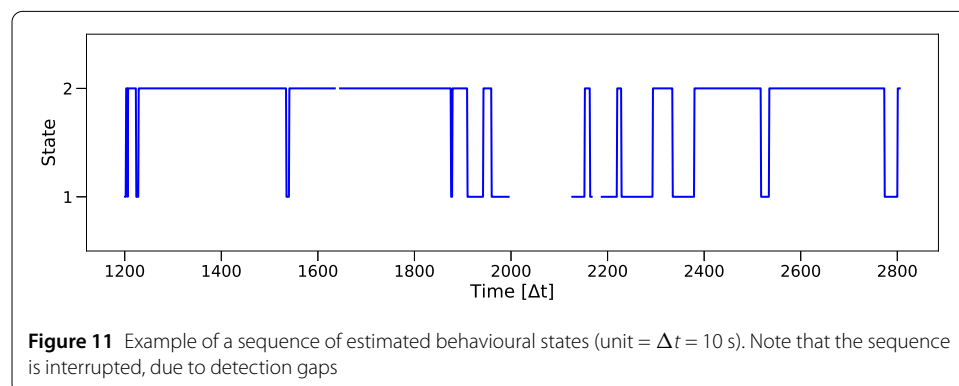
Device selection We select devices whose detection periods span the duration of the event. To this end, we select devices that were detected both in the time period before the start, and in the time period after the end, of the match and DJ show. See Fig. 10 for an illustration. The time from t_0 to T is the time span represented by the data set. The shaded grey area between t_1 and t_2 is the duration of the main show. The shaded red areas are detection periods of device i , which, as required, start before t_1 and end after t_2 . We wish to restrict our attention to movement tracks with minimum amount of gaps. Instead of filtering devices on maximum gap length, we look at the occupation number. The occupation number is the number of time bins between the first and last detection of device i (t_0^i and t_n^i in Fig. 10) which contain detections. We select devices that have an occupation number of at least $1/3$.

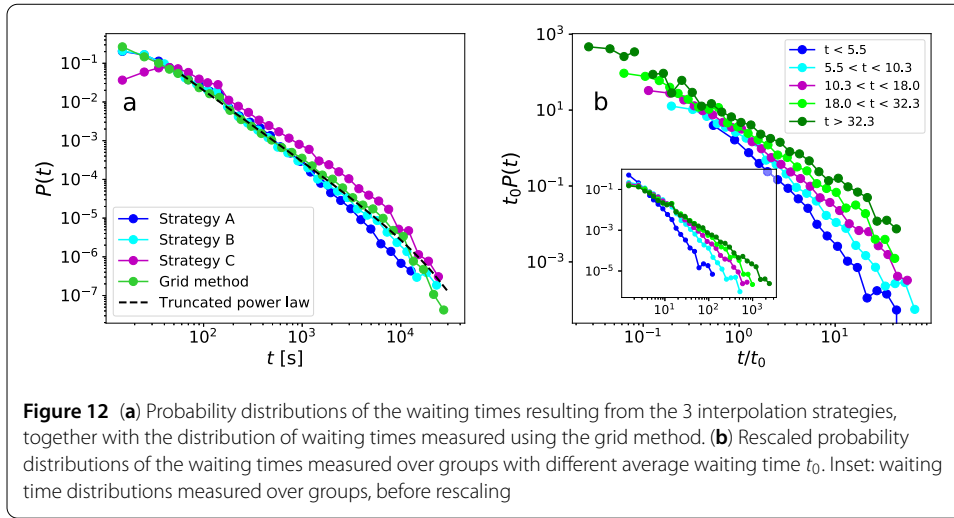
Finally, to filter out devices that did not move at all during the whole event, we select movement tracks for which the MRG exceeds a threshold value at least once. We use an intuitively chosen threshold value ($r_{g,t} > 10$ m) which is large enough to include only tracks for which movement is unambiguous.

Quantile sampling We sample 320 devices which we use for joint estimation. The devices are sampled in proportion to 5-quantiles in the frequency distribution of the observed maximum of $r_{g,t}$ over the devices.

Extracting waiting times The SSM produces temporal sequences of estimated behavioural states, see Fig. 11 for an example. We measure the length of the time periods the sequence dwells in the non-moving state, which defines the waiting times. In Fig. 11 we also see that the sequence is interrupted, due to the detection gaps.

We explore three strategies for extracting waiting times from the estimated behavioural state sequences (see main text Sect. 2.4.3). Here, we compare results of the three strategies with a simple method to extract waiting times directly from the movement tracks. The method of analysis is similar to Boyer et al. (2012) [38]. We discretize the stadium into square cells of size 15×15 m, a size that roughly corresponds to the maximum spatial resolution. Waiting times t are measured as the number of consecutive time intervals (Δt_i) in the same grid cell. We refer to this strategy as grid method. In Fig. 12(a) we compare resulting probability distributions of the waiting times, measured using the different





strategies A–C, and the waiting times measured using the grid method. Note that we apply the grid method to the fully interpolated movement tracks (which we use for strategy C). We also show the maximum likelihood estimate (MLE) fit of the truncated power law distribution. The estimated power law exponent $\alpha = 1.82$ is close to the MLE value of $\alpha = 1.74$ found for the waiting time distributions resulting from strategy C (see main text Sect. 3).

The heterogeneity in waiting times may be the result of heterogeneity in individual movement behaviour. For example, some people may prefer to walk around to explore the stadium, while others prefer to stay in one location. To explore this aspect, we use the method described in [26, 55]. We combine individuals in groups based on their average waiting time t_0 . We create waiting time probability distributions $P(t)$ for each group separately. In Fig. 12(b) (inset) we see that the slope of $P(t)$ is different for each group and decreases with average waiting time. To see whether there is underlying commonality we rescale the distributions as $t_0 P(t/t_0)$. The curves do not collapse into a single curve characterizing all individuals (as in [26, 55]), which would represent a universal characteristic of the system. This indicates there is some irreducible difference between the individual movement patterns. However, the heterogeneity in each separate (group) curve remains (i.e. they do not resolve into multiple exponential distributions). This suggests that the observed heavy-tailed waiting times are a combination of individual heterogeneity and population-based heterogeneity.

Error covariance matrix \mathbf{Q} The movement tracks are not accurate. The state-space modelling approach allows to encode expectation of accuracy in an observation model, and to avoid overfitting to measurement error. To do so, we use fixed values for the error σ^2 in the error covariance matrix

$$\mathbf{Q} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix}. \quad (15)$$

To estimate error values σ^2 we run a second pass of the moving average filter (Equation (14)) with window size $2q + 1 = 15$. Thus,

$$\sigma_i^2 = \text{Var}(\hat{m}_{t,i} - \hat{m}'_{t,i}), \quad (16)$$

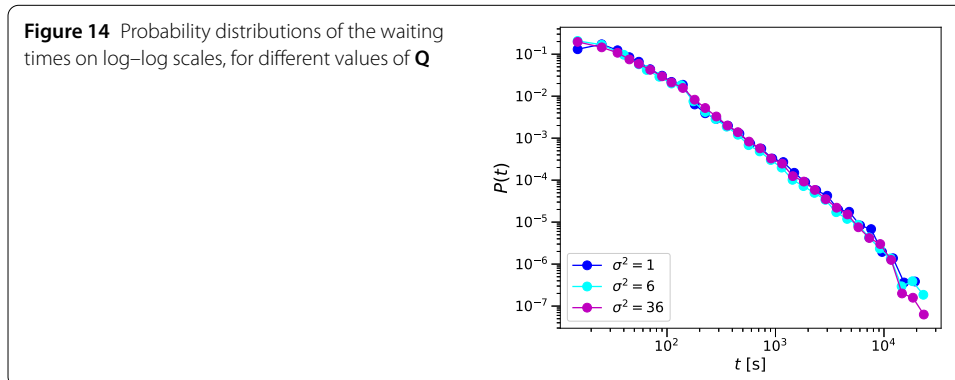
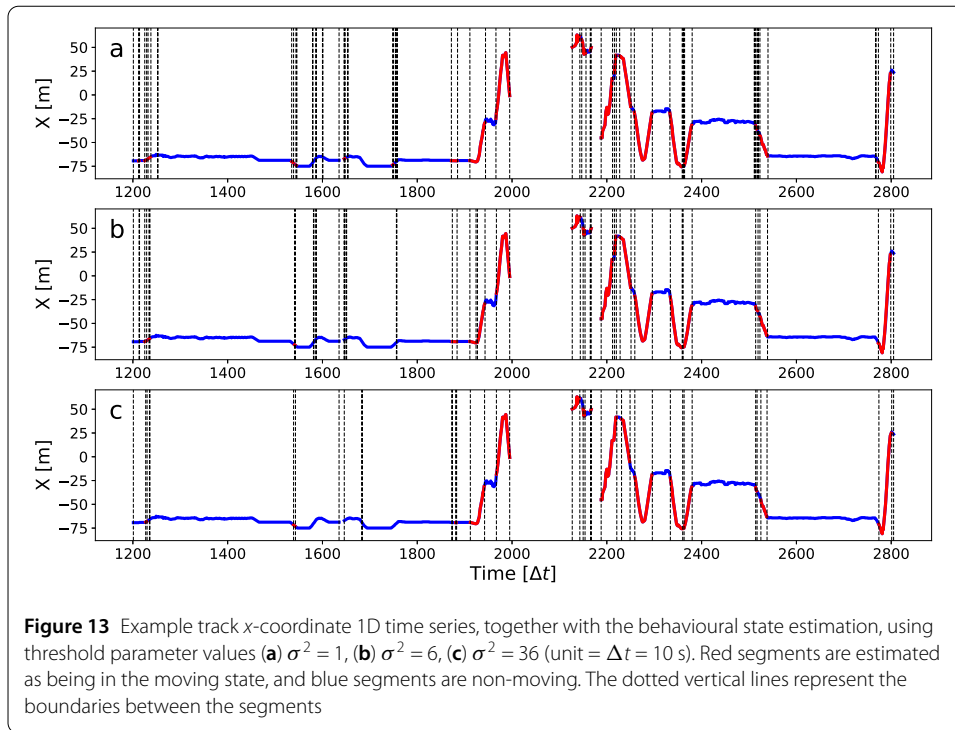


Table 3 Overview of fitting results using different values for \mathbf{Q} . Shown are the MLEs for the truncated power law, which provided the best model for the data, according to model selection based on Akaike weights

Parameter value	t_{\max} (s)	α	t_c (s)
$\sigma^2 = 1$	21,560	1.95	2632
$\sigma^2 = 6$	25,720	2.07	4363
$\sigma^2 = 36$	25,980	1.99	2055

where $\hat{m}'_{t,i}$ is the second pass of the moving average, for each coordinate i of the two-dimensional movement track. From the data we get $\sigma_x^2 = \sigma_y^2 \approx 6$. However, below we show that the state estimation and resulting waiting time distributions are robust against variation in \mathbf{Q} .

Note that the SSM effectively divides the movement tracks into segments based on behaviour, and functions as a path segmentation algorithm. The path segmentation defines

the lengths of the periods the device dwells in each state. Here we illustrate the effect of different choices for the parameter values for \mathbf{Q} on the path segmentation. In Fig. 13(a)–(c) we compare path segmentation results for the example movement track, for parameter values $\sigma^2 = (1, 6, 36)$. Although the number of state switches increases when σ^2 decreases, the changes seem to concentrate in specific regions of the track. The longer periods in each state are not affected. In Fig. 14 we show the resulting probability distributions of the waiting times, measured after fitting the SSM to the device sample using different values for \mathbf{Q} . We see that the results are robust against variation in \mathbf{Q} . The truncated power law distribution provides the best model for all three values of σ^2 , according to model selection based on Akaike weights. In Table 3 we show MLEs of the truncated power law, for different values of σ^2 .

Appendix B: MCMC model fitting

B.1 Prior distributions

In Table 4 we show the prior distributions used in the MCMC model inference for the CGLSSM. Note that the process covariance matrix (see Equation (3)), indexed by behavioural state, $\mathbf{R}(S_t)$, is defined as follows

$$\mathbf{R}(S_t) = \begin{bmatrix} \sigma_s^2 & 0 \\ 0 & \sigma_s^2 \end{bmatrix},$$

where σ_s^2 is the process variance in state $s = 1$ or 2 , and where we have kept covariance values zero. To ensure convergence to different values, for both chains we use starting values ($\sigma_1^2 = 10, \sigma_2^2 = 0.1$). Note that in JAGS the normal distribution is defined by `dnorm(mu, tau)`, where `tau` is the precision τ , which relates to the variance as $1/\sigma^2$.

Table 4 Prior distributions for model parameters of the CGLSSM

Parameter	Description	Prior distribution
τ_1	Process variance in state 1	Uniform(0, 20)
τ_2	Process variance in state 2	Uniform(0, 20)
α_1	Switching probability	Uniform(0, 1)
α_2	Switching probability	Uniform(0, 1)
Φ_1	Probability of state 1 at $t = 1$	Uniform(0, 1)
Φ_2	Probability of state 2 at $t = 1$	Uniform(0, 1)

Table 5 Prior distributions for model parameters of the DCRWS

Parameter	Description	Prior distribution
γ_1	Persistence (state 1)	Beta(5, 2)
γ_2	Persistence (state 2)	Beta(2, 5)
θ_1	Mean turn angle (state 1)	$\pi(2 \text{ Beta}(10, 10) - 1)$
θ_2	Mean turn angle (state 2)	$2\pi \text{ Beta}(10, 10)$
α_1	Switching probability	Uniform(0, 1)
α_2	Switching probability	Uniform(0, 1)
Φ_1	Probability of state 1 at $t = 1$	Uniform(0, 1)
Φ_2	Probability of state 2 at $t = 1$	Uniform(0, 1)
\mathbf{R}	Process variance in x- and y-direction	Wishart($\Sigma, 2$)

In Table 5 we show prior distributions for the DCRWS model. Note that for the Wishart prior

$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Prior distributions for the DCRWS are adopted from Jonsen (2016) [31].

B.2 Convergence diagnostics

Convergence of the model was confirmed by the Gelman–Rubin [56] *potential scale reduction factors* (psrf), which were < 1.01 for all estimated parameters. The psrf values were calculated using the R package `runjags`.

Appendix C: Statistical methods of model comparison

We use statistical methods of Clauset et al. [34] and Edwards et al. [36] for fitting the distributions. We also check results using the Python `powerlaw` package [35], which has implemented the methods from [34].

C.1 Maximum likelihood estimation

The maximum likelihood estimate (MLE) of the parameter λ of the exponential distribution $p(x) = \lambda e^{-\lambda(x-a)}$ is given by

$$\hat{\lambda} = 1 / \left(\sum_{i=1}^n x_i / n - a \right), \quad (17)$$

where n is the number of data points, and a is the lower bound of the fitting range. In this research a is loosely determined as the value after which the decay starts in the empirical frequency distribution.

There are no analytical solutions for the MLEs of the truncated power law distribution, stretched exponential, and log-normal distributions. In this case we numerically minimise the negative log-likelihood function

$$L(\theta) = - \sum_{i=1}^n \log p(x_i | \theta). \quad (18)$$

For the numerical minimisation we use Python library functions (following [35]).

C.2 Akaike model selection

To compute Akaike weights we need the Akaike Information Criterion (AIC)

$$\text{AIC} = 2L(\theta_{\text{mle}}) + 2K \quad (19)$$

for which we require the value of the negative log-likelihood function at the maximum (MLE), and where K is the number of parameters to be estimated [36, 39]. The AIC differences are

$$\Delta_i = \text{AIC}_i - \text{AIC}_{\min}, \quad (20)$$

where AIC_{\min} is the AIC of the model with the minimum AIC, which is considered as the best model. The Akaike weights are given by

$$w_i = \frac{\exp(-\frac{1}{2}\Delta_i)}{\sum_{m=1}^M \exp(-\frac{1}{2}\Delta_m)}, \quad (21)$$

where M is the set of models to be compared.

Acknowledgements

We acknowledge the financial support of the Netherlands eScience Center, under grant number 027.015.001.

Availability of data and materials

The data analysis conducted in the current study is licensed under the contractual agreement that no data are made available outside of the data processing environment of the Johan Cruijff ArenA.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

PR designed the methodology, performed the data analysis and wrote the manuscript, under daily supervision of MHL. SK organised the research project and obtained permission for data collection. SK and PMAS gave advice. All authors read and approved the final manuscript.

Author details

¹Informatics Institute, University of Amsterdam, Science Park 904, 1098XH Amsterdam, The Netherlands. ²Institute for Advanced Study, 1012 GC Amsterdam, The Netherlands.

Publisher's Note

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

Received: 16 November 2020 Accepted: 22 June 2021 Published online: 06 July 2021

References

- Helbing D, Farkas I, Vicsek T (2000) Simulating dynamical features of escape panic. *Nature* 407(6803):487–490
- Moussaid M, Helbing D, Theraulaz G (2011) How simple rules determine pedestrian behavior and crowd disasters. *Proc Natl Acad Sci USA* 108(17):6884–6888
- Bottinelli A, Sumpter DTJ, Silverberg JL (2016) Emergent structural mechanisms for high-density collective motion inspired by human crowds. *Phys Rev Lett* 117(22):228301
- Helbing D, Johansson A, Al-Abideen HZ (2007) The dynamics of crowd disasters: an empirical study. *Phys Rev E* 75:046109
- Johansson A, Batty M, Hayashi K, Bar OA, Marozzi D, Memish ZA (2012) Crowd and environmental management during mass gatherings. *Lancet Infect Dis* 12(2):150–156
- Barrat A, Cattuto C, Colizza V, Gesualdo F, Isella L, Pandolfi E, Pinton J-F, Ravà L, Rizzo C, Romano M, Stehlé J, Tozzi AE, den Broeck WV (2013) Empirical temporal networks of face-to-face human interactions. *Eur Phys J Spec Top* 222:1295–1309
- Goscé L, Barton DAW, Johansson A (2014) Analytical modelling of the spread of disease in confined and crowded spaces. *Sci Rep* 4:4856
- Helbing D, Molnar P (1995) Social force model for pedestrian dynamics. *Phys Rev E* 51(5):4282–4286
- Helbing D, Johansson A, Al-Abideen HZ (2001) Traffic and related self-driven many-particle systems. *Rev Mod Phys* 73(4):1067–1141
- Gutierrez-Roig M, Sagarra O, Oltra A, Palmer JRB, Bartumeus F, Diaz-Guilera A, Perello J (2016) Active and reactive behaviour in human mobility: the influence of attraction points on pedestrians. *R Soc Open Sci* 3:160177
- Rutten P, Lees MH, Klous S, Sloot PMA (2021) Intermittent and persistent movement patterns of dance event visitors in large sporting venues. *Phys A, Stat Mech Appl* 563:125448
- Brockmann D, Hufnagel L, Geisel T (2006) The scaling laws of human travel. *Nature* 439:462–465
- Rhee I, Shin M, Hong S, Lee K, Kim SJ, Chong S (2011) On the Levy-walk nature of human mobility. *IEEE/ACM Trans Netw* 19:630–643
- Song C, Koren T, Wang P, Barabási A-L (2010) Modelling the scaling properties of human mobility. *Nat Phys* 6:818–823
- Alessandretti L, Sapiezynski P, Lehmann S, Baronchelli A (2017) Multi-scale spatio-temporal analysis of human mobility. *PLoS ONE* 12(2):e0171686
- O'Brien WJ, Browman HI, Evans BI (1990) Search strategies of foraging animals. *Am Sci* 78(2):152–160
- Kramer DL, McLaughlin RL (2001) The behavioral ecology of intermittent locomotion. *Am Zool* 41:137–153
- Bénichou O, Loverdo C, Moreau M, Voituriez R (2006) Two-dimensional intermittent search processes: an alternative to Lévy flight strategies. *Phys Rev E* 74:020102
- Portillo IG, Campos D, Mendez V (2011) Intermittent random walks: transport regimes and implications on search strategies. *J Stat Mech* 2011:P02033

20. Nathan R, Getz WM, Revilla E, Holyoak M, Kadmon R, Saltz D, Smouse PE (2008) A movement ecology paradigm for unifying organismal movement research. *Proc Natl Acad Sci USA* 105(49):19052–19059
21. Patterson TA, Parton A, Langrock R, Blackwell PG, Thomas L, King R (2017) Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges. *ASTA Adv Stat Anal* 101:399–438
22. Patterson TA, Thomas L, Wilcox C, Ovaskainen O, Matthiopoulos J (2008) State–space models of individual animal movement. *Trends Ecol Evol* 23(2):87–94
23. Karsai M, Jo HH, Kaski K (2018) Bursty human dynamics. *SpringerBriefs in complexity*. Springer, Berlin
24. Barabási AL (2005) The origin of bursts and heavy tails in human dynamics. *Nature* 435:207–211
25. Kushki A, Plataniotis KN, Venetsanopoulos AN (2012) WLAN positioning systems: principles and applications in location-based services. Cambridge University Press, Cambridge
26. Gonzalez MC, Hidalgo CA, Barabasi AL (2008) Understanding individual human mobility patterns. *Nature* 453:779–782
27. Brockwell PJ, Davis RA (2016) Introduction to time series and forecasting, 3rd edn. Springer, Cham
28. Cappé O, Moulines E, Rydén T (2005) Inference in hidden Markov models. Springer, New York
29. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2014) Bayesian data analysis, 3rd edn. CRC Press, Boca Raton
30. Jonsen ID, Myers RA, James MC (2006) Robust hierarchical state–space models reveal diel variation in travel rates of migrating leatherback turtles. *J Anim Ecol* 75:1046–1057
31. Jonsen I (2016) Joint estimation over multiple individuals improves behavioural state inference from animal movement data. *Sci Rep* 6:20625
32. Breed GA, Jonsen ID, Myers RA, Bowen WD, Leonard ML (2009) Sex-specific, seasonal foraging tactics of adult grey seals (*Halichoerus grypus*) revealed by state-space analysis. *Ecology* 90(11):3209–3221
33. Kivela M, Porter MA (2015) Estimating interevent time distributions from finite observation periods in communication networks. *Phys Rev E* 92:052813
34. Clauset A, Shalizi C, Newman M (2007) Power-law distributions in empirical data. *SIAM Rev* 51:661–703
35. Alstott J, Bullmore E, Plenz D (2014) powerlaw: a Python package for analysis of heavy-tailed distributions. *PLoS ONE* 9(4):e95816
36. Edwards AM, Phillips RA, Watkins NW, Freeman MP, Murphy EJ, Afanasyev V, Buldyrev SV, da Luz MGE, Raposo EP, Stanley HE, Viswanathan GM (2007) Revisiting Levy flight search patterns of wandering albatrosses, bumblebees and deer. *Nature* 449:1044–1048
37. Ramos-Fernandez G, Mateos JL, Miramontes O, Cocho G, Larralde H, Ayala-Orozco B (2004) Levy walk patterns in the foraging movements of spider monkeys (*Ateles geoffroyi*). *Behav Ecol Sociobiol* 55(3):223–230
38. Boyer D, Crofoot MC, Walsh PD (2012) Non-random walks in monkeys and humans. *J R Soc Interface* 9:842–847
39. Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theoretic approach, 2nd edn. Springer, New York
40. Bartumeus F, da Luz MGE, Viswanathan GM, Catalan J (2005) Animal search strategies: a quantitative random-walk analysis. *Ecology* 86(11):3078–3087
41. Turchin P (2015) Quantitative analysis of movement: measuring and modeling population redistribution in animals and plants. Beresta Books
42. Jonsen ID, Flemming JM, Myers RA (2005) Robust state-space modeling of animal movement data. *Ecology* 86(11):2874–2880
43. Johnson DS, London JM, Lea MA, Durban JW (2008) Continuous-time correlated random walk model for animal telemetry data. *Ecology* 89(5):1208–1215
44. Reynolds AM (2010) Bridging the gap between correlated random walks and Levy walks: autocorrelation as a source of Levy walk movement patterns. *J R Soc Interface* 7:1753–1758
45. Plummer M JAGS: just another Gibbs sampler. Version 4.3.0. <https://sourceforge.net/projects/mcmc-jags/files/>
46. Plummer M, Stukalov A, Denwood M (2019) Bayesian graphical models using MCMC. Version 4-10
47. Reynolds AM (2011) On the origin of bursts and heavy tails in animal dynamics. *Physica A* 390:245–249
48. Vazquez A, Racz B, Lukacs A, Barabasi A-L (2007) Impact of non-Poissonian activity patterns on spreading processes. *Phys Rev Lett* 98:158702
49. Bartumeus F, Levin SA (2008) Fractal reorientation clocks: linking animal behavior to statistical patterns of search. *Proc Natl Acad Sci USA* 105(49):19072–19077
50. Campos D, Bartumeus F, Mendez V, Andrade JS Jr, Espadaler X (2016) Variability in individual activity bursts improves ant foraging success. *J R Soc Interface* 13:20160856
51. Humphries NE, Weimerskirch H, Sims DW (2013) A new approach for objective identification of turns and steps in organism movement data relevant to random walk modelling. *Methods Ecol Evol* 4:930–938
52. Tromer RM, Barbosa MB, Bartumeus F, Catalan J, da Luz MGE, Raposo EP, Viswanathan GM (2015) Inferring Levy walks from curved trajectories: a rescaling method. *Phys Rev E* 92:022147
53. Metzner C, Mark C, Steinwachs J, Lautscham L, Stadler F, Fabry B (2015) Superstatistical analysis and modelling of heterogeneous random walks. *Nat Commun* 6:7516
54. Gallotti R, Bazzani A, Rambaldi S, Barthelemy M (2016) A stochastic model of randomly accelerated walkers for human mobility. *Nat Commun* 7:12600
55. Goh K-I, Barabasi A-L (2008) Burstiness and memory in complex systems. *Europhys Lett* 81:48002
56. Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Stat Sci* 7(4):457–472