# Main statistical features for building a pedestrian mobility model

Author: Ferran Larroya Paixà
*Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.*

Advisor: Josep Perelló Palou

**Abstract:** We study the GPS data set of a collection of 260 students from up to 10 schools in the Barcelona Metropolitan Area on their trip from house to school. Participants collected the data with the help of a mobile application. After visualizing and cleaning the data, the main dynamical properties have been studied. We firstly study the diffusion of the movement and then we describe the instantaneous velocity with a log-Normal distribution, proposing a stochastic process to model it. We also obtain two typical stopping times for participants' stops duration. In addition, we calculate the orientation of the movement for each school and for the entire data set. Finally, we describe the change of orientation with a family of symmetric distributions on the circle, comparing the results for the different schools. We thus provide a first analysis to very valuable data to model pedestrian mobility with specific origin-destination journeys. Results found can also help to improve urban planing in the schools surroundings.

## I. INTRODUCTION

The study of complex systems has grown significantly in recent years, becoming a popular scientific research field [1]. Many multidisciplinary research groups composed by physicists, among others, build models trying to explain human behaviour or even financial markets [2,3]. In particular, the study of human mobility has aroused great interest in the last years due to the wide range of applications that it has. It allows improving the life quality of citizens, in terms of traffic jams, urban planning or excess of pollution, among other relevant issues that citizens face in their daily life [4,5].

The topic of this work is pedestrian mobility, with the aim of studying the movement of a collection of students on their journey to school. The work is based on a Citizen Science [6] project called Bee-Path, carried out by the research group OpenSystems UB, Dribia Data Research and Eduscopi [7].

In order to obtain the GPS positions, participants used a mobile application called Bee-Path, developed by OpenSystems UB and Dribia Data Research, available for Android and iOS. It has been used previously in different contexts, for example in the neighbourhood of *Les Corts* (Barcelona) [8] or at Barcelona's annual science festival [2]. This way of collecting data is not very common in human mobility studies, hence we are dealing with very valuable data.

Participants jointly decided to track the path from their houses to the school with the same experimental protocol. Although they also used transport methods such as underground, bus or car; we only focus on the pedestrian mobility. After analysing the collected data, we study several dynamical patterns by means of a statistical analysis. Concerning to the spatial part we focus on diffusion, velocities and we look for the duration of stops. We work with the module of the variables in the two-dimensional space. In reference to the polar part, we study the direction of the movement and the change of orientation. The study seeks to understand how students move to get to school, providing the main statistical features for building a pedestrian mobility model in a future work. The results could also be used to improve the urban planning and the quality of life in the schools surroundings.

## II. RESULTS

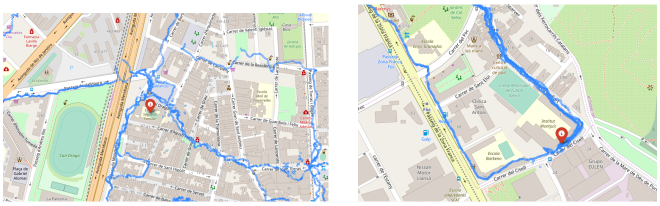
### A. Data acquisition and visualization

The collected data is composed by an anonymized username, date-time, latitude and longitude of every participant. The application is designed to collect data every second. After cleaning up the data by removing the outliers and non-pedestrian users, we ended up with 57,535 GPS locations corresponding to 104 participants from 9 schools of the Barcelona Metropolitan Area. Table I shows the number of points collected from each school.

| School | Points | Participants |
|---|---|---|
| *Institut Montjuïc* | 3,458 | 6 |
| *Institut Verdaguer* | 2,074 | 5 |
| *Col·legi Sant Gabriel de Viladecans* | 17,792 | 32 |
| *Institut Pau Claris* | 3,255 | 8 |
| *Col·legi Sagrada Família Sant Andreu* | 6,247 | 10 |
| *Institut Juan Manuel Zafra* | 7,921 | 17 |
| *Escola Virolai* | 2,890 | 6 |
| *Institut Bellvitge* | 5,929 | 10 |
| *Institut Ferran Tallada* | 7,969 | 10 |

TABLE I: Number of GPS points and individual journeys from each school.

In order to have a first approximation to the data, GPS positions are projected on maps to see how the students arrived at their respective schools. An example is shown in figure 1. The trajectories distribution of FIG.1(a) seems fairly uniform. On the case of FIG.1(b), almost all the trajectories start from the same region due to the fact that *Montjuïc* mountain constrain the space mobility on one side.



(a) (b)

FIG. 1: Projected trajectories of (a) *Col·legi Sagrada Família Sant Andreu (Sant Andreu)* and (b) *Institut Montjuïc (Sants-Montjuïc)*.

### B. Diffusion

We understand the students movement as a 2-dimensional random walk defined by successive discrete random steps of $\Delta t = 1s$ and length $r > 0$ [9]. An important quantity used to measure the spatial extend of random motion is the mean square displacement (MSD). It measures the deviation of the position of a pedestrian with respect to a reference position over time. The scaling of MSD with time, shown in Eq. (1), can be used to categorize the type of diffusive motion. $D$ is the diffusion coefficient and $d$ the spatial dimension ($d = 2$). If $\gamma = 1$, MSD scales linearly with time and ordinary Brownian motion is obtained (normal diffusion). On the other hand, if $\gamma < 1$ the motion is classed as sub-diffusive and if $\gamma > 1$ as super-diffusive.

$$MSD = < r^2(t) > - < r(t) >^2 = 2dDt^\gamma. \qquad (1)$$

We define $r(t)$ as the module of the distance between the position of a pedestrian in a certain time (after $n$ time steps of $\Delta t = 1s$) and the initial position when $t = 0$: $r(t) = |\vec{r}(t) - \vec{r}(t = 0)|$. Then, MSD in a certain time is obtained as the variance of $r(t)$, with the ensemble-average over all pedestrians. We compute the MSD for the first 150 time-steps of $\Delta t = 1s$, which we consider that is statistically enough time to study how MSD scales with time. The shape of MSD variance in a log-log scale shown in figure 2 can be fitted by two lines using Eq. (1), obtaining a super-diffusive motion: up to about 60 seconds with $\gamma = 1.71 \pm 0.01$ and $D = 0.46 \pm 0.002\ m^2/s$ and after that, with $\gamma = 1.934 \pm 0.008$ and $D = 0.171 \pm 0.007\ m^2/s$. This might be modelled with a correlated random walk or with a non-Gaussian model. So next natural statistical feature would be to look at the velocity module.

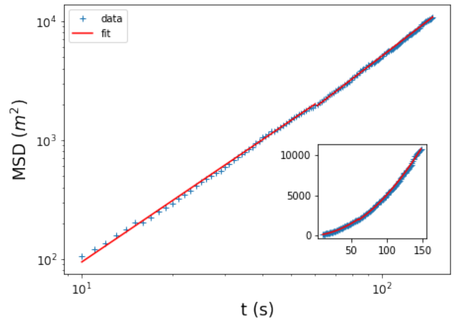



FIG. 2: MSD as a function of time in a log-log scale, for the first 150 seconds of movement, fitted with Eq. (1). The inset shows it in a normal scale.

### C. Instantaneous velocity

Velocity is obtained dividing the distance of two consecutive points by the time difference ($\Delta t = 1\ s$) between them: $v = \frac{|\Delta \vec{r}|}{\Delta t}$. Figure 3 shows the probability density function of the instantaneous velocity obtained from the data points of all the schools. We propose a log-Normal distribution in order to fit the data, based on a study which shows that in several human activities where exists a desired target to be reached by repeated choices, a log-Normal behaviour is found [10]. Eq. (2) shows the proposed fit, where $\mu$ and $\sigma$ are respectively the mean and the standard deviation of the logarithm of the variable.

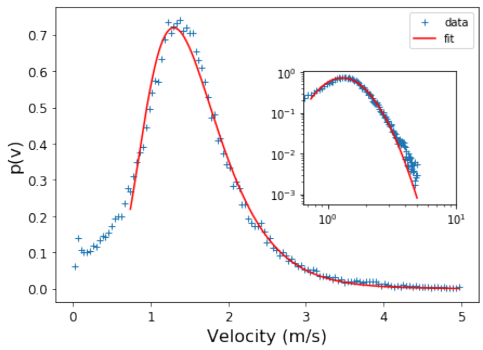



FIG. 3: Probability density function of instantaneous velocity, fitted with Eq. (2). The inset shows it in a log-log scale. The arithmetic mean value is $\overline{v} = 1.49 \pm 0.06\ m/s$ and the standard deviation is $\sigma = 0.7\ m/s$.

$$p(v) = \frac{1}{v \cdot \sigma\sqrt{2\pi}} e^{-\frac{(\ln(v)-\mu)^2}{2\sigma^2}}. \qquad (2)$$

Obtained parameters are $\sigma = 0.365 \pm 0.004$ and $\mu = 0.393 \pm 0.004$. The fit adapts quite well starting from a speed of $0.75\ m/s$ approximately. Below this value, GPS values may overestimate velocity. In Eq. (3) we propose a possible equation to model velocity called Ornstein-Uhlenbeck process, a stochastic process regularly used in financial mathematics [11]. It is the simplest known log-Normal model. We work with the logarithm of the velocity, thereby we define $v = e^u$.

$$du = -\alpha(u - u_o)dt + \kappa dW_t. \tag{3}$$

$W_t$ is the Wiener process and $\alpha, \kappa > 0$ are constants. Over time, the process tends to drift towards its long-term mean becoming a stationary Gaussian-Markov process, with mean $\mu = u_o$ and standard devation $\sigma = \kappa^2/2\alpha$. The Ornstein-Uhlenbeck process can be associated to a noisy relaxation process (for example a Hookean spring) fluctuating stochastically around its mean value.

### D. Stops duration

Another interesting aspect to investigate is the stop duration statistics. We define a stop at a give moment when the distance between two points separated by two time steps (of $\Delta t = 1s$) is less than $L = 1.8\ m$ (see Sect. IV B). Then, we obtain the duration time of each stop. The de-cumulative distribution function for the occurrences of stops duration time can be observed in figure 4. It can be well fitted by a weighted double exponential law with two different characteristic decaying times, as Eq. (4) shows.

$$\Psi(t) = Ae^{-t/a} + (1 - A)e^{-t/b}. \tag{4}$$

Contrary to other studies [2], too long stops are not expected for the activity of going to school, since the goal is to reach a place at a certain time. From the fit we obtain $A = 0.74 \pm 0.02$, $a = 2.95 \pm 0.08\ s$ and $b = 14.2 \pm 0.6\ s$. We believe that $b$ could be the typical time of stops at traffic lights to cross the street, or when someone waits for a friend to go together. On the other hand, $a$ is related to the reorientation of the movement when walking to school.

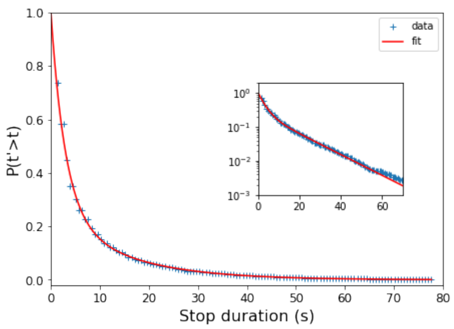



FIG. 4: De-cumulative distribution function of stops duration fitted with Eq. (4). The inset shows the representation in a log scale.

Finally, Eq. (5) shows the mean stopping time obtained from the fit equation.

$$< t >= \int_0^\infty \Psi(t)dt = Aa + (1 - A)b = 5.9 \pm 0.3\,s. \tag{5}$$

### E. Orientation

Going into the polar part of the work, we obtain the instantaneous direction of the movement by calculating the angle between the x-axis and each vector formed by two consecutive points. Figure 5 shows the polar histogram of orientation for every school. East is given by 0°, North by 90°, West by 180° and South by 270°.

The figures can provide information on the structure of the streets and reveal possible difficulties in accessing schools due to the space constraints. In fact, in many schools we can see a privileged axes of movement. For example in FIG.5(e) there is a peak around 315° and in FIG.5(f) the direction of the movement seems fairly uniform (cf. figure 1).

### F. Reorientation

We obtain the change of orientation angle of every consecutive vector by computing the difference of angles between two consecutive orientations, getting the set of angles in the range of $[-\pi, \pi]$.

Turning angle memory is a main feature for modelling random walks. Two well-known probability distributions used in movement modelling of animals are von Mises and wrapped Cauchy [12], both depend on a random variable whose values are angles. In Von Mises there is a preferred orientation and adapts better to Gaussian shapes. On the other hand, wrapped Cauchy is more peaked and has fatter tails, which implies different long-term consequences such as a power law decay or a log-Normal behaviour.

We fit the data to a probability distribution called Jones-Pewsey family of symmetric distributions on the circle [13], shown in Eq. (6), which includes both von Mises when $\psi = 0$ and wrapped Cauchy when $\psi = -1$. $P_{1/\psi}(z)$ is the associated Legendre function of degree $1/\psi$ and order 0, $\mu$ is a measure of location and $\kappa$ measures the concentration. Both $\kappa$ and $\psi$ are related to the degree of pedestrians orientation when exploring a given space. We can see the result of the fit for all the data set in figure 6. The obtained parameters are $\psi = -0.94 \pm 0.02$, $\kappa = 2.34 \pm 0.02$ and $\mu = -0.003 \pm 0.001$.

$$p(\theta) = \frac{(\cosh(\kappa\psi) + \sinh(\kappa\psi)\cos(\theta - \mu))^{1/\psi}}{2\pi P_{1/\psi}(\cosh(\kappa\psi))}. \tag{6}$$

A physics student obtained $\psi = -0.5$ (closer to von Mises) and $\kappa = 3.43$ (more concentrated around the mean) in his work on an experiment in a neighbourhood of Barcelona [8]. Hence our result shows a movement less

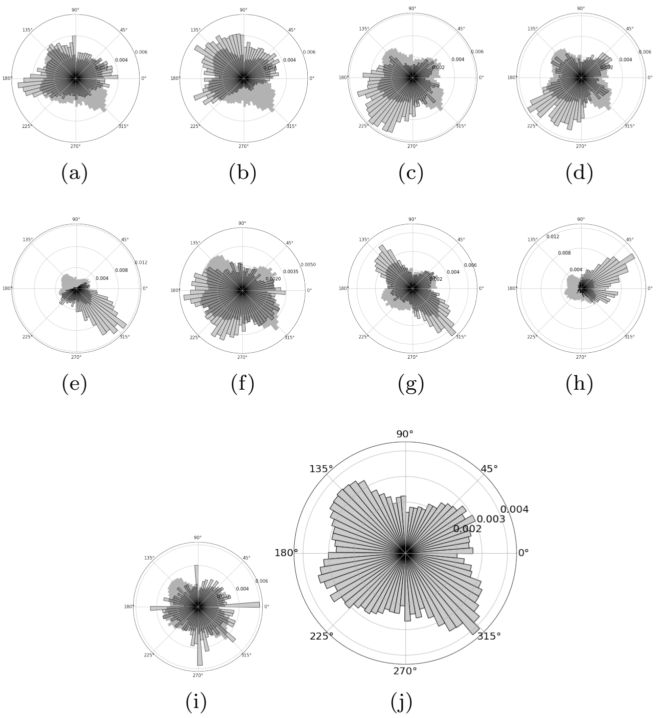



FIG. 5: Polar histogram where radial component measures the probability of an orientation at the corresponding angle. (a) *Institut Bellvitge (L'Hospitalet de Llobregat)*, (b) *Institut Ferran Tallada (Horta-Guinardó)*, (c) *Institut Juan Manuel Zafra (Sant Martí)*, (d) *Institut Pau Claris (Ciutat Vella)*, (e) *Institut Montjuïc (Sants-Montjuïc)*, (f) *Col·legi Sagrada Família Sant Andreu (Sant Andreu)*, (g) *Col·legi Sant Gabriel de Viladecans (Viladecans)*, (h) *Institut Verdaguer (Ciutat Vella)*, (i) *Escola Virolai (Horta-Guinardó)* and (j) All schools.

oriented. In fact, our data follows practically a wrapped Cauchy distribution ($\psi = -1$). Fatter tails imply that the probability of extreme angles is higher compared to a Normal or Von Mises process, and therefore a less oriented movement. This result may be due to the street furniture and the structure of the streets, decreasing the direct orientation towards the school.

Table II shows the values of $\psi$ and $\kappa$ obtained for each school separately. We obtain $\mu \sim 0$ and a behaviour closer to wrapped Cauchy for every school. Then, paying attention to the value of concentration $\kappa$, we can compare in a qualitative way what schools are easier to access. A higher value of $\kappa$ implies a more concentration around $\mu \sim 0$ and therefore a more oriented movement. This can be seen with the two schools in figure 1, where in FIG.1(b) ($\kappa = 2.66$) the movement is more oriented than in FIG.1(a) ($\kappa = 1.97$).

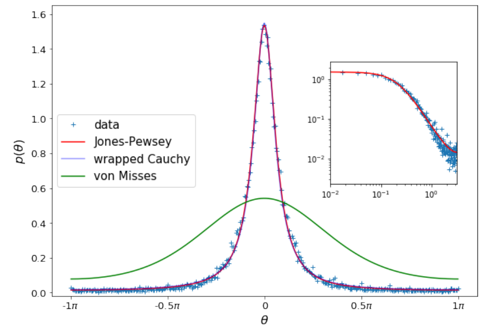



FIG. 6: Probability density function of the instantaneous reorientation fitted with Eq. (6). Two limits have been included (von Mises and wrapped Cauchy). The inset is in a log-log scale.

| School | $\Psi$ | $\kappa$ |
|---|---|---|
| *Escola Virolai* | $-1.51 \pm 0.08$ | $1.64 \pm 0.03$ |
| *Institut Bellvitge* | $-1.13 \pm 0.03$ | $2.13 \pm 0.03$ |
| *Col·legi Sagrada Família S.Andreu* | $-1.07 \pm 0.03$ | $1.97 \pm 0.02$ |
| *Col·legi Sant Gabriel de Viladecans* | $-0.89 \pm 0.02$ | $2.41 \pm 0.03$ |
| *Institut Juan Manuel Zafra* | $-0.87 \pm 0.02$ | $2.15 \pm 0.03$ |
| *Institut Pau Claris* | $-0.81 \pm 0.02$ | $2.38 \pm 0.03$ |
| *Institut Montjuïc* | $-0.80 \pm 0.03$ | $2.66 \pm 0.06$ |
| *Institut Verdaguer* | $-0.77 \pm 0.02$ | $3.1 \pm 0.1$ |
| *Institut Ferran Tallada* | $-0.72 \pm 0.02$ | $2.58 \pm 0.03$ |

TABLE II: Values of $\psi$ and $\kappa$ obtained for each school.

## III. CONCLUSIONS

We have provided the main statistical features for building a pedestrian mobility model in the particular case of students in their journey from house to school. A super-diffusive behaviour for the pedestrians motion have been found. We have been able to explain instantaneous velocity by means a log-Normal distribution, proposing a Ornstein-Uhlenbeck process that is a Gaussian and Markovian model (with exponential decay). In addition, we have described the duration of the stops with a double exponential law, obtaining two typical stopping times. We have also studied the orientation of the movement and the turning angles (reorientation) for each school and for the whole data set. Orientation may provide a glimpse of possible difficulties in accessing schools. Reorientation has been described with a family of symmetric distributions on the circle, resulting practically in a wrapped Cauchy. In some schools the resulting movement is more oriented than in others, hinting at the difficulties of accessing schools due to the structure of the streets and the urban furniture. All these statistical features shall be taken into account when considering new pedestrian models for mobility.

## IV. APPENDIX

### A. Data processing

Python programming language has been used for data processing, using libraries such as GeoPy to calculate the distance between geographical points, NumPy to perform operations and functions, Matplotlib for plotting the data and SciPy for fitting functions to experimental data using the last squares method. We decided to compute the average of every three instantaneous velocities in order to avoid possible errors due to the lack of precision of the GPS, getting smoothed results.

### B. Stop determination

Due to the uncertainty associated to GPS locations, we can not define a stop situation when the distance between consecutive points is zero. For this reason, we propose to fix a point with coordinates of latitude and longitude ($\phi_0$,$\lambda_0$), and compute the distance between that point and another one situated $k$ steps (of $\Delta t = 1s$) before it ($\phi_k$,$\lambda_k$). If that distance is less than a certain value $L$, called *stop threshold parameter*, the fixed point will be labelled as a stop. It is not trivial to choose an appropriate value for $L$ and it may be linked to a certain arbitrariness. For this reason we study the stop frequency as a function of $L$, as shown in figure 7.

Significant difference between $k = 1$ and larger $k$'s allows us to consider that $k = 2$ is good enough to describe the stopping or moving situations while keeping a robust statistics. In the inset of figure 7 we fit $k = 2$ to an analytical function to study the turning points (i.e. a change in function behaviour) that could give a good value for $L$. The chosen function for the fit is a third degree polynomial: $\Psi(L) = aL^3 + bL^2 + cL + d$. From the fit parameters we obtain a turning point in $L = 1.8 \pm 0.1$ $m$ with the second derivative test, which we consider that would be a reasonable value for $L$.

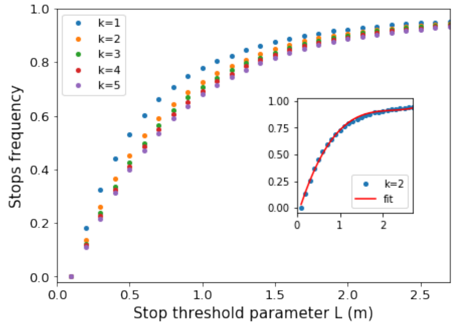



FIG. 7: Stops fraction as a function of the stop threshold parameter $L$ for the five first values of $k$. As expected, stop frequency is lower as the value of $k$ rises.

**Acknowledgments**

I would like to thank my advisor Josep Perelló for the reliance disposed on me and for his guidance throughout the study. Also thank my family and friends for their support and encouragement. Finally, thanks to the students volunteers for participating, providing us the data to perform the study.

[1] Caldarelli, G., Wolf, S., & Moreno, Y. "Physics of humans, physics for society". Nature Physics **14**: 870 (2018).
[2] Gutiérrez-Roig, M., Sagarra, O., Oltra, A., Palmer, J. R., Bartumeus, F., Daz-Guilera, A., & Perelló, J. "Active and reactive behaviour in human mobility: the influence of attraction points on pedestrians". Royal Society open science **3(7)**: 160177 (2016).
[3] Montero, M., & Masoliver, J. "Continuous Time Random Walks with memory and financial distributions". The European Physical Journal B: **90(11)** 207 (2017).
[4] Sugiyama, Y., Fukui, M., Kikuchi, M., Hasebe, K., Nakayama, A., Nishinari, K., Tadaki, S., & Yukawa, S. "Traffic jams without bottlenecksexperimental evidence for the physical mechanism of the formation of a jam". New journal of physics **10(3)**: 033001 (2008).
[5] Setton, E., Marshall, J. D., Brauer, M., Lundquist, K. R., Hystad, P., Keller, P., & Cloutier-Fisher, D. "The impact of daily mobility on exposure to traffic-related air pollution and health effect estimates". Journal of Exposure Science and Environmental Epidemiology **21(1)**: 42-48 (2011).
[6] Sagarra, O., Gutiérrez-Roig, M., Bonhoure, I., & Perelló, J. "Citizen science practices for computational social science research: The conceptualization of pop-up experiments". Frontiers in physics **3**: 93 (2016).
[7] *http://beepath.org/*
[8] Juncà Ramon, Manel. "Pedestrian mobility in a neighbourhood". Facultat de Física Universitat de Barcelona: Physics Degree Thesis (2017).
[9] Barbosa, H., Barthelemy, M., Ghoshal, G., James, C. R., Lenormand, M., Louail, T., Menezes R., Ramasco J.J., Simino F., & Tomasini, M. "Human mobility: Models and applications". Physics Reports **734**: 1-74 (2018).
[10] Gualandi, S., & Toscani, G. "Human behavior and lognormal distribution. A kinetic description". Math. Models Methods Appl. Scie. **29(4)**: 717-753 (2019)
[11] Borovkov, K., & Decrouez, G. "Ornstein–Uhlenbeck Type Processes with Heavy Distribution Tails". Theory of Probability & Its Applications **57(3)**: 396-418 (2013).
[12] Codling E. A., Plank M. J., & Benhamou S. "Random walks models in biology". J. R. Soc. Interface **5(25)**: 813-834 (2008).
[13] Jones, M. C., & Pewsey, A. "A family of symmetric distributions on the circle". Journal of the American Statistical Association **100(472)**: 1422-1428 (2005).