Supplementary Material

SUPPLEMENTARY FIGURES AND TABLES

Mobility represented	$\overline{t}$ (h)	$\overline{\tau}$ (h)	$P(\tau)$	Sampling	Type of result	F_{good}	Notes
Italian cars	0.30	2.49	Exponential	Periodic	Analytical	51%	$\overline{\Delta} = 1.7$h
			Stretched exponential	Periodic	Numerical	39%	$\overline{\Delta} = 1.3$h
			Exponential	Power law	Numerical	27%	–
			Stretched exponential	Power law	Numerical	23%	–
USA CDR data	(0.30)	0.55	Exponential	Periodic	Analytical	24%	$\overline{\Delta} = 0.8$h
			Truncated power law	Periodic	Numerical	15%	$\overline{\Delta} = 0.5$h
			Truncated power law	Power law	Numerical	6%	–
Chinese Geolife Traj.	0.33	0.80	Exponential	Periodic	Analytical	27%	$\overline{\Delta} = 1.0$h
			Empirical	Periodic	Numerical	18%	$\overline{\Delta} = 1.0$h
			Empirical	Empirical ($p = 0$)	Numerical	11%	–
			Empirical	Empirical ($p = 1$)	Numerical	16%	–

TABLE S1. Results' summary.

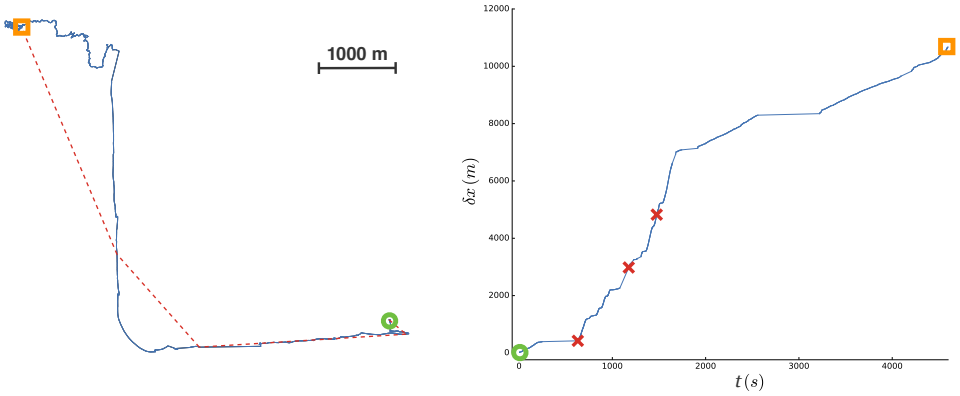


FIG. S1. **Representing trajectories. (Left)** A 2D trajectory from the green circle to the orange square, obtained with GPS readings (solid blue) and sampled trajectory (dashed red). **(Right)** Cumulated distance traveled on the real trajectory (solid blue) and sampling points (red crosses) drawn from a power-law distribution with exponent -1. The sampled trajectory is a gross approximation of the real trajectory, a lot of information being lost in the process.

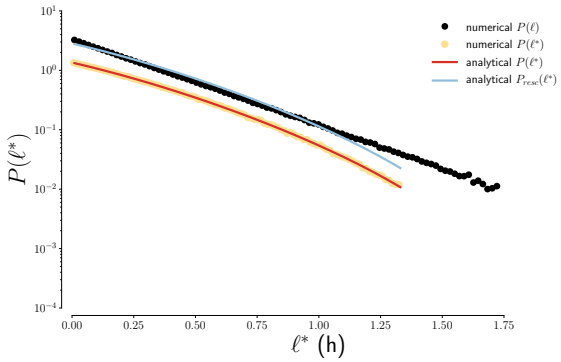


FIG. S2. **Rescaled distribution of travel times.** Optimal sampling of $\bar{t} = 0.30$ h with $\overline{\Delta} = \hat{\Delta} = 1.73$ h. The distribution of sampled travel times (light blue) can be compared with the original exponential distribution (black dots) after re-normalizing the distribution, multiplying it by the factor $(1 - C_0)^{-1}$.

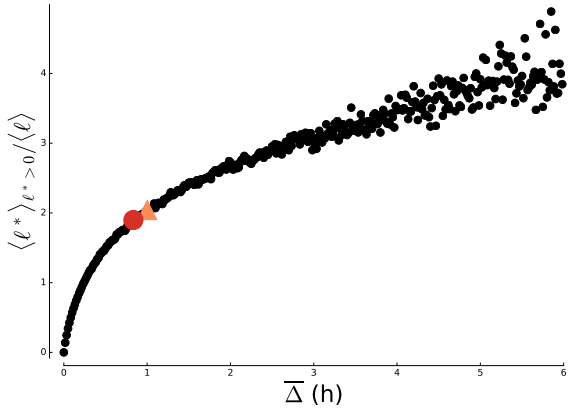


FIG. S3. **Constant sampling on GPS data.** We present results obtained by sampling the GeoLife GPS data with a constant sampling interval Δ. We plot the average sampled move displacement $\langle \ell^* \rangle_{\ell^*>0}$ (computed in 2 dimensions), normalised by the real average move length $\langle \ell \rangle$ as a function of the length of the sampling interval $\overline{\Delta}$. The ideal case $\langle \ell^* \rangle_{\ell^*>0} = \langle \ell \rangle$ (Eq. (9)) is reached for $\overline{\Delta} = 15$ min, while for a short-tailed rest time it is expected to be $\overline{\Delta} = 52$ min (red circle). The fact that the sampling time optimizing F_{good} (orange triangle) corresponds to $\langle \ell^* \rangle_{\ell^*>0}/\langle \ell \rangle \approx 2$ implies that optimal sampling frequencies would represent, in this case, an under-sampling of the trajectory where moves are more frequently joined together than cut by sampling times.

ANALYTICAL CALCULATIONS

Possible sampling scenarios

In order to compute the different statistics presented in the text, we use methods of renewal theory [47-49] along the lines presented in detail in [43], and we refer the interested reader to this last reference.

Sampling can get wrong in seven different ways. The cases are the following. We can have two sampling times falling:

1. in the same rest;

2*a*. in two subsequent rests (the correct way);

2*b*. in two rests separated by more than one move;

3*a*. in a move and in the rest following that move;

3*b*. in a move and in a rest not following that move;

4. in the same move;

5*a*. in a rest and in the move following that rest;

5*b*. in a rest and in a move not following that rest;

6*a*. in two subsequent moves;

6*b*. in two non-subsequent moves.

Case 1 can be identified, since the displacement is $\ell = 0$. Case 2*a* gives a correct evaluation of the move performed, since both sampling are made when the individual is still, and only one movement has been done in that time. Cases 3*a*, 4, 5*a* and 6*a* 'cut' moves, under-estimating the observed displacements and leading to an over-estimate of the number of moves. Cases 2*b*, 3*b*, 5*b* and 6*b* 'join' together multiple moves, thus yielding over-estimated displacements and under-estimated number of moves.

General setting

The random trajectory consists in an alternation of moves with durations $t_1, t_2, t_3, \ldots$, where the position $x(\theta)$ increases with unit velocity ($v = 1$), and of rests with durations $\tau_1, \tau_2, \tau_3, \ldots$, where $x(\theta)$ stays constant. The walker starts from $x = 0$ at time $\theta = 0$. The move durations t_k and the rest durations τ_k are drawn from two given continuous distributions $f(t)$ and $g(\tau)$.

We are interested in the distribution $P_{\theta_1,\theta_2}(\ell)$ of the distance

$$\ell = x(\theta_2) - x(\theta_1) \tag{S1}$$

traveled by the walker between two fixed times θ_1 and θ_2, and in various related quantities. An exact expression for the distribution $P_{\theta_1,\theta_2}(\ell)$ can be derived by analytical means, for arbitrary distributions $f(t)$ and $g(\tau)$, by using techniques from renewal theory [47–49]. The key quantity is the triple Laplace transform

$$\begin{aligned}\mathcal{L}(r,s,u) = &\int_0^\infty \mathrm{e}^{-r\theta_1}\,\mathrm{d}\theta_1 \int_{\theta_1}^\infty \mathrm{e}^{-s\theta_2}\,\mathrm{d}\theta_2 \\ &\times \int_0^\infty \mathrm{e}^{-u\ell}\,P_{\theta_1,\theta_2}(\ell)\,\mathrm{d}\ell.\end{aligned} \tag{S2}$$

This quantity can be evaluated as a sum over six sectors (see above discussion and Fig. S4):

1. θ_1 and θ_2 belong to the same τ_n;
2. θ_1 belongs to τ_m while θ_2 belongs to τ_n;
3. θ_1 belongs to t_m while θ_2 belongs to τ_n;
4. θ_1 and θ_2 belong to the same t_n;
5. θ_1 belongs to τ_m while θ_2 belongs to t_n;
6. θ_1 belongs to t_m while θ_2 belongs to t_n.

Let us illustrate the method on the example of sector 2. For fixed integers $m \geq 1$ and $n \geq m+1$, we have

$$\begin{aligned}\theta_1 &= \Theta_1 + B_1, \\ \theta_2 &= \Theta_2 + B_2, \\ \Theta_1 &= (t_1 + \cdots + t_m) + (\tau_1 + \cdots + \tau_{m-1}), \\ \Theta_2 &= (t_1 + \cdots + t_n) + (\tau_1 + \cdots + \tau_{n-1}), \\ x(\theta_1) &= t_1 + \cdots + t_m, \\ x(\theta_2) &= t_1 + \cdots + t_n, \\ \ell &= t_{m+1} + \cdots + t_n,\end{aligned} \tag{S3}$$

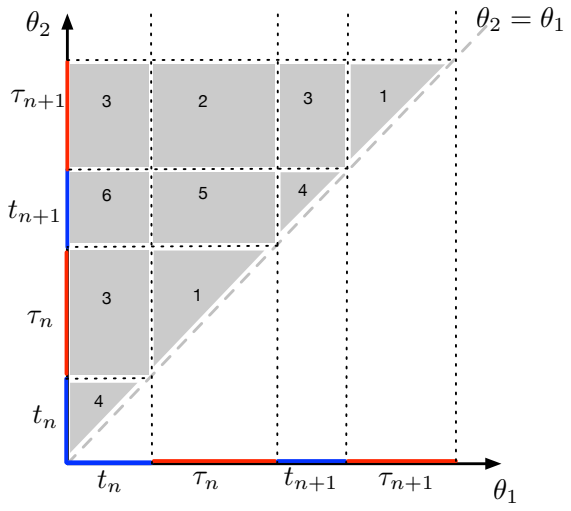


FIG. S4. **Schematic representation of the different sectors.** In the plane (θ_1, θ_2) we represent the different sectors according to the numbering defined in the text.

with $0 < B_1 < \tau_m$ and $0 < B_2 < \tau_n$. The contribution of sector 2 with fixed m and n to $\mathcal{L}(r,s,u)$ therefore reads

$$\mathcal{L}_2^{(m,n)}(r,s,u) = \Big\langle \mathrm{e}^{-r\Theta_1 - s\Theta_2 - u\ell} \times \int_0^{\tau_m} \mathrm{e}^{-rB_1}\,\mathrm{d}B_1 \int_0^{\tau_n} \mathrm{e}^{-sB_2}\,\mathrm{d}B_2 \Big\rangle. \tag{S4}$$

Hereafter we borrow conventions and notations from Ref. [43]. In particular, $\langle \dots \rangle$ denotes an average over the random process, i.e., over all the move durations t_k and rest durations τ_k.

The explicit evaluation of (S4) involves three steps.

• First, performing the two integrals leads to the expression

$$\mathcal{L}_2^{(m,n)}(r,s,u) = \left\langle \mathrm{e}^{-r\Theta_1 - s\Theta_2 - u\ell} \frac{1-\mathrm{e}^{-r\tau_m}}{r} \frac{1-\mathrm{e}^{-s\tau_n}}{s} \right\rangle, \tag{S5}$$

which only involves the t_k and τ_k.

• Second, averaging independently over all the t_k and τ_k, we obtain

$$\mathcal{L}_2^{(m,n)}(r,s,u) = \widehat{f}(r+s)^m \widehat{f}(s+u)^{n-m} \widehat{g}(r+s)^{m-1} \times \widehat{g}(s)^{n-m-1} \frac{\widehat{g}(s) - \widehat{g}(r+s)}{r} \frac{1-\widehat{g}(s)}{s} \tag{S6}$$

in terms of the Laplace transforms (characteristic functions)

$$\widehat{f}(s) = \langle \mathrm{e}^{-st} \rangle = \int_0^\infty f(t)\,\mathrm{e}^{-st}\,\mathrm{d}t,$$
$$\widehat{g}(s) = \langle \mathrm{e}^{-s\tau} \rangle = \int_0^\infty g(\tau)\,\mathrm{e}^{-s\tau}\,\mathrm{d}\tau. \tag{S7}$$

• Third, the entire contribution of sector 2 reads

$$\mathcal{L}_2(r,s,u) = \sum_{m=1}^\infty \sum_{n=m+1}^\infty \mathcal{L}_2^{(m,n)}(r,s,u), \tag{S8}$$

where the sums boil down to geometric sums. This leads to

$$\mathcal{L}_2(r,s,u) = \frac{\widehat{f}(s+u)}{1-\widehat{f}(s+u)\widehat{g}(s)}\,\frac{\widehat{f}(r+s)}{1-\widehat{f}(r+s)\widehat{g}(r+s)} \times \frac{\widehat{g}(s)-\widehat{g}(r+s)}{r}\,\frac{1-\widehat{g}(s)}{s}. \tag{S9}$$

The contributions of the five other sectors can be evaluated along the same lines. We thus obtain

$$\begin{aligned}
\mathcal{L}_1(r,s,u) &= \frac{\widehat{f}(r+s)}{1-\widehat{f}(r+s)\widehat{g}(r+s)} \\
&\quad\times \frac{r+s\widehat{g}(r+s)-(r+s)\widehat{g}(s)}{rs(r+s)}, \\
\mathcal{L}_3(r,s,u) &= \frac{1}{1-\widehat{f}(s+u)\widehat{g}(s)}\,\frac{1}{1-\widehat{f}(r+s)\widehat{g}(r+s)} \\
&\quad\times \frac{\widehat{f}(s+u)-\widehat{f}(r+s)}{r-u}\,\frac{1-\widehat{g}(s)}{s}, \\
\mathcal{L}_4(r,s,u) &= \frac{1}{1-\widehat{f}(r+s)\widehat{g}(r+s)} \\
&\quad\times \frac{u-r+(r+s)\widehat{f}(s+u)-(s+u)\widehat{f}(r+s)}{(r+s)(u-r)(s+u)}, \\
\mathcal{L}_5(r,s,u) &= \frac{1}{1-\widehat{f}(s+u)\widehat{g}(s)}\,\frac{\widehat{f}(r+s)}{1-\widehat{f}(r+s)\widehat{g}(r+s)} \\
&\quad\times \frac{\widehat{g}(s)-\widehat{g}(r+s)}{r}\,\frac{1-\widehat{f}(s+u)}{s+u}, \\
\mathcal{L}_6(r,s,u) &= \frac{\widehat{g}(s)}{1-\widehat{f}(s+u)\widehat{g}(s)}\,\frac{1}{1-\widehat{f}(r+s)\widehat{g}(r+s)} \\
&\quad\times \frac{\widehat{f}(r+s)-\widehat{f}(s+u)}{u-r}\,\frac{1-\widehat{f}(s+u)}{s+u}.
\end{aligned} \tag{S10}$$

Steady state

From now on we focus our attention onto the steady state of the process, obtained by letting the first time θ_1 go to infinity, keeping the time difference

$$\Delta = \theta_2 - \theta_1 \tag{S11}$$

fixed. This steady state is well-defined if the distributions $f(t)$ and $g(\tau)$ decay fast enough for the mean values $\overline{t}$ and $\overline{\tau}$ to be finite. In the opposite situation, where either $\overline{t}$ or $\overline{\tau}$ or even both are divergent, the process never reaches a steady state. It rather exhibits various non-stationary features, usually referred to as aging or weak ergodicity breaking [50]. We assume henceforth that $\overline{t}$ and $\overline{\tau}$ are finite.

The quantity of most interest is the steady-state distribution $P_\Delta(\ell)$. Its double Laplace transform

$$L(s,u) = \int_0^\infty \mathrm{e}^{-s\Delta}\,\mathrm{d}\Delta \int_0^\infty \mathrm{e}^{-u\ell}\,P_\Delta(\ell)\,\mathrm{d}\ell \tag{S12}$$

is the limit of the product $(r+s)\mathcal{L}(r,s,u)$ as $r \to -s$. We thus obtain

$$L(s,u) = \frac{N(s,u)}{(\overline{t}+\overline{\tau})s^2(s+u)^2(1-\widehat{f}(s+u)\widehat{g}(s))}, \tag{S13}$$

with

$$\begin{aligned}
N(s,u) &= s(s+u)(s\overline{t}+(s+u)\overline{\tau})(1-\widehat{f}(s+u)\widehat{g}(s)) \\
&\quad - u^2(1-\widehat{f}(s+u))(1-\widehat{g}(s)).
\end{aligned} \tag{S14}$$

The distribution $P_\Delta(\ell)$ has the general form

$$P_\Delta(\ell) = C_0(\Delta)\,\delta(\ell) + C_1(\Delta)\,\delta(\ell-\Delta) + P_\Delta^{\rm cont}(\ell), \tag{S15}$$

with two delta functions at the endpoints $\ell = 0$ and $\ell = \Delta$, and a non-trivial continuous piece in-between. The delta function at $\ell = 0$ corresponds to sector 1 (θ_1 and θ_2 belong to the same rest duration τ_n). The Laplace transform $\widehat{C}_0(s)$ of the associated amplitude $C_0(\Delta)$ reads

$$\widehat{C}_0(s) = \frac{\widehat{g}(s) + s\overline{\tau} - 1}{(\overline{t}+\overline{\tau})s^2}, \tag{S16}$$

hence

$$C_0(\Delta) = \frac{1}{\overline{t}+\overline{\tau}} \int_\Delta^\infty (\tau - \Delta) g(\tau)\,{\rm d}\tau. \tag{S17}$$

Similarly, the delta function of $P_\Delta(\ell)$ at $\ell = \Delta$ corresponds to sector 4 (θ_1 and θ_2 belong to the same move duration t_n). The associated amplitude reads

$$C_1(\Delta) = \frac{1}{\overline{t}+\overline{\tau}} \int_\Delta^\infty (t - \Delta) f(t)\,{\rm d}t. \tag{S18}$$

As Δ increases from 0 to infinity, the above amplitudes decrease monotonically from $C_0(0) = \overline{\tau}/(\overline{t}+\overline{\tau})$ and $C_1(0) = \overline{t}/(\overline{t}+\overline{\tau})$ to zero, whereas the weight of the continuous part $P_\Delta^{\rm cont}(\ell)$ increases from zero to one.

The mean value of ℓ has the remarkably simple expression

$$\langle \ell \rangle_\Delta = \frac{\overline{t}}{\overline{t}+\overline{\tau}}\,\Delta. \tag{S19}$$

The second moment of ℓ reads in Laplace space

$$\begin{aligned}\int_0^\infty {\rm e}^{-s\Delta} \langle \ell^2 \rangle_\Delta\,{\rm d}\Delta = {} & \frac{2\overline{t}}{(\overline{t}+\overline{\tau})s^3} \\ & - \frac{2(1-\widehat{f}(s))(1-\widehat{g}(s))}{(\overline{t}+\overline{\tau})s^4(1-\widehat{f}(s)\widehat{g}(s))}.\end{aligned} \tag{S20}$$

This formula can be inverted in the regime where Δ is much larger than $\overline{t}$ and $\overline{\tau}$, yielding

$$\langle \ell^2 \rangle_\Delta - \langle \ell \rangle_\Delta^2 \approx \frac{(\overline{\tau^2} - \overline{\tau}^2)\overline{t}^2 + (\overline{t^2} - \overline{t}^2)\overline{\tau}^2}{(\overline{t}+\overline{\tau})^3}\,\Delta + K. \tag{S21}$$

The linear growth of the variance of ℓ with Δ testifies that the continuous part $P_\Delta^{\rm cont}(\ell)$ satisfies an approximate central limit theorem at large Δ, where the measured displacement ℓ is the sum of a typically large number of elementary moves. The constant K, corresponding to the first correction to this limit, is given by a combination of moments of t and τ, which can be either positive or negative.

Another quantity which is used in the main text is the fraction $P_{\rm good}(\Delta)$ of the moves which are correctly sampled. These events correspond to the observation times θ_1 and θ_2 belonging to two consecutive rests surrounding the move under consideration, i.e., to sector 2 with $n = m+1$. It is also useful to introduce the normalised fraction of correctly sampled moves,

$$F_{\rm good}(\Delta) = \frac{P_{\rm good}(\Delta)}{1 - C_0(\Delta)}, \tag{S22}$$

where the denominator is nothing but the probability of measuring a non-zero displacement.

In the steady-state of the process, we obtain in Laplace space

$$\widehat{P}_{\rm good}(s) = \frac{\widehat{f}(s)(1-\widehat{g}(s))^2}{(\overline{t}+\overline{\tau})s^2}, \tag{S23}$$

whereas the Laplace transform of $C_0(\Delta)$ is given by (S16).

Exponential distributions

When the distributions of the move and rest durations are exponential, with respective parameters $a = 1/\overline{t}$ and $b = 1/\overline{\tau}$, i.e., $f(t) = a\,\mathrm{e}^{-at}$, $g(\tau) = b\,\mathrm{e}^{-b\tau}$, $\widehat{f}(s) = a/(s+a)$, $\widehat{g}(s) = b/(s+b)$, the above expressions simplify drastically, and so many observables can be evaluated in closed form.

Eqs. (S13), (S14) read

$$L(s,u) = \frac{(a+b)^2 + (a+b)s + au}{(a+b)(s^2 + (a+b+u)s + bu)}. \tag{S24}$$

We thus recover (S19), i.e.,

$$\langle \ell \rangle_\Delta = \frac{b}{a+b}\,\Delta, \tag{S25}$$

as well as the following expression

$$\langle \ell^2 \rangle_\Delta = \langle \ell \rangle_\Delta^2 + \frac{2ab}{(a+b)^3}\,\Delta + \frac{2ab}{(a+b)^4}\left(\mathrm{e}^{-(a+b)\Delta} - 1\right) \tag{S26}$$

for the second moment of the measured displacement. In this example, the constant $K = -2ab/(a+b)^4$ is negative.

With the notations of the main text, i.e., in terms of $\overline{\tau}, \overline{t}, \ell^*, v, \overline{\Delta}$, the first two moments read

$$\langle \ell^* \rangle = \frac{v\overline{\Delta}}{1 + \overline{\tau}/\overline{t}} \tag{S27}$$

and

$$\begin{aligned}\langle \ell^{*2} \rangle = {} & \frac{v^2\overline{\Delta}^2}{(1+\overline{\tau}/\overline{t})^2} + \frac{2v^2\overline{\tau}^2\overline{\Delta}}{\overline{t}(1+\overline{\tau}/\overline{t})^3} \\ & + \frac{2v^2\overline{\tau}^3}{\overline{t}(1+\overline{\tau}/\overline{t})^4}\left(\exp\left(-\frac{(\overline{t}+\overline{\tau})\overline{\Delta}}{\overline{t}\overline{\tau}}\right) - 1\right).\end{aligned} \tag{S28}$$

The full distribution $P_\Delta(\ell)$ can also be obtained in closed form. In a first step, performing the inverse transform of the expression (S24) from s to Δ, we obtain an expression for the Laplace transform

$$L_\Delta(u) = \int_0^\infty \mathrm{e}^{-u\ell} P_\Delta(\ell)\,\mathrm{d}\ell, \tag{S29}$$

namely

$$\begin{aligned}L_\Delta(u) = {} & \mathrm{e}^{-(a+b+u)\Delta/2} \\ & \times \left(\cosh R + \frac{(a+b)^2 + (a-b)u}{2(a+b)}\frac{\sinh R}{R}\Delta\right),\end{aligned} \tag{S30}$$

with

$$R = \frac{\Delta}{2}\sqrt{(a-b+u)^2 + 4ab}. \tag{S31}$$

In a second step, the expression (S30) can be inverse transformed from u to ℓ, yielding a result of the general form (S15), with

$$C_0(\Delta) = \frac{a}{a+b}\,\mathrm{e}^{-b\Delta}, \quad C_1(\Delta) = \frac{b}{a+b}\,\mathrm{e}^{-a\Delta} \tag{S32}$$

and

$$\begin{aligned}P_\Delta^{\mathrm{cont}}(\ell) = {} & \frac{2ab}{a+b}\,\mathrm{e}^{-a\ell - b(\Delta-\ell)} \\ & \times \left(I_0(x) + (a(\Delta-\ell) + b\ell)\frac{I_1(x)}{x}\right),\end{aligned} \tag{S33}$$

with

$$x = 2\sqrt{ab\ell(\Delta - \ell)}, \tag{S34}$$

and where I_0 and I_1 are modified Bessel functions.

The expression (S23) reads

$$\widehat{P}_{\rm good}(s) = \frac{a^2 b}{(a+b)(a+s)(b+s)^2}. \tag{S35}$$

We have therefore

$$\begin{aligned} P_{\rm good}(\Delta) = {} & \frac{a^2 b}{(a+b)(a-b)^2} \\ & \times \left(\mathrm{e}^{-a\Delta} + ((a-b)\Delta - 1)\mathrm{e}^{-b\Delta}\right) \end{aligned} \tag{S36}$$

and

$$F_{\rm good}(\Delta) = \frac{a^2 b}{(a-b)^2} \, \frac{\mathrm{e}^{-a\Delta} + ((a-b)\Delta - 1)\mathrm{e}^{-b\Delta}}{a + b - a\,\mathrm{e}^{-b\Delta}}. \tag{S37}$$

The normalised fraction $F_{\rm good}(\Delta)$ of correctly sampled moves starts growing as $(a\Delta)^2/2$ at small Δ, whereas it falls off exponentially at large Δ. It therefore reaches a non-trivial maximum $\hat{F}_{\rm good}$ for an optimal value $\hat{\Delta}$ of Δ (see Fig. 3 (bottom)). In the limit $b/a \to 0$, the maximal value $\hat{F}_{\rm good}$ reaches unity (see Fig. 4 (top)). It is attained for

$$\hat{\Delta} \approx \frac{2}{\sqrt{ab}}. \tag{S38}$$

It however drops from this perfect value very rapidly, with a square-root singularity of the form

$$\hat{F}_{\rm good} \approx 1 - 2\sqrt{\frac{b}{a}}. \tag{S39}$$

For $a = b$, the expression (S37) simplifies to

$$F_{\rm good}(\Delta) = \frac{(a\Delta)^2}{2(2\mathrm{e}^{a\Delta} - 1)}. \tag{S40}$$

The maximal value is already as small as $\hat{F}_{\rm good} \approx 0.14602$. It is reached for $\hat{\Delta} \approx 1.84141/a$.

NUMERICAL ANALYSIS

Random sampling and long-tailed pause distributions

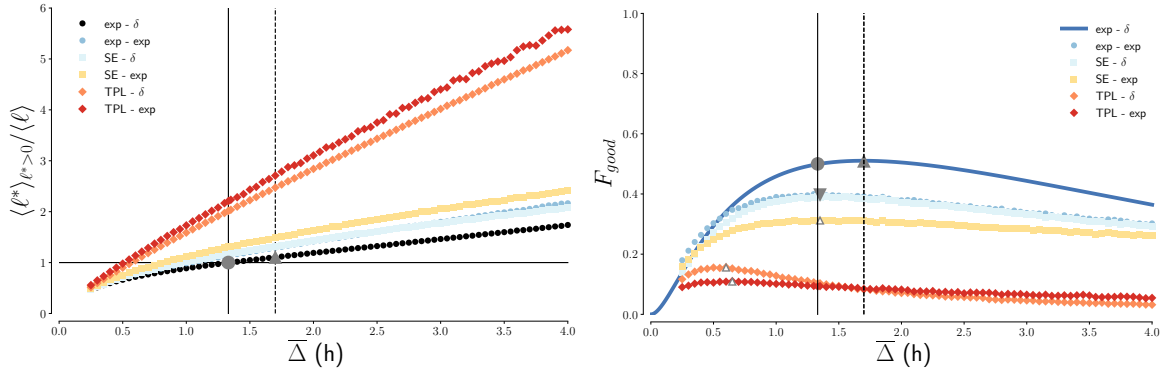


FIG. S5. **Effect of sampling (periodic and Poisson process) with different rest distribution.** Similar to what we depicted in Fig. 3 (represented here as black circles in the left panel and as a blue line in the right panel). In all scenarios studied, we fix $\bar{t} = 0.30$ h and consider different alternative forms for $P(\tau)$ (exponential (exp–) or Stretched Exponential (SE–) or Truncated Power Law (TPL–)). We use either constant (–δ) or exponentially distributed (–exp) sampling times. The grey circle represents the value associated with $\mathring{\Delta}$ and the filled grey up-triangle that of $\hat{\Delta}$ for the (exp – δ) scenario. **(Left)** We compare the average value $\langle \ell^* \rangle_{\ell^*>0}$ with $\langle \ell \rangle = \overline{\ell} = v\bar{t}$. All definitions of the optimal sampling time different from $\mathring{\Delta}$ (defined by imposing the identity between $\langle \ell^* \rangle$ and $\overline{\ell}$) are necessarily associated with values smaller than $\mathring{\Delta}$ for long-tailed rest distributions. For the SE case, a sampling time of $\approx$ 1h correctly estimates the average. If rests are distributed as a TPL, the large fraction of short rests leads to the concatenation of subsequent trips, even with this relatively short sampling time. This result illustrates a first incongruence in the work by Song et al. [4], where individual moves and rests are reconstructed with a sampling rate of 1 h, identifying a rest distribution (the TPL studied here) for which we predict that only half of the movements would be captured (because $\langle \ell^* \rangle / \overline{\ell} \approx 2$ and consequently $n^*/n \approx 1/2$). **(Right)** In all scenarios the fraction F_{good} of correctly sampled trips is lower than the value (51%) given by Eq. (S37) and studied extensively in this paper. In particular, trajectories with long rest times (SE– and TPL–) yield worse results than the peaked case (exp–). At the same time, using exponentially distributed sampling times (–exp) systematically yields worse results than constant sampling (–δ). The optimal value $\hat{\Delta}$ (filled grey up triangle and dashed line) over-estimates the position of the peak for long-tailed rests. A possible realistic scenario for human mobility is represented by a down triangle, where trajectories with SE rest distribution reach a maximum F_{good} of 39% when sampled with a constant $\overline{\Delta} = \check{\Delta} = 1.33$ h (80 min). If the pause distribution is TPL, F_{good} barely reaches the 15% mark, illustrating a second incongruence in the work by Song et al. [4]. (In the right panel, all triangle markers indicate the maximum of the associated curve.)

As stated in the main text, real sampling problems can be more complex than the idealised case defined by Eqs. (1), (2) and (3). In particular: (i) the rest time distribution can be broad; (ii) sampling times can be random variables; (iii) speed can be a random variable (travel durations have been seen to have a short-tailed distribution). We start by studying the effects of points (i) and (ii), while we discuss point (iii) in the following section.

Available data on rest durations suggest that their distribution is long-tailed. Different fits have been proposed for the latter distribution. Here we consider a Truncated Power Law (TPL), used for interpreting mobile phone data [4], and a Stretched Exponential (SE), used for interpreting private vehicles' parking times [7]. As for the sampling time, we can introduce randomness with an exponential distribution of inter-event times (Poisson process). Alternatively, if we want to represent the sampling process associated with communication, we use a power-law distribution with exponent -1 [3, 38]. Since this distribution is not integrable, it is necessarily defined on an interval between some $\Delta_{\min}$ and $\Delta_{\max}$.

In Figs. S5 and S6, we combine different rest distributions and sampling time distributions. We see that, in all the possible scenarios, F_{good} is below the best value $\hat{F}_{\text{good}} = 51\%$. In Fig. S5 (right) we show the fraction F_{good} for exponential (exp), TPL or SE rest time distributions and with sampling times Δ distributed as a delta function ($P(\Delta) = \delta(\Delta - \overline{\Delta})$) or an exponential distribution ($P(\Delta) = (1/\overline{\Delta}) \exp(-\Delta/\overline{\Delta})$), associated with a Poisson process where a sampling can happen at each moment with uniform rate. The solid blue line, together with the gray circle and the up triangle markers, represent the same information as displayed in Fig. 2 (d). All the other curves for $F_{\text{good}}(\overline{\Delta})$

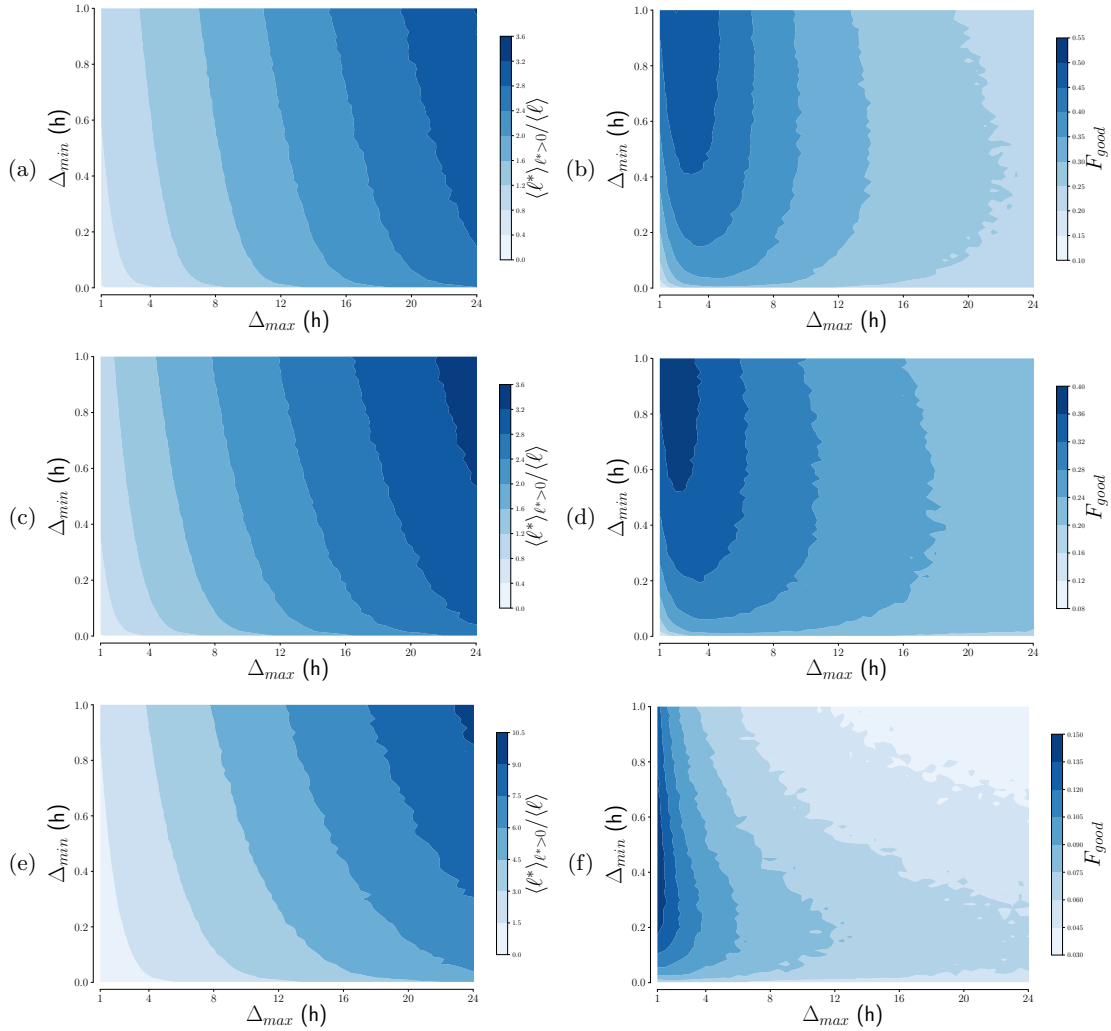


FIG. S6. **Effect of sampling with power-law inter-event times and different rest distributions.** Using sampling times distributed similarly to communication patterns, samplings get very bad for Exponential **(a, b)**, Stretched Exponential **(c, d)** or Truncated Power Law **(e, f)** rest distributions. The quality of the sampling depends on how we choose the minimal $\Delta_{\min}$ and maximal $\Delta_{\max}$ inter-event times. The left panels show the error in the estimate of the average $\langle \ell \rangle$. The right panels represent the fraction F_{good} of correctly sampled moves. With a very conservative choice of $\Delta_{\min} = 5$ min and $\Delta_{\max} = 12$ h, the value F_{good} for the exponential rest distribution drops from $\approx 51\%$ to $\approx 27\%$. For long-tailed rests we are limited to a maximal $F_{\text{good}} \approx 23\%$ (panel (d), Stretched Exponential) when sampling trajectories consistent with the rest and move times of humans with inter-event times consistent with mobile phones CDR data. The value drops below 6% in the same conditions for the Truncated Power Law (panel (f)).

are bell-shaped, their maxima have heights $< 51\%$, and these maxima are reached for a $\overline{\Delta}$ close to the expected $\hat{\Delta}$. Therefore, all variations introduced here only yield worse samplings. With a down triangle we show that, when the rest times are distributed as a Stretched Exponential, as suggested by car mobility data, the optimal sampling time would be $\check{\Delta} \approx 1.33$ h, but with only 39% of trips correctly sampled. If rest times have instead a TPL distribution, as estimated by mobile phone data [4], the sampling is very poor, with $F_{\text{good}} < 15\%$.

Panels (b) (d) and (f) of Fig. S6 correspond to a power-law sampling, where the final result of course depends on $\Delta_{\min}$ and $\Delta_{\max}$. F_{good} never goes over the optimal value found for constant sampling, which is naturally reached when $\Delta_{\min}$ and $\Delta_{\max}$ get close to $\hat{\Delta}$. However, since in an individual behaviour we can easily have a whole day without communication, the values of F_{good} we expect would be more of the order of the values reached on the right half of each contour graph. We consider as reference values (here as for our analysis of the GeoLife trajectory) $\Delta_{\min} = 5$ min

and $\Delta_{\max} = 12$ h. This very conservative choice yields $F_{\text{good}} = 27\%$ for the exponential rest distribution.

For both constant and random sampling times, the SE distribution yields values of F_{good} better than the TPL. This is expected, since larger values of $\overline{\tau}$ are associated with a better sampling (see Fig. 4 (top)), and confirms our choice of the characteristic times for vehicular mobility as the best-case scenario for our study. We therefore identify as 'optimistic' values for F_{good} for a realistic distribution of rests (the Stretched Exponential) the value 39% for constant sampling and 23% for power-law sampling. This last value is, again, computed with $\Delta_{\min} = 5$ min and $\Delta_{\max} = 12$ h.

In conclusion of this section, we cannot help but remark the incongruence between the rest time distribution identified from CDR data (the TPL studied here) and the sampling time of 1 h used to identify mobility patterns. For such a rest time distribution we predict $\langle \ell^* \rangle / \overline{\ell} \approx 2$ for a sampling time of 1 h. This ratio suggests that the trajectory is largely under-sampled, with only about half the trips correctly identified. Since rests would also be consequently under-counted, and thus over-estimated in duration, the rest time distribution estimated cannot be correct either.

Effect of velocity and spatial embedding

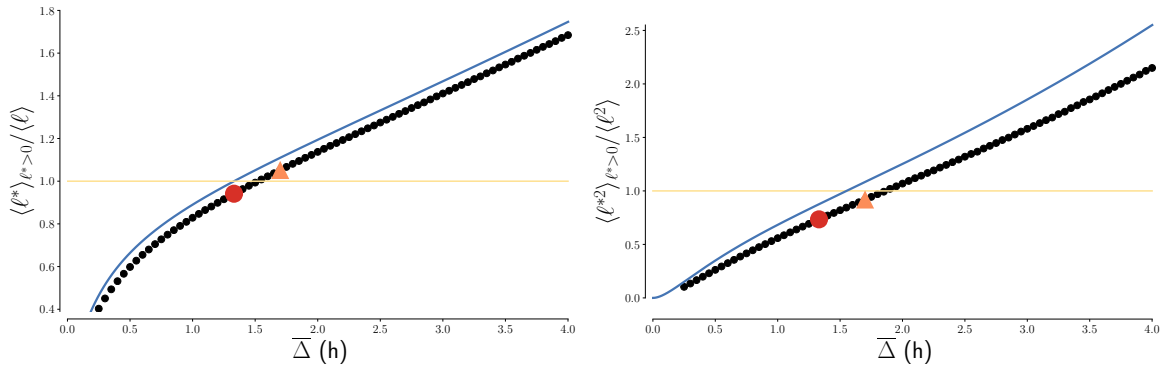


FIG. S7. **Moments of $P(\ell^*)$ with speed variability in the peaked scenario. (Left)** First moment. **(Right)** Second moment. We introduce variability in the speed distribution $P(v)$, using a random acceleration model [7]. Comparing with the analytical results of Eq. (8) and Eq. (S28) (blue lines), we can appreciate that realistic speeds introduce only minor changes in estimating the moments. The red circle indicates the result using the sampling time $\mathring{\Delta}$ that matches the mean values ($\langle \ell^* \rangle = \overline{\ell}$). The orange triangle shows the values associated with $\hat{\Delta}$, the sampling time maximizing F_{good}. The first sampling time $\mathring{\Delta}$ yields a slightly under-estimated first and second moment, with a deviation of 5% and 20% respectively, with respect to the moments of the displacement distribution in the simulated trajectories (yellow solid line).

We have seen that the optimal sampling times defined in the main text do not depend on the nature of $P(v)$, nor on the dimensionality of space (and thus of the vector speed v). Even when introducing these two factors, the fraction of moves that have been correctly identified remains unchanged. Nevertheless, the shape of the distribution $P(\ell^*)$ of sampled distances necessarily depends on speed and spatial embedding. We illustrate this by using, on top of the conditions set by Eqs. (1), (2) and (3), a random acceleration mobility model that induces a correlation between travel time and speed consistent with real data at a national level [7]. This model assumes that trips have an exponentially distributed durations with average value $\bar{t}$ and start at a base speed v_0. At any time of the first half of the trip the speed can be incremented of a value Δv with a constant probability π. Then, the second half of the trip is characterised by an identical process, but with descending speeds. The value of the four parameters of this model have been set to those used to reconstruct the speed distribution of cars in Italy ($\bar{t} = 0.30$h, $v_0 = 17.9$km/h, $\Delta v = 41.8$ km/h, $\pi = 2.12$ jumps/h) [7]. The results of Fig. S7 are to be compared with those of Fig. 3 (top) in the main text. In this case, the sampling time $\hat{\Delta}$ that is expected to match the first moment for constant speed under-estimates the average displacement by about 5%, although in general this deviation will depend on the type of movement studied..

In Fig. S7 (left) we show the effect of sampling trajectories with speed variability. We observe that, when the sampling time distribution is broad, we have the larger deviations from the original distribution. This same phenomenon can also be seen in Fig. S5 (left), where we show that the mean value $\langle \ell^* \rangle$ is significantly larger when the rest distribution is broad. The issue comes by the fact that with broad distributions we have several instances where the sampling time is large with respect to the average rest, and therefore more jumps are joined together. When

this happens, the dimensionality and the nature of the turning angle distribution becomes important. Fig. S8 would differ significantly by introducing this further element. Indeed, at low frequencies one would have to integrate over many re-orientations [29]. These re-orientations are neglected in our one-dimensional picture, that as a consequence over-estimates this sum. A possible multi-dimensional model could be the worm-like chain. On top of this, since the human tendency of returning home limits the space explored [3], the size of these summed displacements could be further reduced.

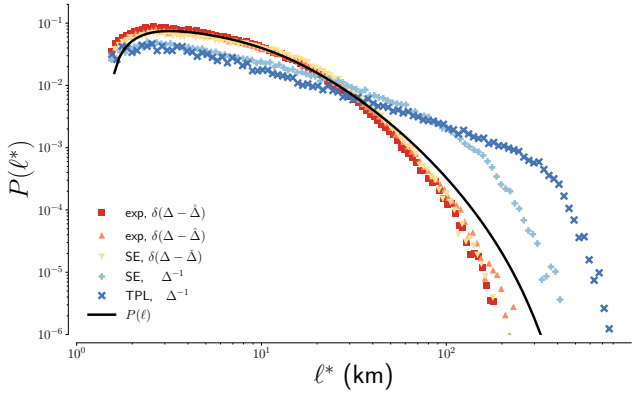


FIG. S8. **Sampling accelerated trajectories.** Using a random acceleration model [7], we study how the displacement distribution changes with the sampling. The result depends on the rest time distribution (exponential (exp), stretched exponential (SE) and truncated power law (TPL)) and on the sampling time distribution (peaked $\delta(\Delta - \check{\Delta})$ or with a long tail (PL)). The optimal sampling times ($\hat{\Delta}$ and $\mathring{\Delta}$, defined in the main text and $\check{\Delta}$ defined in the preceding section) give reasonably good results, while long-tailed sampling creates sizeable deviations in the displacement distribution.

CORRELATIONS BETWEEN CALLS AND RESTS IN EMPIRICAL SAMPLING

In the main text we study how the statistical properties of GPS trajectories of individuals are affected when we sample them with an inter-call distribution extracted from mobile phone data (CDR). It can however be objected that the times at which calls are made are correlated to the rests. Hence the fraction of correctly sampled trips might be higher than the one we obtain without correlations.

Moreover, more refined method of trajectory extraction are based on the idea of identifying stays (where the user is performing an activity) and pass-by's (locations where the call is made during a travel) [42]. Normally, one needs at least two calls in the same stay to identify it correctly as an activity. The goal of this method consists in filtering out calls made during rests. An ideal algorithm that perfectly identifies calls done during moves would then be equivalent to a perfect correlation between rests and calls.

For these reasons, we study the effect of correlations on the fraction of correctly sampled trajectories. We introduce a parameter p to quantify the extent of correlations between calls and rests. Any call that is performed during a move is excluded with probability p. When $p = 0$ calls and rests are de-correlated and calls can happen at any point in time, while they are perfectly correlated, and all calls necessarily happen when at rest, when $p = 1$.

The results are presented on Fig. S9: when $p = 0$ we find the same value $F_{\text{good}} = 11\%$ as presented in the main text. When $p = 1$ we find $F_{\text{good}} = 16\%$, which is indeed better than without correlations, yet not large enough to invalidate our conclusions, nor to support the current 'stay point identification' method as sufficient for reconstructing mobility patterns.

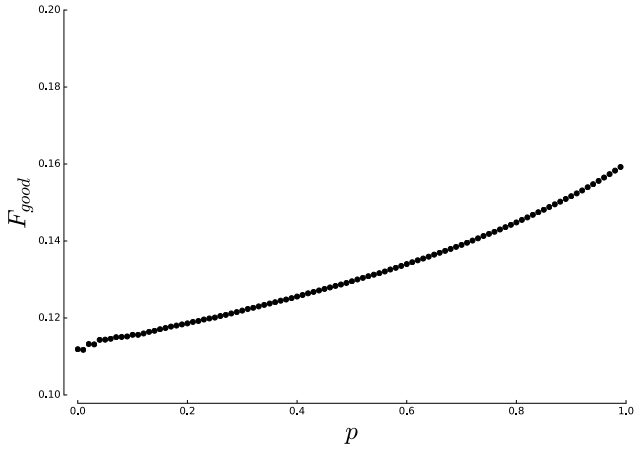


FIG. S9. **Effect of correlations on the fraction of correctly sampled trajectories.** We quantify the effect of correlations between calls and rests on the fraction of rightly sampled trips F_{good}. When $p = 0$ calls and rests are uncorrelated and we find $F_{\text{good}} = 11\%$, as shown in the main text. When $p = 1$ calls only happen during rests and we find $F_{\text{good}} = 16\%$. Our conclusions therefore hold disregarding of whether there are correlations between calls and movements, or if this correlation is induced by filtering out calls done during moves.