RESEARCH

Supplementary material for "An end-to-end statistical process with mobile network data for Official Statistics"

David Salgado[1,2*†], Luis Sanguiao[1†], Bogdan Oancea[3†], Sandra Barragán[1†] and Marian Necula[4†]

*Correspondence: david.salgado.fernandez@ine.es
[1]Dept. Methodology and Development of Statistical Production, Statistics Spain (INE), Paseo de la Castellana, 183, Madrid, Spain
Full list of author information is available at the end of the article
†The views expressed in this working paper are those of the authors and do not necessarily reflect the views of their affiliating institutions.

1 Introduction

A more complete list of published articles with statistical uses of mobile network data is provided in the references [1–73]. A great deal of unpublished work can also be found. We specifically recommend the conference series NetMob (`http://www.netmob.org`).

2 Data description

2.1 Scenario data

All input and output data for the simulator can be found at [URL]. Input data for the simulator are specified as xml files (`persons.xml`, `simulation.xml`, `antennas.xml`) and a wkt file for the irregular polygon (territory map). For the scenario used in the article, we have selected:

- `simulation.xml` contains general parameters for the simulation: see table 1.
- `persons.xml` contains general parameters for the displacement patterns: see table 2.
- `antennas.xml` contains parameters to configure each antenna: see table 3. We have configured 70 antennas with the marginal distributions included in table 4.

Output data from the simulator are obtained in csv format (we comment only the basic ones):

- `persons.csv` contains the real evolution of each individual, i.e. the ground truth. For each time instant t and each individual k, the simulator records the position coordinates x and y, the tile and the device(s) carried by the individual.
- `SignalMeasure_MNO1.csv` contains for each antenna the RSS in each tile of the reference grid.
- `AntennaInfo_MNO_MNO1.csv` contains the time sequence of connections. For each time instant t and each device, the simulator records the antenna to which it is connected, its true position coordinates (x, y) and tile, and a network event code for the type of connection.

For details about other parameters and files see [74].

3 Geolocation of mobile devices

3.1 Model construction

We include the mathematical details to compute the posterior location probabilities from the input data. This is conducted in steps:

1 Time discretization and padding.
2 Construction of the transition model.
3 Construction of the emission model.
4 Construction of the initial state (prior) distribution.
5 Computation of the likelihood function.
6 Parameter estimation (likelihood maximization).
7 Application of the forward-backward algorithm.

3.1.1 Time discretization and padding

We shall work in discrete times. To do this we need to relate three parameters, namely (i) the tile dimension l (we assume a square grid for simplicity), (ii) the time increment Δt between two consecutive time instants, and (iii) an upper bound $v_{\max}$ for the velocity of the individuals in the population. As we argued in the main text, we impose that in the time interval Δt, the device d at most can displace from one tile to an adjacent tile. Under this condition, we can trivially set $\Delta t \lesssim \frac{l}{v_{\max}}$. For example, if $v_{\max} = 150\text{km/h} \approx 42\text{ms}^{-1}$, then $\Delta t \lesssim \frac{100}{42} \approx 2\text{s}$. Conversely, if the time increment Δt is fixed, then the maximum distance in terms of the number of tiles will be $\lceil \frac{v_{\max} \cdot \Delta t}{l} \rceil$, which expresses the number of time instants to insert in the time sequence to guarantee the maximum one-tile distance restriction.

If in the dataset the device d is detected at longer time periods, e.g. once in a minute, then we artificially introduce missing values at intervals Δt between every two observed values. This artificial non-response allows us to work with parsimonious models easier to estimate instead of using more complex transition matrices.

Notice that we have used an a priori value for $v_{\max}$, but we can also possibly make an estimation using the observed values $\mathbf{E}_d(t)$ and geometrical considerations about the respective coverage areas and their mutual distance.

Additionally, each observed time instance t is approximated to its closest multiple integer of Δt. Thus, we will have as input data a sequence of time instants at multiples $t_n = \Delta t \cdot n, (n \geq 0)$ and a randomly alternate sequence of missing values and of observed event variables $\mathbf{E}_{t_n}$ (hereafter for ease of notation we drop out any reference to mobile device d since we are only focusing on one device).

3.1.2 Construction of the transition model

Now we specify a model for the transition between tiles (states) $\{T_i\}_{i=1,\dots,N_T}$. For ease of explanation and notation, let us change the notation of each tile T_i to a two-dimensional index $T_{(i,j)}$. Accordingly, each tile will be specified in this section by a pair of integer coordinates. The correspondence between both enumerations is arbitrary, but fixed once it has been chosen. We again assume time homogeneity for simplicity. Thus, $\mathbb{P}\left(T_{(r,s)} \middle| T_{(i,j)}\right)$ will denote $\mathbb{P}\left(T_{(r,s)}(t_n + \Delta t) \middle| T_{(i,j)}(t_n)\right)$ for any $t_n = 0, 1, \dots$ We assume a square regular grid for simplicity.

Now, we make use of our preceding imposition by which an individual can at most reach an adjacent tile in time Δt. Thus,

$$\mathbb{P}\left(T_{(r,s)}\middle|T_{(i,j)}\right) = 0 \qquad \max\{|r-i|, |s-j|\} \geq 2, \qquad r,s,i,j = 1,\ldots,\sqrt{N_T}. \tag{1a}$$

Now, we assume that we have no further auxiliary information to model these transitions and impose rectangular isotropic conditions:

$$\begin{aligned} \mathbb{P}\left(T_{(i\pm1,j)}\middle|T_{(i,j)}\right) = \mathbb{P}\left(T_{(i,j\pm1)}\middle|T_{(i,j)}\right) &= \theta_1 \qquad i,j = 1,\ldots,\sqrt{N_T}, &(1b)\\ \mathbb{P}\left(T_{(i\pm1,j\pm1)}\middle|T_{(i,j)}\right) &= \theta_2 \qquad i,j = 1,\ldots,\sqrt{N_T}. &(1c)\end{aligned}$$

The last set of conditions is row-stochasticity:

$$\begin{aligned} \sum_{r,s=1}^{N_T} \mathbb{P}\left(T_{(r,s)}\middle|T_{(i,j)}\right) &= 1, \qquad i,j = 1,\ldots,\sqrt{N_T}, &(1d)\\ \mathbb{P}\left(T_{(r,s)}\middle|T_{(i,j)}\right) &\geq 0, \qquad i,j,r,s = 1,\ldots,\sqrt{N_T}. \end{aligned}$$

Now back to the original notation for tiles and using the usual notation for the transition matrix $A = [a_{ij}]$, with $a_{ij} = \mathbb{P}\left(T_{jt}\middle|T_{it}\right)$, conditions (1) amounts to having a highly sparse transition matrix A with up to 4 terms equal to θ_1 and θ_2 (each) per row and diagonal entries guaranteeing row-stochasticity.

For the generic case of a square grid with size N_T, we have

$$A(\theta_1,\theta_2) = \begin{bmatrix} D_1(\theta_1,\theta_2) & M(\theta_1,\theta_2) & O & O & \cdots & O \\ M(\theta_1,\theta_2) & D_2(\theta_1,\theta_2) & M(\theta_1,\theta_2) & O & \cdots & O \\ O & M(\theta_1,\theta_2) & D_2(\theta_1,\theta_2) & M(\theta_1,\theta_2) & \cdots & O \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ O & O & O & M(\theta_1,\theta_2) & D_2(\theta_1,\theta_2) & M(\theta_1,\theta_2) \\ O & O & O & O & M(\theta_1,\theta_2) & D_1(\theta_1,\theta_2) \end{bmatrix}, \tag{2}$$

where

$$D_1(\theta_1,\theta_2) = \begin{pmatrix} 1-2\theta_1-\theta_2 & \theta_1 & 0 & 0 & \cdots & 0 \\ \theta_1 & 1-3\theta_1-2\theta_2 & \theta_1 & 0 & \cdots & 0 \\ 0 & \theta_1 & 1-3\theta_1-2\theta_2 & \theta_1 & \cdots & 0 \\ 0 & 0 & \theta_1 & 1-3\theta_1-2\theta_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1-2\theta_1-\theta_2 \end{pmatrix}_{N_T \times N_T},$$

$$D_2(\theta_1,\theta_2) = \begin{pmatrix} 1-3\theta_1-2\theta_2 & \theta_1 & 0 & 0 & \cdots & 0 \\ \theta_1 & 1-4\theta_1-4\theta_2 & \theta_1 & 0 & \cdots & 0 \\ 0 & \theta_1 & 1-4\theta_1-4\theta_2 & \theta_1 & \cdots & 0 \\ 0 & 0 & \theta_1 & 1-4\theta_1-4\theta_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1-3\theta_1-2\theta_2 \end{pmatrix}_{N_T \times N_T},$$

$$M(\theta_1,\theta_2) = \begin{pmatrix} \theta_1 & \theta_2 & 0 & 0 & \cdots & 0 \\ \theta_2 & \theta_1 & \theta_2 & 0 & \cdots & 0 \\ 0 & \theta_2 & \theta_1 & \theta_2 & \cdots & 0 \\ 0 & 0 & \theta_2 & \theta_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \theta_1 \end{pmatrix}_{N_T \times N_T},$$

$$O = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 \end{pmatrix}_{N_T \times N_T}.$$

Notice that $A(\theta_1, \theta_2)$ fulfills all restrictions (1). Indeed, in our proposed implementation, in order to seek future generalization, we will work with a generic block-tridiagonal matrix (2), where the restrictions (1a) leading to 0 have been included, and complemented with the rest of restrictions (1b)-(1d) in matrix form. Thus, we write $C \cdot \text{vec}(\tilde{A}) = \mathbf{b}$, where $\text{vec}(\tilde{A})$ stands for the non-null elements of A in vector form. The rows of $[C\ \mathbf{b}]$ encode each of the restrictions (1b), (1c), and (1d). For example, $a_{12} = \theta_1$ and $a_{21} = \theta_1$ produce a row like this

$$C_i \cdot \text{vec}(\tilde{A}) = \begin{bmatrix} \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & 0 & 1 & 0 & \cdots & 0 & -1 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \cdot \left(\cdots \cdot a_{12} \cdot \cdots \cdot a_{21} \cdot \cdots \right)^T = b_i = 0. \quad (3)$$

Thus, in our software implementation to test this proposal, we have considered a block-tridiagonal matrix like (2) together with a set of linear restrictions of the form $C \cdot \text{vec}\left(\tilde{A}\right) = \mathbf{b}$.

3.1.3 Construction of the emission model

The emission model is specified by the HMM emission probabilities $b_{ik} = \mathbb{P}\left(\mathbf{E}_{t_n} = \mathbf{E}_k \middle| T_{t_n} = i\right)$, where $\mathbf{E}_k$ is a possible value for the observed event variables

$\mathbf{E}_{t_n}$ and i denotes the tile index. We assume time homogeneity. This conditional probability is computed using the radio wave propagation model of our choice:

$$\begin{aligned} b_{ik}^{\mathrm{RSS}} &\propto \mathrm{RSS}(d(\mathbf{E}_k, T_i)) && (4) \\ b_{ik}^{\mathrm{SDM}} &\propto \mathrm{SDM}(d(\mathbf{E}_k, T_i)), && (5) \end{aligned}$$

where $d(\mathbf{E}_k, T_i)$ stands for the distance between the antenna generating the event $\mathbf{E}_k$ and tile T_i. The proportional constant is fixed to normalize the probability functions.

Up to this point we have as input data the sequence of observed and missing values $a_{t_n} \in \{0, 1, \ldots, N_A\}$ for $t_n = 0, 1, \ldots, T$. We already have the emission matrix B, too.

3.1.4 Construction of the initial state (prior) distribution

For illustrative purposes, we consider two choices: (i) uniform prior, i.e. $\pi_i^{\mathrm{uniform}} = \frac{1}{N_T}$ and (ii) $\pi_i^{\mathrm{network}} \propto \sum_k (\mathrm{RSS}(d(\mathbf{E}_k, T_i)))$ (where RSS is expressed in watts) or $\pi_i^{\mathrm{network}} \propto \sum_k (\mathrm{SDM}(d(\mathbf{E}_k, T_i)))$, depending on the emission model.

3.1.5 Computation of the likelihood

The likelihood is trivially computed using the numerical proviso of setting emission probabilities equal to 1 when there is a missing value in the observed variables (e.g. due to time padding). The general expression for the likelihood is

$$\begin{aligned} L(\mathbf{E}) &= \sum_{i_0=1}^{N_T} \cdots \sum_{i_T=1}^{N_T} \mathbb{P}\left(T_{t_0} = i_0\right) \prod_{n=1}^{N} \mathbb{P}\left(T_{t_n} = i_n | T_{t_{n-1}} = i_{n-1}\right) \mathbb{P}\left(E_{t_n} \middle| T_{t_n} = i_n\right) \\ &= \sum_{i_0=1}^{N_T} \cdots \sum_{i_T=1}^{N_T} \mathbb{P}\left(T_{t_0} = i_0\right) \prod_{n=1}^{N} a_{i_{n-1} i_n}(\boldsymbol{\theta}) b_{i_n k_{t_n}} && \text{(6a)} \end{aligned}$$

Notice that the emission probabilities only contribute numerically providing no parameter whatsoever to be estimated.

3.1.6 Parameter estimation

The estimation of the unknown parameters $\boldsymbol{\theta}$ is conducted maximizing the likelihood. The restrictions coming from the transition model (1) makes the optimization problem not trivial. Notice that the EM algorithm is not useful. Instead, we provide a taylor-made solution seeking for future generalizations with more realistic choices of transition probabilities incorporating land use information. Formally, the optimization problem is given by:

$$\begin{aligned} \max \quad & h(\mathbf{a}) \\ \text{s.t.} \quad & C \cdot \mathbf{a} = \mathbf{b} && (7) \\ & a_k \in [0, 1], \end{aligned}$$

where $\mathbf{a}$ stands for the nonnull entries of the transition probability matrix A, the objective function $h(\mathbf{a})$ is derived from the likelihood L expressed in terms of the nonnull entries of the transition matrix A, and the system $C \cdot \mathbf{a} = \mathbf{b}$ expresses the sets of restrictions from the transition model (1) not involving null rhs terms (restrictions (1b), (1c), and (1d)).

Let us quantify the number of variables and restrictions in order to propose an abstract procedure possibly generalized to other situations. We illustrate this procedure for a square regular grid of size N_T. The number of zeroes in the transition matrix A can be computed as follows:

- There exist 4 rows in A corresponding to the 4 vertices in the grid. Each of these rows contains $N_T - 4$ zeroes.
- There exist 4 sets of $\sqrt{N_T} - 2$ rows in A corresponding to boundary tiles not being vertices. Each of these rows contains $N_T - 6$ zeroes.
- There exist $(\sqrt{N_T} - 2)^2$ rows in A corresponding to this same number of inner tiles. Each of these rows contains $N_T - 9$ zeroes.

Thus, the total number of zeroes in A is given by $4 \times (N_T - 4) + 4 \times (\sqrt{N_T} - 2) \times (N_T - 6) + (\sqrt{N_T} - 2)^2 \times (N_T - 9) = N_T^2 - 9 \cdot N_T + 12\sqrt{N_T} - 4$. The number of non-null components of $\mathbf{a}$ in problem (7) is $d = 9 \cdot N_T - 12\sqrt{N_T} + 4$.

The number of restrictions n_r not involving zeroes depends very sensitively on the particular transition model chosen for the displacements. In the rectangular isotropic model considered above, we need to identify the number of entries (i) equal to θ_1, (ii) equal to θ_2, and (iii) in the diagonal (thus guaranteeing the row-stochasticity restriction). Using the same counting procedure as above, the number of entries equal to θ_1 will be given by $4 \times 2 + 4 \times (\sqrt{N_T} - 2) \times 3 + (\sqrt{N_T} - 2)^2 \times 4 = 4 \cdot N_T - 4\sqrt{N_T}$. Since θ_1 is a free parameters we get $4 \cdot N_T - 4\sqrt{N_T} - 1$ rows. For θ_2, we get $4 \times (\sqrt{N_T} - 1)^2 - 1$ rows. From the row-stochasticity restriction we get N_T rows. Thus, the matrix C will have dimensions $n_r \times d$, with $n_r = 4 \cdot N_T - 4\sqrt{N_T} - 1 + 4 \times (\sqrt{N_T} - 1)^2 - 1 + N_T = 9 \cdot N_T - 12\sqrt{N_T} + 2$. Notice that $d - n_r = 2$, as expected, since we have two free parameters θ_1 and θ_2.

The abstract optimization problem is thus

$$
\begin{aligned}
\max \quad & h(\mathbf{a}) \\
\text{s.t.} \quad & C \cdot \mathbf{a} = \mathbf{b} \ , \\
& \mathbf{a} \in [0,1]^d,
\end{aligned}
\tag{8}
$$

where $C \in \mathbb{R}^{n_r \times d}$ and $\mathbf{b} \in \mathbb{R}^d$. The objective function $h(\mathbf{a})$ is indeed a polynomial in the non-null entries $\mathbf{a}$. This problem can be further simplified using the matrix QR decomposition. Write $C = Q \cdot R$, where Q is an orthogonal matrix of dimensions $n_r \times n_r$ and R is an upper triangular matrix of dimensions $n_r \times d$. Then we can rewrite the linear system as $R \cdot \mathbf{a} = Q^T \cdot \mathbf{b}$ and we can linearly solve variables $a_1, \ldots, a_{n_r}$ in terms of variables $a_{n_r+1}, \ldots, a_d$:

$$
\begin{pmatrix} a_1 & \cdots & a_{n_r} \end{pmatrix}^T = \tilde{C}_{n_r \times (d-n_r)} \begin{pmatrix} a_{n_r+1} & \cdots & a_d \end{pmatrix}^T .
$$

The system (8) then reduces to

$$\begin{aligned} \max \quad & \tilde{h}(a_{n_r+1}, \ldots, a_d) \\ \text{s.t.} \quad & 0 \leq \tilde{C} \cdot \begin{pmatrix} a_{n_r+1} & \cdots & a_d \end{pmatrix}^T \leq 1. \end{aligned} \tag{9}$$

In our current software implementation we resort to general-purpose optimizers. It remains for future work to find an optimised algorithm to solve (9). The solution $\mathbf{a}^*$ to problem (8) will be introduced in the transition probability matrix, which will thus be denoted by $\widehat{A}$.

3.1.7 Application of the forward-backward algorithm

Once the HMM has been fitted, we can readily apply the well-known forward-backward algorithm [see e.g. 75] to compute the target location probabilities γ_{it} and γ_{tij}. No novel methodological content is introduced at this point. For our implementation, we have used the scaled version of the algorithm (see [75]).

3.1.8 Software implementation

To carry out the computation described above upon the synthetic scenario generated by the network event data simulator we have used the prototyping R package called `destim` developed for these purposes by the authors [76]. This package contains a specific implementation of the rectangular geolocation model described in the preceding sections.

3.2 Model evaluation

The center of location probabilities and the root mean squared dispersion can be obtained naturally from a bias-variance decomposition of a mean squared distance. Let us denote by $\mathbf{R}_{dt} \in \{\mathbf{r}_i^{(c)}\}_{i=1,\ldots,N_T}$ the random vector for the position of a device according to the distribution of posterior location probabilities γ_{dti}. Let us shortly denote $\bar{\mathbf{R}}_{dt} \equiv \mathbb{E}\mathbf{R}_{dt} = \sum_{i=1}^{N_T} \gamma_{dti}\mathbf{r}_i^{(c)}$. Let us also denote the true position of device d at time t by $\mathbf{r}_{dt}^*$. Then, we can decompose

$$\begin{aligned} \text{msd}_{dt} \equiv \mathbb{E}\|\mathbf{R}_{dt} - \mathbf{r}_{dt}^*\|^2 &= \mathbb{E}\|(\mathbf{R}_{dt} - \bar{\mathbf{R}}_{dt}) + (\bar{\mathbf{R}}_{dt} - \mathbf{r}_{dt}^*)\|^2 \\ &= \mathbb{E}\left[\langle \mathbf{R}_{dt} - \bar{\mathbf{R}}_{dt}, \mathbf{R}_{dt} - \bar{\mathbf{R}}_{dt}\rangle\right] + \\ &\quad 2 \cdot \mathbb{E}\left[\langle \mathbf{R}_{dt} - \bar{\mathbf{R}}_{dt}, \bar{\mathbf{R}}_{dt} - \mathbf{r}_{dt}^*\rangle\right] + \\ &\quad \mathbb{E}\left[\langle \bar{\mathbf{R}}_{dt} - \mathbf{r}_{dt}^*, \bar{\mathbf{R}}_{dt} - \mathbf{r}_{dt}^*\rangle\right] \\ &= \text{rmsd}_{dt}^2 + \text{b}_{dt}^2. \end{aligned} \tag{10}$$

This decomposition motivates the definition of the figures of merit proposed in the main text. We can also compare directly the mean squared distance (see figure 1). The overall performance is similar for the four models.

4 Device duplicity

4.1 The double-device emission model

To apply formulas for the computation of the device duplicity probabilities we need to compute the likelihood for the HMM model described above for each device separately and for each pair of devices according to figure 7 in the main text. To do this, we just need to have a new emission model producing a double event data sequence. The emission probabilities in this augmented model are computed using the original emission probabilities:

$$\mathbb{P}\left(\mathbf{E}_{dt}, \mathbf{E}_{d't} \middle| T_{dt}, \mathbf{I}^{\mathrm{aux}}\right) = \mathbb{P}\left(\mathbf{E}_{dt} \middle| T_{dt}, \mathbf{I}^{\mathrm{aux}}\right) \cdot \mathbb{P}\left(\mathbf{E}_{d't} \middle| T_{d't}, \mathbf{I}^{\mathrm{aux}}\right). \tag{11}$$

Once these emission probabilities are computed, the computation of the likelihoods $\ell_{dd'}$ runs similar to the single-device case.

The computation depends on prior choice of the parameters λ_d, i.e. the ratio between the prior probability of no duplicity to the prior probability of duplicity. For the computation in the main text, this was chosen according to the parameters in the network event data simulator. In practice, this is not the case, but the MNO can provide a prior estimation of the number of subscribers with more than one device. In any case, we run the computation of $p_d^{(2)}$ for all d and checked the number of true/false positive/negative cases obtained. This is represented in figure 2, where we observed that around the chosen value, the classification is robust.

4.2 Software implementation

To carry out the computation described above upon the synthetic scenario generated by the network event data simulator we have used the prototyping R package called `deduplication` developed for these purposes by the authors [77]. This package implements the computation of the device duplicity probabilities described in the main text, including the computation of the double-device emission model for the underlying HMM. This package contains another two deduplication procedures based on pairwise comparisons and trajectory comparisons. In this work we have included only the alternative producing the best disambiguating method on our scenario.

5 Statistical filtering

To apply the proposed trajectory indicators to the synthetic scenario generated by the network event data simulator we have profusely used the R package `trajr` [78], with slight modifications on some functions to adequate to our trajectories.

6 Aggregation of individuals detected by a network

The core of the aggregation module is the generation of random multidimensional variates according to the Poisson-multinomial distribution as a sum of categorical (multinoulli) variables. This is directly implemented in the prototyping R package called `aggregation` developed for these purposes by the authors [79]. This package takes as input both the posterior location probabilities γ_{dti}, the device duplicity probabilities $p_d^{(2)}$ for all devices d, and the spatial aggregation of tiles i into larger territorial units and produce n random multidimensional variates according to the Poisson-multinomial distribution defined by equation (17) in the main text. The package also implements the similar computation for the origin-destination matrix according to equation (21) in the main text.

7 Inference

The different models proposed for the inference module have been directly implemented using standard distributions in base R and package `extraDistr` [80], except for the continuous mixtures integrating the full hierarchy of levels for the observation and/or the state processes. The credible interval computations, both for the inference and the aggregation module, have been carried out using the R package `bayestestR` [81]. All credible intervals included in this work are high-density intervals [see e.g. 82].

Declaration

Availability of data and materials

Data, scripts, and source code are freely available at [URL].

Competing interests

The authors declare that they have no competing interests.

Funding

This work is part of ongoing projects at Statistics Spain (INE) and Statistics Romania (INS) in joint collaboration with the European Statistical System under Grant Agreement Number 847375-2018-NL-BIGDATA (ESSnet on Big Data II).

Authors' contributions

All authors have contributed equally.

Acknowledgements

The authors acknowledge M.Á. Martínez-Vidal, S. Lorenzo, M. Suárez-Castillo, R. Radini, T. Tuoto, M. Offermans, M. Tennekes, S. Hadam, and F. Ricciato for invaluable insights and debates.

Author details

[1]Dept. Methodology and Development of Statistical Production, Statistics Spain (INE), Paseo de la Castellana, 183, Madrid, Spain. [2]Dept. Statistics and Operations Research, Complutense University of Madrid, Plaza de las Ciencias, 3, Madrid, Spain. [3]Dept. Business Administration, University of Bucharest, 90 Panduri Street, Bucharest, Romania. [4]Dept. Innovative Tools in Official Statistics, Statistics Romania (INS), 16 Libertatii Bvd, Bucharest, Romania.

References

1. Cáceres, N., Wideberg, J.P., ıtez, F.G.B.: Deriving origin–destination data from a mobile phone network. IET Intelligent Transport Systems **1**(1), 15 (2007). doi:10.1049/iet-its:20060020
2. Ahas, R., Aasa, A., Ülar Mark, Pae, T., Kull, A.: Seasonal tourism spaces in Estonia: Case study with mobile positioning data. Tourism Management **28**(3), 898–910 (2007). doi:10.1016/j.tourman.2006.05.010
3. Candia, J., González, M.C., Wang, P., Schoenharl, T., Madey, G., Barabási, A.-L.: Uncovering individual and collective human dynamics from mobile phone records. Journal of Physics A: Mathematical and Theoretical **41**(22), 224015 (2008). doi:10.1088/1751-8113/41/22/224015
4. González, M.C., Hidalgo, C.A., Barabási, A.-L.: Understanding individual human mobility patterns. Nature **453**(7196), 779–782 (2008). doi:10.1038/nature06958
5. Farrahi, K., Gatica-Perez, D.: Daily routine classification from mobile phone data. In: Popescu-Belis, A., Stiefelhagen, R. (eds.) Machine Learning for Multimodal Interaction. Lecture Notes in Computer Science, vol. 5237, pp. 173–184. Springer, Berlin Heidelberg (2008). doi:10.1007/978-3-540-85853-9_16
6. Ahas, R., Aasa, A., Roose, A., Mark, U., Silm, S.: Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. Tourism Management **29**(3), 469–486 (2008). doi:10.1016/j.tourman.2007.05.014
7. Eagle, N., Pentland, A., Lazer, D.: Inferring friendship network structure by using mobile phone data. Proceedings of the National Academy of Sciences **106**(36), 15274–15278 (2009). doi:10.1073/pnas.0900282106
8. Ahas, R., Silm, S., Järv, O., Saluveer, E., Tiru, M.: Using mobile positioning data to model locations meaningful to users of mobile phones. Journal of Urban Technology **17**(1), 3–27 (2010). doi:10.1080/10630731003597306

9. Farrahi, K., Gatica-Perez, D.: Probabilistic mining of socio-geographic routines from mobile phone data. IEEE Journal of Selected Topics in Signal Processing **4**(4), 746–755 (2010). doi:10.1109/jstsp.2010.2049513
10. Sevtsuk, A., Ratti, C.: Does urban mobility have a daily routine? learning from the aggregate data of mobile networks. Journal of Urban Technology **17**(1), 41–60 (2010). doi:10.1080/10630731003597322
11. Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Martonosi, M., Rowland, J., Varshavsky, A.: Identifying important places in people's lives from cellular network data. In: Lecture Notes in Computer Science, pp. 133–151. Springer, ??? (2011). doi:10.1007/978-3-642-21726-5_9
12. Becker, R.A., Cáceres, R., Hanson, K., Loh, J.M., Urbanek, S., Varshavsky, A., Volinsky, C.: A Tale of One City: Using Cellular Network Data for Urban Planning. IEEE Pervasive Computing **10**(4), 18–26 (2011). doi:10.1109/MPRV.2011.44
13. Steenbruggen, J., Borzacchiello, M.T., Nijkamp, P., Scholten, H.: Mobile phone data from GSM networks for traffic parameter and urban spatial pattern assessment: a review of applications and opportunities. GeoJournal **78**(2), 223–243 (2011). doi:10.1007/s10708-011-9413-y
14. Couronné, T., Smoreda, Z., Raimond, A.-M.O.: Chatty mobiles: individual mobility and communication patterns. CoRR **abs/1301.6553** (2011)
15. Soto, V., Frias-Martinez, V., Virseda, J., Frias-Martinez, E.: Prediction of socioeconomic levels using cell phone records. In: User Modeling, Adaption and Personalization, pp. 377–388. Springer, ??? (2011). doi:10.1007/978-3-642-22362-4_35
16. Blumenstock, J.E.: Inferring patterns of internal migration from mobile phone call records: evidence from rwanda. Information Technology for Development **18**(2), 107–125 (2012). doi:10.1080/02681102.2011.643209
17. Cáceres, R., Rowland, J., Small, C., Urbanek, S.: Exploring the use of urban greenspace through cellular network activity. In: The Second Workshop on Pervasive Urban Applications (PURBA), in Conjunction with Pervasive (2012). http://www.kiskeya.net/ramon/work/pubs/purba12.pdf
18. Phithakkitnukoon, S., Smoreda, Z., Olivier, P.: Socio-geography of human mobility: A study using longitudinal mobile phone data. PLoS ONE **7**(6), 39253 (2012). doi:10.1371/journal.pone.0039253
19. Palmer, J.R.B., Espenshade, T.J., Bartumeus, F., Chung, C.Y., Ozgencil, N.E., Li, K.: New approaches to human mobility: Using mobile phones for demographic research. Demography **50**(3), 1105–1128 (2012). doi:10.1007/s13524-012-0175-z
20. Ferrari, L., Mamei, M., Colonna, M.: Discovering events in the city via mobile network analysis. Journal of Ambient Intelligence and Humanized Computing **5**(3), 265–277 (2012). doi:10.1007/s12652-012-0169-0
21. Smoreda, Z., Olteanu-Raimond, A.-M., Couronné, T.: Spatiotemporal data from mobile phones for personal mobility assessment. In: Zmud, J., Lee-Gosselin, M., Munizaga, M., Carrasco, J.A. (eds.) Transport Survey MethodsM Best Practice for Decision Making. Emerald, ??? (2013)
22. Becker, R., Volinsky, C., Cáceres, R., Hanson, K., Isaacman, S., Loh, J.M., Martonosi, M., Rowland, J., Urbanek, S., Varshavsky, A.: Human mobility characterization from cellular network data. Communications of the ACM **56**(1), 74 (2013). doi:10.1145/2398356.2398375
23. Calabrese, F., Diao, M., Lorenzo, G.D., Ferreira, J., Ratti, C.: Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. Transportation Research Part C: Emerging Technologies **26**, 301–313 (2013). doi:10.1016/j.trc.2012.09.009
24. Demissie, M.G., de Almeida Correia, G.H., Bento, C.: Exploring cellular network handover information for urban mobility analysis. Journal of Transport Geography **31**, 164–170 (2013). doi:10.1016/j.jtrangeo.2013.06.016
25. Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F.R., Gaughan, A.E., Blondel, V.D., Tatem, A.J.: Dynamic population mapping using mobile phone data. Proceedings of the National Academy of Sciences **111**(45), 15888–15893 (2014). doi:10.1073/pnas.1408439111
26. Louail, T., Lenormand, M., Ros, O.G.C., Picornell, M., Herranz, R., Frias-Martinez, E., Ramasco, J.J., Barthelemy, M.: From mobile phone data to the spatial structure of cities. Scientific Reports **4**(1) (2014). doi:10.1038/srep05276
27. Li, W., Cheng, X., Duan, Z., Yang, D., Guo, G.: A framework for spatial interaction analysis based on large-scale mobile phone data. Computational Intelligence and Neuroscience **2014**, 1–11 (2014). doi:10.1155/2014/363502
28. Iqbal, M.S., Choudhury, C.F., Wang, P., González, M.C.: Development of origin–destination matrices using mobile phone call data. Transportation Research Part C: Emerging Technologies **40**, 63–74 (2014). doi:10.1016/j.trc.2014.01.002
29. Hoteit, S., Secci, S., Sobolevsky, S., Ratti, C., Pujolle, G.: Estimating human trajectories and hotspots through mobile phone data. Computer Networks **64**, 296–307 (2014). doi:10.1016/j.comnet.2014.02.011
30. Calabrese, F., Ferrari, L., Blondel, V.D.: Urban sensing using mobile phone network data: A survey of research. ACM Computing Surveys **47**(2), 1–20 (2014). doi:10.1145/2655691
31. Chi, G., Thill, J.-C., Tong, D., Shi, L., Liu, Y.: Uncovering regional characteristics from mobile phone data: A network science approach. Papers in Regional Science **95**(3), 613–631 (2014). doi:10.1111/pirs.12149
32. Ahas, R., Aasa, A., Yuan, Y., Raubal, M., Smoreda, Z., Liu, Y., Ziemlicki, C., Tiru, M., Zook, M.: Everyday space–time geographies: using mobile phone-based sensor data to monitor urban activity in harbin, paris, and tallinn. International Journal of Geographical Information Science **29**(11), 2017–2039 (2015). doi:10.1080/13658816.2015.1063151
33. Alexander, L., Jiang, S., Murga, M., González, M.C.: Origin-destination trips by purpose and time of day inferred from mobile phone data. Transportation Research Part C: Emerging Technologies **58**, 240–250 (2015). doi:10.1016/j.trc.2015.02.018
34. Blondel, V.D., Decuyper, A., Krings, G.: A survey of results on mobile phone datasets analysis. EPJ Data Science **4**(1) (2015). doi:10.1140/epjds/s13688-015-0046-0
35. Horn, C., Kern, R.: Deriving public transportation timetables with large-scale cell phone data. Procedia Computer Science **52**, 67–74 (2015). doi:10.1016/j.procs.2015.05.026
36. Steenbruggen, J., Tranos, E., Nijkamp, P.: Data from mobile phone operators: A tool for smarter cities? Telecommunications Policy **39**(3-4), 335–346 (2015). doi:10.1016/j.telpol.2014.04.001

37. Doyle, J., Hung, P., Farrell, R., McLeone, S.: Population mobility dynamics estimatedfrom mobile telephony data. Journal of Urban Technology **21**, 109–132 (2014)
38. Picornell, M., Ruiz, T., Lenormand, M., Ramasco, J.J., Dubernet, T., Frías-Martínez, E.: Exploring the potential of phone call data to characterize the relationship between social network and travel behavior. Transportation **42**(4), 647–668 (2015). doi:10.1007/s11116-015-9594-1
39. Xu, Y., Shaw, S.-L., Zhao, Z., Yin, L., Fang, Z., Li, Q.: Understanding aggregate human mobility patterns using passive mobile phone location data: a home-based approach. Transportation **42**(4), 625–646 (2015). doi:10.1007/s11116-015-9597-y
40. Çolak, S., Alexander, L.P., Alvim, B.G., Mehndiratta, S.R., González, M.C.: Analyzing cell phone location data for urban travel. Transportation Research Record: Journal of the Transportation Research Board **2526**, 126–135 (2015). doi:10.3141/2526-14
41. Janecek, A., Valerio, D., Hummel, K.A., Ricciato, F., Hlavacs, H.: The cellular network as a sensor: From mobile phone data to real-time road traffic monitoring. IEEE Transactions on Intelligent Transportation Systems **16**(5), 2551–2572 (2015). doi:10.1109/tits.2015.2413215
42. Trasarti, R., Olteanu-Raimond, A.-M., Nanni, M., Couronné, T., Furletti, B., Giannotti, F., Smoreda, Z., Ziemlicki, C.: Discovering urban and country dynamics from mobile phone data with spatial correlation patterns. Telecommunications Policy **39**(3-4), 347–362 (2015). doi:10.1016/j.telpol.2013.12.002
43. Tranos, E., Nijkamp, P.: Mobile phone usage in complex urban systems: a space–time, aggregated human activity study. Journal of Geographical Systems **17**(2), 157–185 (2015). doi:10.1007/s10109-015-0211-9
44. Douglass, R.W., Meyer, D.A., Ram, M., Rideout, D., Song, D.: High resolution population estimates from telecommunications data. EPJ Data Science **4**(1) (2015). doi:10.1140/epjds/s13688-015-0040-6
45. Dobra, A., Williams, N.E., Eagle, N.: Spatiotemporal detection of unusual human population behavior using mobile phone data. PLOS ONE **10**(3), 0120449 (2015). doi:10.1371/journal.pone.0120449
46. Bonnel, P., Hombourger, E., Olteanu-Raimond, A.-M., Smoreda, Z.: Passive mobile phone dataset to construct origin-destination matrix: Potentials and limitations. Transportation Research Procedia **11**, 381–398 (2015). doi:10.1016/j.trpro.2015.12.032
47. Bajardi, P., Delfino, M., Panisson, A., Petri, G., Tizzoni, M.: Unveiling patterns of international communities in a global city using mobile phone data. EPJ Data Science **4**(1) (2015). doi:10.1140/epjds/s13688-015-0041-5
48. Pappalardo, L., Vanhoof, M., Gabrielli, L., Smoreda, Z., Pedreschi, D., Giannotti, F.: An analytical framework to nowcast well-being using mobile phone data. International Journal of Data Science and Analytics **2**(1-2), 75–92 (2016). doi:10.1007/s41060-016-0013-2
49. Ponieman, N.B., Sarraute, C., Minnoni, M., Travizano, M., Zivic, P.R., Salles, A.: Mobility and sociocultural events in mobile phone data records. AI Communications **29**(1), 77–86 (2016). doi:10.3233/AIC-150687
50. Chua, A., Servillo, L., Marcheggiani, E., Moere, A.V.: Mapping cilento: Using geotagged social media data to characterize tourist flows in southern italy. Tourism Management **57**, 295–310 (2016). doi:10.1016/j.tourman.2016.06.013
51. Raun, J., Ahas, R., Tiru, M.: Measuring tourism destinations using mobile tracking data. Tourism Management **57**, 202–212 (2016). doi:10.1016/j.tourman.2016.06.006
52. Lu, S., Fang, Z., Zhang, X., Shaw, S.-L., Yin, L., Zhao, Z., Yang, X.: Understanding the representativeness of mobile phone location data in characterizing human mobility indicators. ISPRS International Journal of Geo-Information **6**(1), 7 (2017). doi:10.3390/ijgi6010007
53. Panigutti, C., Tizzoni, M., Bajardi, P., Smoreda, Z., Colizza, V.: Assessing the use of mobile phone data to describe recurrent mobility patterns in spatial epidemic models. Royal Society Open Science **4**(5), 160950 (2017). doi:10.1098/rsos.160950
54. Bwambale, A., Choudhury, C.F., Hess, S.: Modelling trip generation using mobile phone data: A latent demographics approach. Journal of Transport Geography (2017). doi:10.1016/j.jtrangeo.2017.08.020
55. Ricciato, F., Widhalm, P., Pantisano, F., Craglia, M.: Beyond the "single-operator, CDR-only" paradigm: An interoperable framework for mobile phone network data analyses and population density estimation. Pervasive and Mobile Computing **35**, 65–82 (2017). doi:10.1016/j.pmcj.2016.04.009
56. Song, X., Ouyang, Y., Du, B., Wang, J., Xiong, Z.: Recovering individual's commute routes based on mobile phone data. Mobile Information Systems **2017**, 1–11 (2017). doi:10.1155/2017/7653706
57. Tolouei, R., Psarras, S., Prince, R.: Origin-destination trip matrix development: Conventional methods versus mobile phone data. Transportation Research Procedia **26**, 39–52 (2017). doi:10.1016/j.trpro.2017.07.007
58. Fiadino, P., Ponce-Lopez, V., Torrero-Gonzalez, J.A., Torrent-Moreno, M., D'Alconzo, A.: Call detail records for human mobility studies. In: Proceedings of the Workshop on Big Data Analytics and Machine Learning for Data Communication Networks - Big-DAMA '17. ACM Press, ??? (2017). doi:10.1145/3098593.3098601
59. Furno, A., Fiore, M., Stanica, R., Ziemlicki, C., Smoreda, Z.: A tale of ten cities: Characterizing signatures of mobile traffic in urban areas. IEEE Transactions on Mobile Computing **16**(10), 2682–2696 (2017). doi:10.1109/tmc.2016.2637901
60. Celik, S.C., Incel, O.D.: Semantic place prediction from crowd-sensed mobile phone data. Journal of Ambient Intelligence and Humanized Computing **9**(6), 2109–2124 (2017). doi:10.1007/s12652-017-0549-6
61. Jiang, S., Ferreira, J., González, M.C.: Activity-based human mobility patterns inferred from mobile phone data: A case study of singapore. IEEE Transactions on Big Data **3**(2), 208–219 (2017). doi:10.1109/tbdata.2016.2631141
62. Masso, A., Silm, S., Ahas, R.: Generational differences in spatial mobility: A study with mobile phone data. Population, Space and Place **25**(2), 2210 (2018). doi:10.1002/psp.2210
63. Anda, C., nez Medina, S.A.O., Fourie, P.: Multi-agent urban transport simulations using OD matrices from mobile phone data. Procedia Computer Science **130**, 803–809 (2018). doi:10.1016/j.procs.2018.04.139
64. Graells-Garrido, E., Caro, D., Parra, D.: Inferring modes of transportation using mobile phone data. EPJ Data Science **7**(1) (2018). doi:10.1140/epjds/s13688-018-0177-1
65. Thuillier, E., Moalic, L., Lamrous, S., Caminada, A.: Clustering weekly patterns of human mobility through mobile phone data. IEEE Transactions on Mobile Computing **17**(4), 817–830 (2018).

doi:10.1109/tmc.2017.2742953
66. Sørensen, A.Ø., Bjelland, J., Bull-Berg, H., Landmark, A.D., Akhtar, M.M., Olsson, N.O.E.: Use of mobile phone data for analysis of number of train travellers. Journal of Rail Transport Planning & Management **8**(2), 123–144 (2018). doi:10.1016/j.jrtpm.2018.06.002
67. Li, Z., Yu, L., Gao, Y., Wu, Y., Song, G., Gong, D.: Identifying temporal and spatial characteristics of residents' trips from cellular signaling data: Case study of beijing. Transportation Research Record: Journal of the Transportation Research Board **2672**(42), 81–90 (2018). doi:10.1177/0361198118793495
68. Liu, Z., Ma, T., Du, Y., Pei, T., Yi, J., Peng, H.: Mapping hourly dynamics of urban population using trajectories reconstructed from mobile phone records. Transactions in GIS **22**(2), 494–513 (2018). doi:10.1111/tgis.12323
69. Wang, Z., He, S.Y., Leung, Y.: Applying mobile phone data to travel behaviour research: A literature review. Travel Behaviour and Society **11**, 141–155 (2018). doi:10.1016/j.tbs.2017.02.005
70. Chen, J., Pei, T., Shaw, S.-L., Lu, F., Li, M., Cheng, S., Liu, X., Zhang, H.: Fine-grained prediction of urban population using mobile phone location data. International Journal of Geographical Information Science **32**(9), 1770–1786 (2018). doi:10.1080/13658816.2018.1460753
71. Batran, M., Mejia, M., Kanasugi, H., Sekimoto, Y., Shibasaki, R.: Inferencing human spatiotemporal mobility in greater maputo via mobile phone big data mining. ISPRS International Journal of Geo-Information **7**(7), 259 (2018). doi:10.3390/ijgi7070259
72. Bachir, D., Khodabandelou, G., Gauthier, V., Yacoubi, M.E., Puchinger, J.: Inferring dynamic origin-destination flows by transport mode using mobile phone data. Transportation Research Part C: Emerging Technologies **101**, 254–275 (2019). doi:10.1016/j.trc.2019.02.013
73. Demissie, M.G., Phithakkitnukoon, S., Kattan, L., Farhan, A.: Understanding human mobility patterns in a developing country using mobile phone data. Data Science Journal **18** (2019). doi:10.5334/dsj-2019-001
74. Oancea, B., Necula, M., Sanguiao, L., Salgado, D., Barragán, S.: A simulator for network event data. Technical report, Statistics Romania (INS) and Statistics Spain (INE) (December 2019). Deliverable I.2 of Work Package I of ESSnet on Big Data II. https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/images/b/b9/WPI_Deliverable_I2_Data_Simulator_-_A_simulator_for_network_event_data.pdf
75. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, ??? (2006)
76. Sanguiao, L., Barragán, S., Salgado, D.: Destim: R Package for Mobile Devices Position Estimation. (2020). R package version 0.1.0. https://github.com/Luis-Sanguiao/destim
77. Oancea, B., Barragán, S., Salgado, D.: Deduplication: R Package for Deduplicating Mobile Device Counts Into Population Individual Counts. (2020). R package version 0.1.0. https://github.com/bogdanoancea/deduplication
78. McLean, D.J., Volponi, M.A.S.: trajr: An r package for characterisation of animal trajectories. Ethology **124**(6), 440–448 (2018). doi:10.1111/eth.12739
79. Oancea, B., Barragán, S., Salgado, D.: Aggregation: An R Package to Produce Probability Distributions of Aggregate Number of Mobile Devices. (2020). R package version 0.1.0. https://github.org/bogdanoancea/aggregation
80. Wolodzko, T.: extraDistr: Additional Univariate and Multivariate Distributions. (2019). R package version 1.8.11. https://CRAN.R-project.org/package=extraDistr
81. Makowski, D., Ben-Shachar, M., Lüdecke, D.: bayestestR: Describing effects and their uncertainty, existence and significance within the bayesian framework. Journal of Open Source Software **4**(40), 1541 (2019). doi:10.21105/joss.01541
82. Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A.: Bayesian Data Analysis. Taylor & Francis Ltd, ??? (2013)

Figures

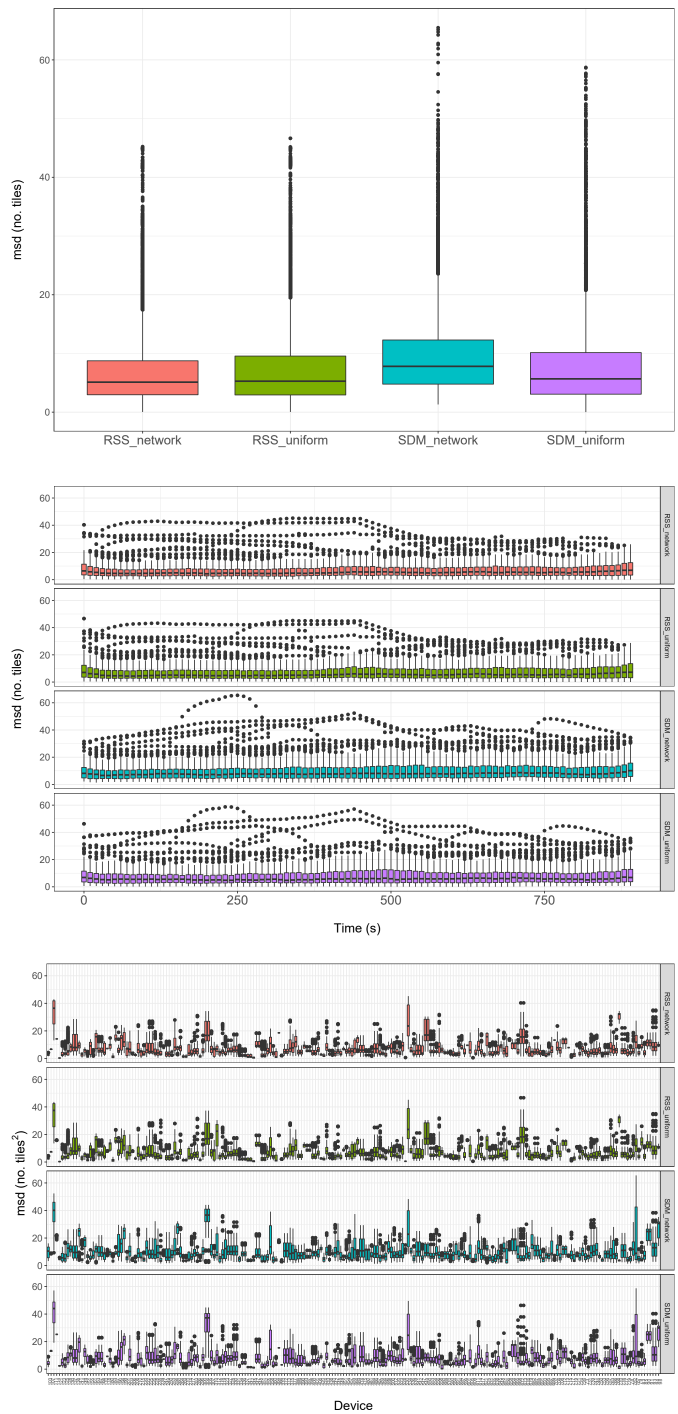

Figure 1 Mean squared distance. (Top) Distribution of mean squared distances msd_{dt} for models RSS and SDM with uniform and network priors. (Middle) Time evolution of distributions of mean squared distances msd_{dt} for the same models. (Bottom) Distribution per device of mean square distance msd_{dt} for all times t for the same models.

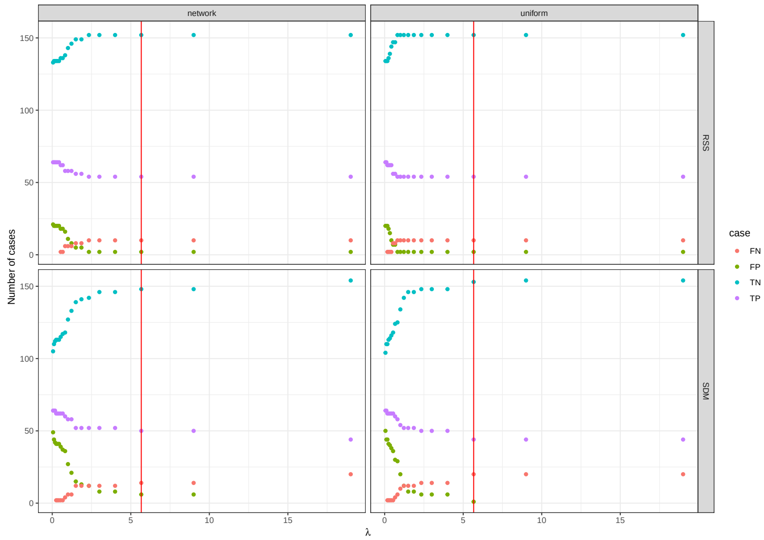

Figure 2 Device duplicity classification for different $\boldsymbol{\lambda}$. Number of true/false positive/negative cases in the classification of device duplicity for different values of the prior ratio $\lambda_d = \lambda$ for all devices d. The red vertical lines represent the value used in the main text.

Tables

Table 1 Simulation parameters. Generic parameters included in `simulation.xml`.

Time (s)	
start_time	0
end_time	900
time_increment	10
time_stay	20
interval_btw_stays	120

MNO	
name	MNO1
device_share	0.35

Others	
displacement	random walk w/ drift
connection_type	strength
connection_threshold	-85dBm
grid_tile_dimensions	250 m×250 m

Table 2 Persons parameters. Parameters included in `persons.xml` (not exhaustive).

Persons	
num_persons	500
speed_walk	3 ms^{-1}
speed_car	16 ms^{-1}

Table 3 Antennas parameters. Parameters per antenna included in `antennas.xml`.

Antenna	
MNO_name	MNO1
max_connections	56
power	10
attenuationfactor	3.8
type	omnidirectional
Smin (thrsh_RSS)	-85 dBm
Qmin (thrsh_SDM)	0.3
S_{mid}	-76 dBm
S_{steep}	0.5
coords	(500 m, 10000 m)

Table 4 Antenna configuration parameters. Marginal distributions of network configuration parameters included in `antenna.xml`.

Parameter	min	q1	q2	mean	q3	max
Power (W)	5.000	10.000	10.000	9.574	10.000	10.000
Path Loss	3.800	3.900	3.900	3.939	4.000	4.000
Radius CoverArea (m)	1121.353	1333.521	1530.999	1483.766	1603.719	1947.483
S_{steep}	0.500	0.900	0.900	0.959	0.900	3.000
S_{mid} (dBm)	-94.000	-80.000	-80.000	-80.871	-79.000	-76.000