



## Research

**Cite this article:** Lima A, Stanojevic R, Papagiannaki D, Rodriguez P, González MC. 2016 Understanding individual routing behaviour. *J. R. Soc. Interface* **13**: 20160021. <http://dx.doi.org/10.1098/rsif.2016.0021>

Received: 11 January 2016

Accepted: 18 February 2016

**Subject Category:**

Life Sciences – Physics interface

**Subject Areas:**

mathematical physics

**Keywords:**

human mobility, complex systems, transportation, city science

**Author for correspondence:**

Antonio Lima

e-mail: [antol@mit.edu](mailto:antol@mit.edu)

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsif.2016.0021> or via <http://rsif.royalsocietypublishing.org>.

# Understanding individual routing behaviour

Antonio Lima<sup>1</sup>, Rade Stanojevic<sup>2</sup>, Dina Papagiannaki<sup>2</sup>, Pablo Rodriguez<sup>2</sup> and Marta C. González<sup>1</sup>

<sup>1</sup>Civil and Environmental Engineering Department, Massachusetts Institute of Technology, Cambridge, MA, USA

<sup>2</sup>Telefonica Research, Barcelona, Spain

AL, 0000-0002-4473-5311

Knowing how individuals move between places is fundamental to advance our understanding of human mobility (González *et al.* 2008 *Nature* **453**, 779–782. (doi:10.1038/nature06958)), improve our urban infrastructure (Prato 2009 *J. Choice Model.* **2**, 65–100. (doi:10.1016/S1755-5345(13)70005-8)) and drive the development of transportation systems. Current route-choice models that are used in transportation planning are based on the widely accepted assumption that people follow the minimum cost path (Wardrop 1952 *Proc. Inst. Civ. Eng.* **1**, 325–362. (doi:10.1680/ipeds.1952.11362)), despite little empirical support. Fine-grained location traces collected by smart devices give us today an unprecedented opportunity to learn how citizens organize their travel plans into a set of routes, and how similar behaviour patterns emerge among distinct individual choices. Here we study 92 419 anonymized GPS trajectories describing the movement of personal cars over an 18-month period. We group user trips by origin–destination and we find that most drivers use a small number of routes for their routine journeys, and tend to have a preferred route for frequent trips. In contrast to the cost minimization assumption, we also find that a significant fraction of drivers' routes are not optimal. We present a spatial probability distribution that bounds the route selection space within an ellipse, having the origin and the destination as focal points, characterized by high eccentricity independent of the scale. While individual routing choices are not captured by path optimization, their spatial bounds are similar, even for trips performed by distinct individuals and at various scales. These basic discoveries can inform realistic route-choice models that are not based on optimization, having an impact on several applications, such as infrastructure planning, routing recommendation systems and new mobility solutions.

## 1. Introduction

The high urban population density [1] poses new critical challenges in designing the cities of the future. Among those, traffic congestion is one of the most pressing issues. Under increasing mobility demand, the intricate task of improving existing infrastructure to allow swift mobility in the city requires special efforts. Technology can be used to collect data about humans interacting with their built environment. Converting unstructured data into knowledge requires specialized methods that extract meaningful information about individual preferences.

In the previous decade, we have learned valuable aspects of human mobility, mainly from large scale data mined from mobile phone networks. Individuals' visit patterns are highly predictable, presenting unique and slow exploration habits [2–7]. Mobile phone traces still remain too coarse, both in space and in time, and are unsuitable to investigate details of human choices. On the other hand, the rapidly increasing popularity of devices equipped with location sensors offers unprecedented possibilities to study individual mobility at a finer-grained level. This new lens enriches our understanding of

human behaviour, and allows us to examine each movement in detail and to better comprehend routing decisions, at the root of vehicular traffic.

Route-choice modelling is the process of estimating the number of vehicles using a link in the road network and it is a fundamental step in transportation forecasting [8]. Given some knowledge of travel demand, models associate individuals with the path they will probably follow during their journey. Urban travel demand has traditionally been estimated by upscaling travel diary surveys [9] and, more recently, through analysis of mobile phone data [10–14]. Route assignment techniques are based on the widely accepted assumption that individuals choose the route that minimizes a cost, usually distance, travel time and/or fuel consumption. The true utilization of a road link is assumed to be similar to that obtained under deterministic user equilibrium, or Wardrop's equilibrium [15]. In stochastic user equilibrium [16,17], a random component is added in the expected travel times, in order to introduce heterogeneity in the routes and to represent travellers' preferences unknown to the modeller. The sets of feasible routes are either obtained by the two methods described above: deterministic shortest path and stochastic shortest path. More recently, probabilistic approaches and constrained enumeration algorithms have also been used for this purpose. In probabilistic approaches [18], a network link is chosen depending on its distance from the shortest path, according to a generalized cost function. Enumeration methods [19,20] rely on the assumption that travellers choose routes according to behavioural rules other than the minimum cost path. However, empirical results have shown that users choose multiple routes over origin–destination pairs, reporting that most choices deviate significantly from the shortest time path [20–23]. Detours can happen for several reasons, like picking up or dropping off a passenger, having a short break at a favourite place or avoiding unpleasant areas (because of high traffic, crime, aesthetic reasons, etc.). Previous literature has tried to identify causes behind deviation from cost minimization, finding that factors that influence this are several and related to many aspects, for example, the initial straightness of the route [24,25], the relative topography [26,27], the presence of landmarks and anchor points [28–30], the direction and other aspects [31] influenced by estimation errors [32].

Investigating the reason behind each detour might be a daunting task, given that the strategies used by drivers are much more diverse than route-choice models assume [33,34]. Instead, we would first instead try to quantify how often these detours occur and how large they are. The hypothesis that we check is that, regardless of the reason behind the detour, a clear physical limit dictates whether a possible deviation that is being taken into consideration by the driver will be ultimately taken or ignored by the driver. In other words, how to quantify these deviations in a set of universal rules in order to be able to synthesize them. Doing that would allow us to inform probabilistic and enumeration route-choice models. Namely, given the daily routine of individuals in different cities with long-term observations, how can we generate heterogeneous yet feasible rules related to alternative route choices?

To that end, we use GPS traces generated by 526 private cars over an 18-month period, and explore how their routing behaviour unfolds in four cities. We investigate how many routes they use and how often they use each of them. We

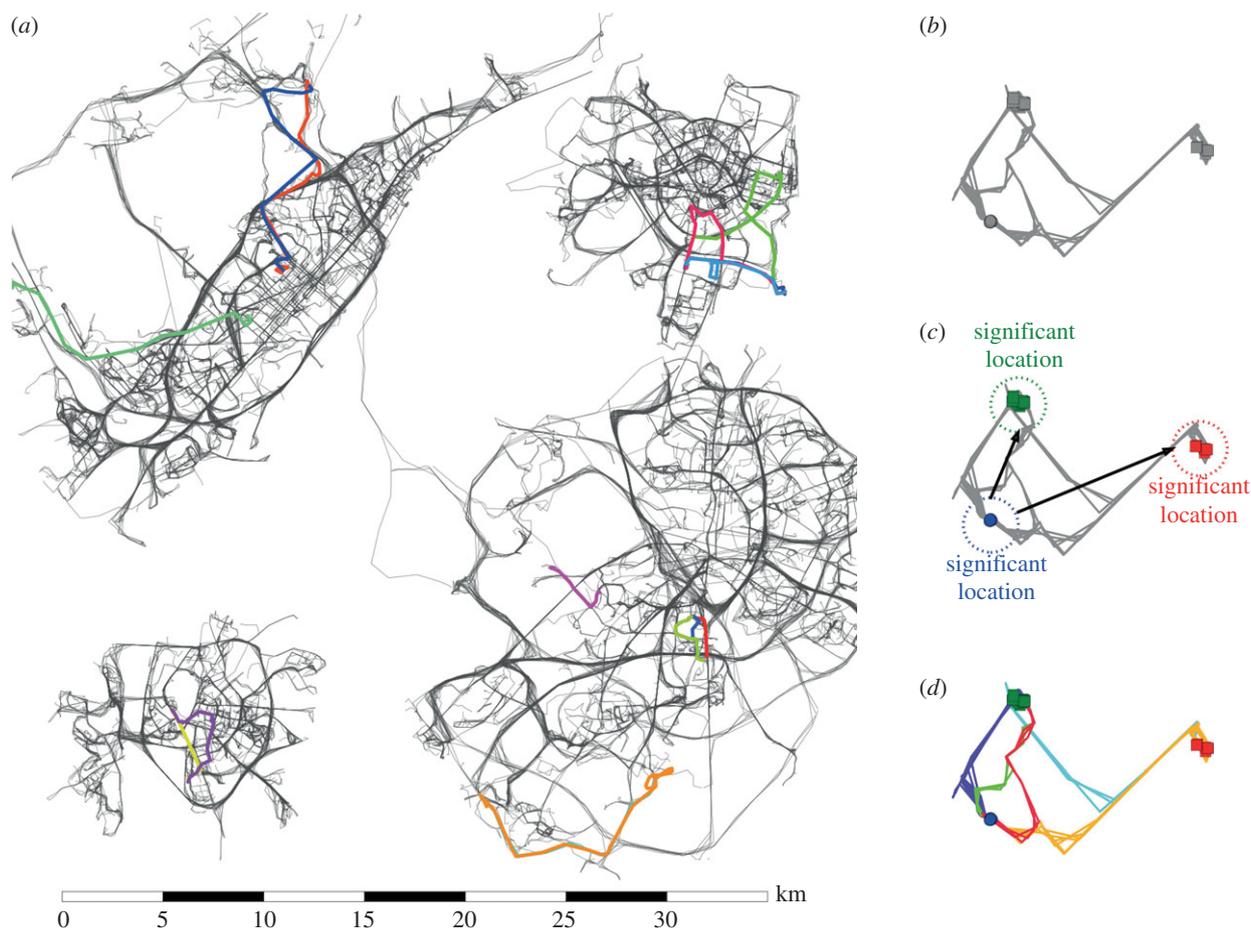
also consider whether these routes are shortest paths and evaluate how far they typically go. We finally give a spatial characterization of routing behaviour. These findings can be used in existing or new models of route choice.

## 2. Results

Firstly, we convert the unstructured sequences of time-referenced positions coming from GPS devices into a meaningful set of locations, trips and route choices [35,36], as shown in figure 1*a*. We describe a *trajectory* as a finite sequence of  $(t, x)$  tuples, where  $t$  represents a time value and  $x$  a location vector. The source and the destination of the trajectory are the first and last point of the sequence, respectively. We call *significant place* a geographical region that a person goes to several times. Significant places detected in this study have a diameter smaller than 600 m, which is compatible with choosing parking spots in the proximity of a destination. Several trips performed by a user between the same pair of significant places together define a *routine trip*. The distribution of significant places in each city is shown in the electronic supplementary material, figure S1. Finally, depending on how spatially similar the trajectories of a routine trip are, they can be grouped into one or more *route choices* (see an illustration in figure 1*b–d* and more details in Material and methods).

The first question we answer is: how many different routes do drivers use in their routine trips? In figure 2*a*, we plot the histogram of the number of routes used for each routine trip. The histograms are surprisingly similar among diverse cities. Independent of the urban settlement under consideration, different individuals prefer to use a limited number of routes, and a third of them use only one route. This is a notable result, considering that these trips span an 18-month period. We can safely conclude that users organize their routine trips through only a few preferred route options, where the number of choices follows a lognormal distribution with parameters  $\mu = 0.71$  and  $\sigma = 2.22$ . The lognormal distribution, linked to a multiplicative random process, is ubiquitous in social science [37] and has been found also in the distribution of single-mode distance trips [38]. In this case, it may arise from the set of unknown random variables that determine individual route choices.

Next, for routine trips that have used more than one route, are some of them chosen more often than others? In order to answer this question, we use a normalized Gini coefficient  $G_n$ , corrected to have meaningful values when the number of routes is small (see Material and methods). A value close to 0 (maximum equality) suggests that routes are evenly used. A value close to 1 (maximum inequality) suggests that the user is strongly biased towards one route for that routine trip and that the alternate routes have been seldom used. In figure 2*d* we plot the normalized Gini, for routine trips that have at least two route choices, computed on the number of times that the route has been used. In general, routine trips have high values of the Gini coefficient with a median value of 0.6, suggesting that people tend to have a dominant route. Moreover, a mild correlation between the Gini values and the number of trips made suggests an adaptation process: when an individual repeats a journey more than 20 times, a preferred route tends to dominate their route choices. By contrast, we found both the number



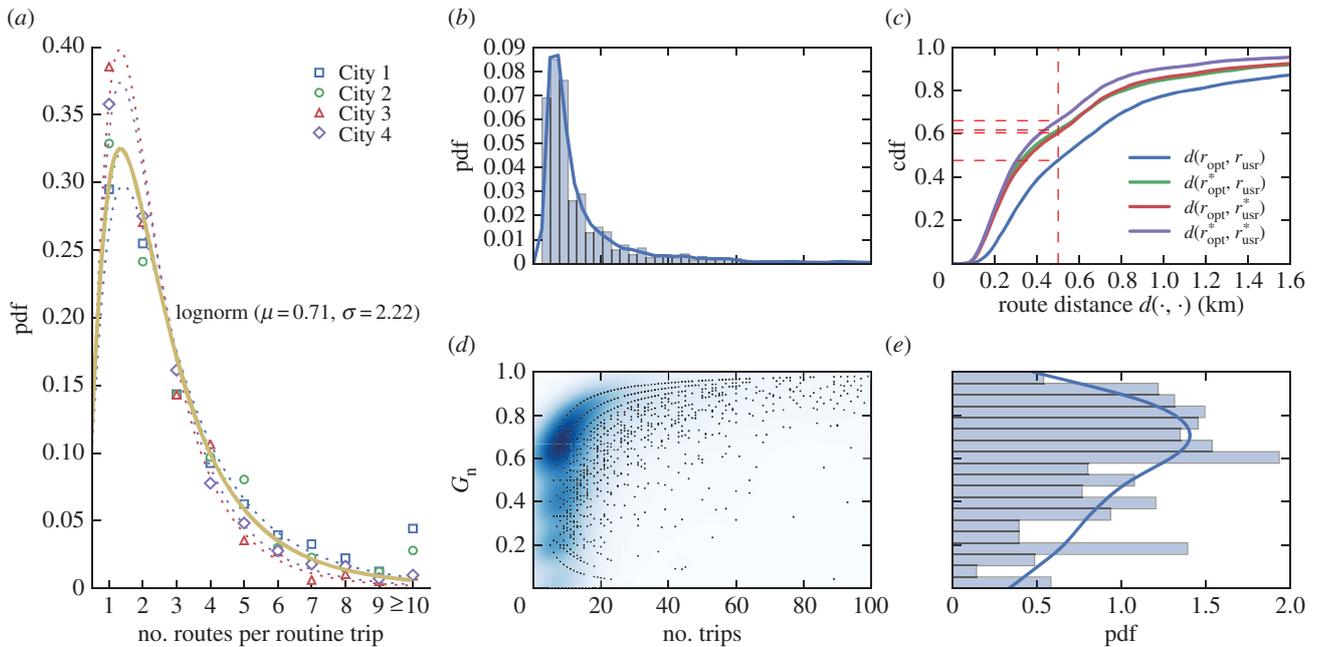
**Figure 1.** From trajectories to route choices. (a) A sample of the trajectories analysed from the four cities, shown in grey, outline their urban road networks. Coloured trajectories spanning between the same pair of points represent seven routine trips. In each routine trip, a coloured line represents a distinct route choice. (b) A set of trajectories belonging to a car. Each trajectory starts at the circle marker and ends at a square marker. (c) By clustering the endpoints of the trips, we find three significant places. Two routine trips are shown with a solid black arrow. (d) We finally discover, for each routine trip, the different route choices performed by the driver. In this example one routine trip has three route choices (purple, green, red), the other has two (cyan, orange). (Online version in colour.)

of routes and  $G_n$  to be uncorrelated with the most common time and day of the week of the routine trip.

Finally, what are the characteristics of a dominant route? Previous research has assumed that drivers prefer routes minimizing some cost function, directly connected to travel time, fuel consumption or distance. We compare the routes taken by the user with the routes suggested by a popular online routing service. The service provides up to three alternative routes, accounting for expected travel times and traffic conditions. In order to compare these recommended optimal routes with the routes actually chosen, we measure the maximum distance between a user's GPS positions and the recommended path (see Material and methods for further information). In figure 2c, we show the distribution of these distances, in four cases: when comparing only the top optimal route with the most used route; when comparing the optimal route with all the user's routes; when comparing the three suggested optimal routes to the dominant route used by the user and, finally, when comparing all suggested routes with all the user's routes. In the last three cases, only the pairs of routes that deviate the least are considered. In about 53% of the cases, the dominant route chosen by the user is not the first optimal choice. For about 34% of the user routines none of the routes are compatible with the optimal choices, indicating that preferred routes do not minimize the travel cost. In electronic supplementary material, figure S3, we see that this result is valid independently

of the distance between origin and destination and that as this distance increases, the chosen route is further away from the optimal route. A previous study at a smaller scale had also found similar results that reject the shortest-path assumption [39]. It is noteworthy that in this paper we define 'optimal routes' as those suggested by a very popular online routing service that takes into account typical traffic conditions based on historical data. We stress this does not mean these routes are optimal in absolute terms. In this case, the suggested routes are optimal according to the service used, based on their traffic estimation model and routing algorithm.

Next, our goal is to determine how far away individuals are willing to go while undertaking their trip. To that end, we study the probability density function  $\Phi(x, y)$  of the route locations, normalized with respect to the source and the destination. We transform trajectories to a common reference frame of coordinates for all trajectories. The goal is to see how paths unfold and how far they usually go from their endpoints, regardless of their geographical position and of the trip length. In figure 3b, we see that most of the deviations are small with respect to the source–destination endpoints. In particular, we find that the majority of the positions recorded are contained within an elliptical area, having as the two foci the first and last point of the trip (figure 3c; electronic supplementary material, figure S5). This result suggests that while individuals commonly take detours due to personal



**Figure 2.** Individual patterns of route choice. (a) The distribution of the number of routes used for a routine trip. For most routine trips this number is low, despite the fact that these trips span over a period of up to 18 months. The markers show the empirical histograms about routine trips grouped by city. The solid curve shows the best lognormal fit, obtained on aggregated data generated in all four cities. (b) The probability density distribution of number of trips performed in a routine trip; the solid line is a kernel density estimation. (b,d,e) Share the axes and are on the same scale. (c) Maximum point distance between the optimal route  $r_{opt}$ , as suggested by the online routing service, and the favourite user route  $r_{usr}$ . For the other three curves, we consider all the alternative routes returned by the service and all the routes ever used by the driver, choosing for each element the route that deviates the least from its counterpart, respectively  $r_{opt}^*$  and  $r_{usr}^*$ . Notably, 34% of the routes chosen are not any of the shortest paths and over 53% of the preferred routes are not optimal. (d) The number of trips performed during a routine journey versus the normalized Gini coefficient related to how many times each route choice is used. The two quantities show a weak correlation (Pearson's  $r = 0.48$ ,  $p = 4.2e-255$ ). The more a driver travels between two locations, the more likely it is for them to have a route of preference. (e) The probability density distribution of the normalized Gini coefficient  $G_n$ ; the solid line shows a kernel density estimation. (Online version in colour.)

preferences or characteristics of the street network [14,40], these detours are well bounded. The emergence of an elliptical shape is not surprising. Keeping in mind that an ellipse is the loci of the points  $P$  such that the sum of the distances to the two focal points  $F_1$  and  $F_2$  is constant ( $d(F_1, P) + d(F_2, P) = a$ ), this result shows that the detour that people are willing to take is bounded. Trips that require larger detours are rare, as they are unlikely to be undertaken, or they might be split into two distinct trips.

In order to further investigate this hypothesis and formally quantify the detours, we calculate two quantities for each trip: the geodesic distance between source and destination  $f$ ; and  $a$ , the maximum value of the sum of the distance to the source and to the destination from any points along the path taken by the user. Finding these values is equivalent to identifying an idealized ellipse that fully contains all the paths taken by the driver. The eccentricity of the ellipse  $e = f/a$  indicates how far from the geodesic this path goes. In the unlikely case where the endpoints lie on the same straight street and the driver takes the shortest route,  $f = a$ , the eccentricity takes the maximal value of 1, and the ellipse degenerates into a straight line. At the other extreme, a value of eccentricity close to 0 indicates that the path taken is very far from the endpoints, the ellipse tends to look like a circle in the target space and the two endpoints are close to each other compared with the path taken by the driver while moving between them.

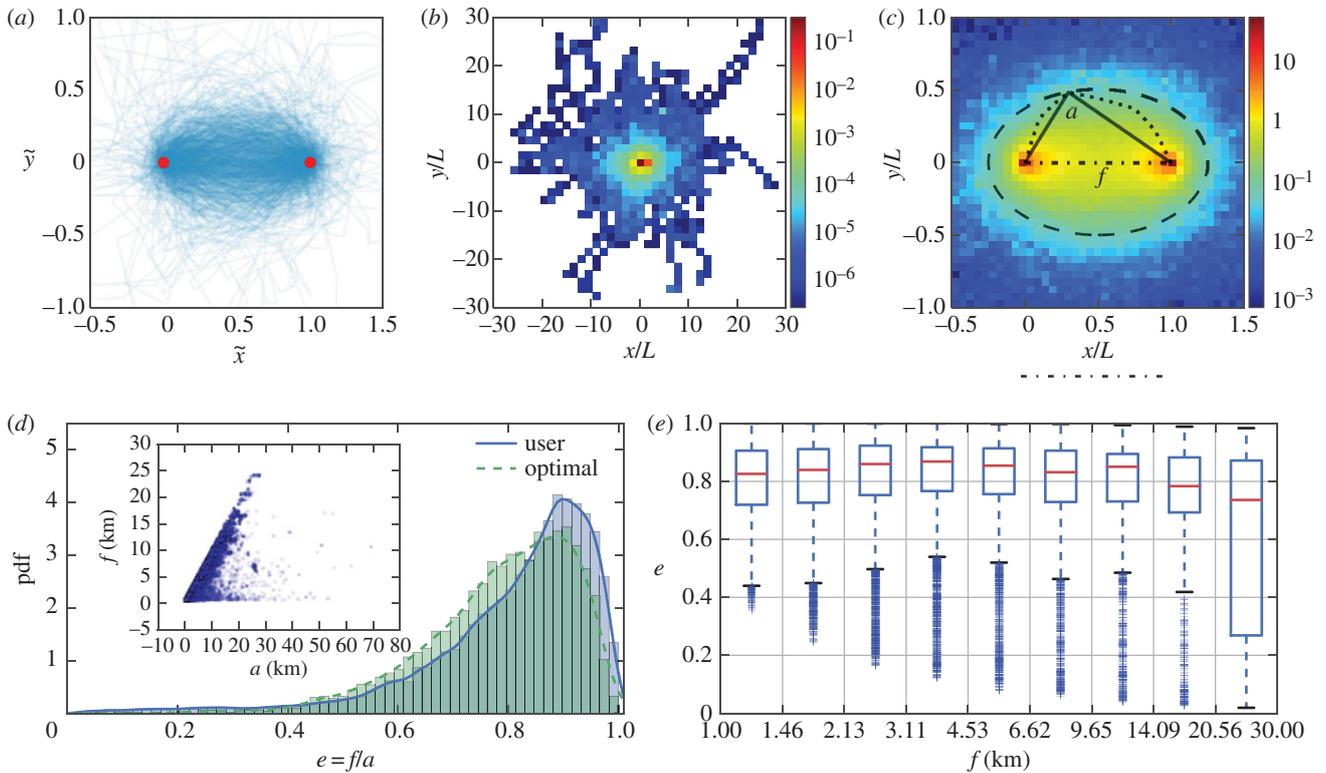
Generally the straight route is not a viable option because of physical obstacles. Drivers deviate from that idealized shortest path according to the underlying street network and personal routing preferences. While these two phenomena are hard to treat, we find that routing detours are well

approximated by an ellipse with high values of eccentricity (figure 3d; electronic supplementary material, figure S4). Large deviations are rare; we speculate that they are caused by intermediate destinations that the driver intends to reach before the final destination (e.g. giving a ride to somebody and dropping them off). Interestingly, the value of the eccentricity does not change considerably with distance between the endpoints (figure 3e), suggesting that, in an urban setting, the space of the routing alternatives is proportional to the effective distance travelled. Whether this result also holds for trips at longer distances, such as inter-city journeys, is to be investigated in future analyses.

It is worth mentioning that ellipses have been previously used to understand the spatial extent of activity spaces and chained trips [41–43]. To the best of our knowledge, instead, this is the first work that uses ellipses to quantify detours of single non-chained trips.

### 3. Discussion

We have discovered a set of behavioural rules that capture individual behaviour in an urban environment. They are independent of the urban layout and were obtained by methods that are agnostic of the underlying street network. The rules establish the basic ingredients of realistic route-choice models. Once a travel plan is established for a user, a dominant route must be assigned. This choice should be spatially bounded within an elliptic shape of high eccentricity, as observed in the experimental distribution, opportunely scaled so that the origin and the destination



**Figure 3.** The boundaries of human routes. Coordinates are projected to a cartesian coordinate system using the spatial reference system EPSG:2062. Each trajectory  $(x, y)$  is roto-translated and scaled into  $(\tilde{x}, \tilde{y})$  so that the source and destination of each trip are  $(0, 0)$  and  $(1, 0)$ , respectively. (a) A 1% random sample of the trips, shown as partially transparent lines connecting consecutive  $(\tilde{x}, \tilde{y})$  positions. (b) The probability density function  $\Phi(\tilde{x}, \tilde{y})$  of the trajectory positions during a journey. Significant detours in all directions are uncommon but not unheard of. (c) Ninety-five per cent of the positions are within the region shown in the figure. The figure shows a sample trajectory, as a dotted line, the ellipse that fully contains it, as a dashed line, the focal distance as a dash-dot line and the major axis as a solid line. (d) The probability density function of the eccentricity of the ellipse containing each trip, shown as a solid blue line, and for comparison, the same quantity measured for the optimal trips, shown as a dashed green line. While both groups of trajectories are characterized by high-eccentricity, optimal trips are slightly less eccentric than actual user trips, suggesting that the former deviate slightly more from the ideal origin-destination straight line. (Online version in colour.)

are the foci of the ellipse. Although the choice can be driven by a distance/cost function from the main axis of the ellipse, it does not have to be deterministically chosen as the path that minimizes a travel cost, as we have seen this does not typically reflect personal routing choices. Finally, individuals could choose alternate routes, within the ellipse, with probability inversely proportional to how often the person travels between the endpoints.

A new science of cities is emerging [44], heavily fuelled by the massive data generated by numerous sensors, inherently interdisciplinary, motivated by the need to improve people's lives and counteract the negative effects of the increasing urban population (such as traffic congestion and pollution, to name the most urgent). The findings generated by this urban science can be successfully used to design simple yet innovative solutions [45,46] that can help cities of today turn into smart cities of tomorrow.

## 4. Material and methods

### 4.1. GPS data

The dataset contains information about the trajectories followed by 526 users in an undisclosed European country over a period of 18 months. The trajectories followed by the cars were collected by GPS devices installed on them. Each trajectory is composed by periodic location updates, taken every 60 s, starting when the driver turns the engine on, until it is turned off. We remove

inconsistent data points that are collected when the number of satellites available is lower than 4, and we remove sudden GPS jumps that are inconsistent with average travel speeds higher than  $110 \text{ km h}^{-1}$ . All user IDs were given in anonymized form.

### 4.2. Significant locations extraction

We extracted each user's significant locations by clustering the starting and ending point of each trajectory. The geographical distance between points was computed using the Haversine formula. The clustering was performed using the mean shift algorithm. This clustering method detects groups of points that are dispersed around a centre, according to a Gaussian distribution. By choosing the bandwidth parameter  $\gamma = 0.025$ , we find clusters of points that are distant from each other at most by 600 m. The sensitivity analysis of  $\gamma$  is shown in the electronic supplementary material, figure S2. These points can be reasonably different parking spots used to reach the same final destination, located at walking distance.

### 4.3. Distance between trajectories

In order to compare trajectories, which in general are defined by a heterogeneous number of points, we use the dynamic time warping (DTW) algorithm, traditionally used in speech recognition and shape analysis. Given two paths  $A = [a_1, a_2, \dots, a_N]$  and  $E = [b_1, b_2, \dots, b_M]$ , specified as sequences of geographical points of different length, we first find an alignment such that the following recursive definition, for  $i = 1 \dots N - 1, j = 1 \dots$

$M-1$ , is minimized:

$$W(A_i, B_j) = d(a_i, b_j) + \min \begin{cases} W(A_{i+1}, B_{j+1}) \\ W(A_{i+1}, B_j) \\ W(A_i, B_{j+1}) \end{cases}, \quad (4.1)$$

where  $A_i$  and  $B_j$  are subsequences containing all the elements  $1 \dots i$  from  $A$  and  $1 \dots j$  from  $B$ , respectively; the element-wise distance  $d$  is here considered to be the Haversine distance. The algorithm tries to match each point in  $A$  with a point in  $B$ , taking into consideration the sequence order. Initially, the two starting points are associated; then the algorithm advances one of the two trajectories, or both, depending on which pair of points minimizes the element-wise distance; the algorithm proceeds until both endpoints are reached. Once the alignment is found between the two trajectories, we consider the maximum distance between all the matched pairs of points.

#### 4.4. Route detection

Clustering of trajectories in a routine trip is performed using the DBSCAN algorithm on the maximum distance in the DTW-aligned trajectories, obtained as previously described. This clustering method has the advantage of not needing to specify the number of groups. However, it is necessary to choose two parameters:  $B$ , the minimum number of trajectories necessary to form a route and  $\epsilon$ , the maximum distance to consider an element part of the cluster. We set  $B = 1$ , so that a single different trajectory is considered as a distinct route choice. We obtained the best clustering results with a choice of  $\epsilon = 0.5$  km; such a value is reasonable, considering that a car travelling at  $30 \text{ km h}^{-1}$  covers that distance during the sample period of 60 s.

It is also worth mentioning that an alternative to the method we devised and used is represented by map-matching and segment-by-segment comparison of routes. However, we decided not to use this option for several reasons. Firstly, map-matching requires full-knowledge of the urban network, making implementation and reproducibility of the results harder. Secondly, as map-matching best performs at higher sampling-rates, in this case, it will introduce additional undesired noise and bias. Finally and most importantly, we are not interested in achieving

maximum precision: while the rate of 60 s rate period might seem high, the detour that a driver is able to make during this time period is quite limited, at most one block away ( $\approx 150$  m) considering an average speed of  $30 \text{ km h}^{-1}$ . We consider any smaller deviation too small to be considered a different route choice.

#### 4.5. Normalized Gini coefficient

The Gini coefficient  $G$  is a statistical index of dispersion of values, typically used in economics to quantify the inequality of income among people. Its value is bounded between  $0 \leq G \leq 1 - 1/N$ , where  $N$  is the size of the population; the coefficient is null for perfect equality and maximum for complete inequality. We use the Gini coefficient to quantify, for a routine trip, how similar the usage frequencies are among all the routes employed at least once by the user. In order to compare this index on routine trips with a heterogeneous number of routes  $N$ , which is typically small, we consider a variant of the Gini coefficient, normalized by the maximum value the Gini index  $G$  obtainable with a number  $N$  of routes:

$$G_n = \frac{G}{1 - 1/N}. \quad (4.2)$$

As a consequence  $G_n = 1$  for perfect inequality, regardless of the number of elements considered.

**Authors' contributions.** A.L. performed the analysis, designed the research and wrote the paper. D.P., P.R. and R.S. provided general advice and helped in the design of the project and data collection. M.C.G. designed and guided the research and wrote the paper.

**Competing interests.** The authors declare that they have no competing interests.

**Funding.** The research was partly funded by the MIT Portugal Program, the DOT New England UTC Year 25 and the MIT-Brazil MISTI seed grant programme. A.L. acknowledges the Vest Scholarship programme. The contents of the paper do not necessarily reflect the position or the policy of funding parties.

**Acknowledgement.** We acknowledge Serdar Çolak and Jameson Toole for important comments and suggestions.

## References

- Cohen JE. 2003 Human population: the next half century. *Science* **302**, 1172–1175. (doi:10.1126/science.1088665)
- González MC, Hidalgo CA, Barabási A-L. 2008 Understanding individual human mobility patterns. *Nature* **453**, 779–782. (doi:10.1038/nature06958)
- Brockmann D, Hufnagel L, Geisel T. 2006 The scaling laws of human travel. *Nature* **439**, 462–465. (doi:10.1038/nature04292)
- Schneider CM, Belik V, Couronné T, Smoreda Z, González MC. 2013 Unravelling daily human mobility motifs. *J. R. Soc. Interface* **10**, 20130246. (doi:10.1098/rsif.2013.0246)
- Song C, Qu Z, Blumm N, Barabási A-L. 2010 Limits of predictability in human mobility. *Science* **327**, 1018–1021. (doi:10.1126/science.1177170)
- de Montjoye Y-A, Hidalgo CA, Verleyen M, Blondel VD. 2013 Unique in the crowd: the privacy bounds of human mobility. *Sci. Rep.* **3**, 1376. (doi:10.1038/srep01376)
- Louail T, Lenormand M, Picornell M, Cantú OG, Herranz R, Frias-Martinez E, Ramasco JJ, Barthelemy M. 2015 Uncovering the spatial structure of mobility networks. *Nat. Commun.* **6**, 6007. (doi:10.1038/ncomms7007)
- Prato CG. 2009 Route choice modeling: past, present and future research directions. *J. Choice Model.* **2**, 65–100. (doi:10.1016/S1755-5345(13)70005-8)
- Stophor PR, Greaves SP. 2007 Household travel surveys: where are we going? *Transp. Res. A, Pol.* **41**, 367–381.
- Wang P, Hunter T, Bayen AM, Schechtner K, González MC. 2012 Understanding road usage patterns in urban areas. *Sci. Rep.* **2**, 1001. (doi:10.1038/srep01001)
- Alexander L, Jiang S, Murga M, González MC. 2015 Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transp. Res. C, Emerg. Technol.* **58**, 240–250. (doi:10.1016/j.trc.2015.02.018)
- Toole JL, Çolak S, Alhasoun F, Evsukoff A, Gonzalez MC. 2014 The path most travelled: mining road usage patterns from massive call data. (<http://arxiv.org/abs/1403.0636>)
- Çolak S, Alexander LP, Alvim BG, Mehndiretta SR, Gonzalez MC. 2015 Analyzing cell phone location data for urban travel: current methods, limitations and opportunities. *Transport. Res. Rec.* **2526**, 126–135. (doi:10.3141/2526-14)
- Barthélemy M. 2011 Spatial networks. *Phys. Rep.* **499**, 1–101. (doi:10.1016/j.physrep.2010.11.002)
- Wardrop JG. 1952 Some theoretical aspects of road traffic research. *Proc. Inst. Civ. Eng.* **1**, 325–362. (doi:10.1680/ipeds.1952.11362)
- Daganzo CF, Sheffi Y. 1977 On stochastic models of traffic assignment. *Transp. Sci.* **11**, 253–274. (doi:10.1287/trsc.11.3.253)
- Cascetta E, Russo F, Vitetta A. 1997 Stochastic user equilibrium assignment with explicit path enumeration: comparison of models and algorithms. In *Transportation systems, 1997, Chania, Greece, 16–18 June* (eds M Papageorgiou, A Pouliezios), pp. 1031–1037. Oxford, UK: Elsevier. See <http://trid.trb.org/view.aspx?id=505812>

18. Frejinger E, Bierlaire M, Ben-Akiva M. 2009 Sampling of alternatives for route choice modeling. *Transp. Res. B, Methodol.* **43**, 984–994. (doi:10.1016/j.trb.2009.03.001)
19. Prato C, Bekhor S. 2006 Applying branch-and-bound technique to route choice set generation. *Transp. Res. Rec.* **1985**, 19–28. (doi:10.3141/1985-03)
20. Levinson D, Zhu S. 2013 A portfolio theory of route choice. *Transp. Res. C, Emerg. Technol.* **35**, 232–243. (doi:10.1016/j.trc.2013.03.001)
21. Jan O, Horowitz AJ, Peng Z-R. 2000 Using global positioning system data to understand variations in path choice. *Transp. Res. Rec.* **1725**, 37–44. (doi:10.3141/1725-06)
22. Selten R, Chmura T, Pitz T, Kube S, Schreckenberg M. 2007 Commuters route choice behaviour. *Games Econ. Behav.* **58**, 394–406. (doi:10.1016/j.geb.2006.03.012)
23. Nielsen OA, Daly A, Frederiksen RD. 2002 A stochastic route choice model for car travellers in the Copenhagen region. *Netw. Spat. Econ.* **2**, 327–346. (doi:10.1023/A:1020895427428)
24. Bailenson JN, Shum MS, Uttal DH. 2000 The initial segment strategy: a heuristic for route selection. *Mem. Cognit.* **28**, 306–318. (doi:10.3758/BF03213808)
25. Hochmair HH, Karlsson V. 2004 Investigation of preference between the least-angle strategy and the initial segment strategy for route selection in unknown environments. In *Spatial cognition IV. Reasoning, action, interaction*, no. 3343 in Lecture notes in computer science (eds C Freksa, M Knauff, B Krieg-Brückner, B Nebel, T Barkowsky), pp. 79–97. Berlin, Germany: Springer.
26. Brunyé TT, Mahoney CR, Gardony AL, Taylor HA. 2010 North is up(hill): route planning heuristics in real-world environments. *Mem. Cognit.* **38**, 700–712. (doi:10.3758/MC.38.6.700)
27. Brunyé TT, Andonova E, Meneghetti C, Noordzij ML, Pazzaglia F, Wienemann R, Mahoney CR, Taylor HA. 2012 Planning routes around the world: international evidence for southern route preferences. *J. Environ. Psychol.* **32**, 297–304. (doi:10.1016/j.jenvp.2012.05.003)
28. Golledge RG, Smith TR, Pellegrino JW, Doherty S, Marshall SP. 1985 A conceptual model and empirical analysis of children's acquisition of spatial knowledge. *J. Environ. Psychol.* **5**, 125–152. (doi:10.1016/S0272-4944(85)80014-1)
29. Manley EJ, Addison JD, Cheng T. 2015 Shortest path or anchor-based route choice: a large-scale empirical analysis of minicab routing in London. *J. Transp. Geography* **43**, 123–139. (doi:10.1016/j.jtrangeo.2015.01.006)
30. Manley EJ, Orr SW, Cheng T. 2015 A heuristic model of bounded route choice in urban areas. *Transp. Res. C Emerg. Technol.* **56**, 195–209. (doi:10.1016/j.trc.2015.03.020)
31. Brunyé TT, Collier ZA, Cantelon J, Holmes A, Wood MD, Linkov I, Taylor HA. 2015 Strategies for selecting routes through real-world environments: relative topography, initial route straightness, and cardinal direction. *PLoS ONE* **10**, e0124404. (doi:10.1371/journal.pone.0124404)
32. Hochmair H, Frank AU. 2000 Influence of estimation errors on wayfinding-decisions in unknown street networks—analyzing the least-angle strategy. *Spat. Cogn. Comput.* **2**, 283–313. (doi:10.1023/A:1015566423907)
33. Chorus CG, Arentze TA, Timmermans HJP. 2008 A random regret-minimization model of travel choice. *Transp. Res. B, Methodol.* **42**, 1–18. (doi:10.1016/j.trb.2007.05.004)
34. Senk P. 2010 Route choice under the microscope. *Transp. Res. Rec.* **2156**, 56–63. (doi:10.3141/2156-07)
35. Giannotti F, Nanni M, Pinelli F, Pedreschi D. 2007 Trajectory pattern mining. In *Proceedings of the 13th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, KDD '07*, pp. 330–339. San Jose, CA: ACM.
36. Zheng Y, Zhou X. 2011 *Computing with spatial trajectories*. Berlin, Germany: Springer Science & Business Media.
37. Limpert E, Stahel WA, Abbt M. 2001 Log-normal distributions across the sciences: keys and clues on the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal: that is the question. *BioScience* **51**, 341–352. (doi:10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2)
38. Kölbl R, Helbing D. 2003 Energy laws in human travel behaviour. *New J. Phys.* **5**, 48. (doi:10.1088/1367-2630/5/1/348)
39. Zhu S, Levinson D. 2015 Do people use the shortest path? An empirical test of Wardrop's first principle. *PLoS ONE* **10**, e0134322. (doi:10.1371/journal.pone.0134322)
40. Porta S, Crucitti P, Latora V. 2006 The network analysis of urban streets: a primal approach. *Environ. Plan. B Plan. Des.* **33**, 705–725. (doi:10.1068/b32045)
41. Newsome TH, Walcott WA, Smith PD. 1998 Urban activity spaces: illustrations and application of a conceptual model for integrating the time and space dimensions. *Transportation* **25**, 357–377. (doi:10.1023/A:1005082827030)
42. Schönfelder S, Axhausen KW. 2003 Activity spaces: measures of social exclusion? *Transp. Policy* **10**, 273–286. (doi:10.1016/j.tranpol.2003.07.002)
43. Buliung RN, Kanaroglou PS. 2006 Urban form and household activity-travel behavior. *Growth Change* **37**, 172–199. (doi:10.1111/j.1468-2257.2006.00314.x)
44. Zheng Y, Capra L, Wolfson O, Yang H. 2014 Urban computing: concepts, methodologies, and applications. *ACM Trans. Intell. Syst. Technol.* **5**, 38:1–38:55.
45. Strano E, Shai S, Dobson S, Barthelemy M. 2015 Multiplex networks in metropolitan areas: generic features and local effects. *J. R. Soc. Interface* **12**, 20150651. (doi:10.1098/rsif.2015.0651)
46. Domenico MD, Lima A, González MC, Arenas A. 2015 Personalized routing for multitudes in smart cities. *EPJ Data Sci.* **4**, 1–11. (doi:10.1140/epjds/s13688-015-0038-0)