

Research



CrossMark  
click for updates

**Cite this article:** Lenormand M, Picornell M, Cantú-Ros OG, Louail T, Herranz R, Barthelemy M, Frías-Martínez E, San Miguel M, Ramasco JJ. 2015 Comparing and modelling land use organization in cities. *R. Soc. open sci.* **2**: 150449.

<http://dx.doi.org/10.1098/rsos.150449>

Received: 1 September 2015

Accepted: 2 November 2015

One contribution to a special feature 'City analytics: mathematical modelling and computational analytics for urban behaviour'.

**Subject Category:**

Physics

**Subject Areas:**

biomathematics/mathematical physics

**Keywords:**

land use, network, human mobility, population distribution, call detail record

**Author for correspondence:**

Maxime Lenormand

e-mail: [maxime@ifisc.uib-csic.es](mailto:maxime@ifisc.uib-csic.es)

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsos.150449> or via <http://rsos.royalsocietypublishing.org>.

# Comparing and modelling land use organization in cities

Maxime Lenormand<sup>1</sup>, Miguel Picornell<sup>2</sup>,  
Oliva G. Cantú-Ros<sup>2</sup>, Thomas Louail<sup>3,4</sup>,  
Ricardo Herranz<sup>2</sup>, Marc Barthelemy<sup>3,5</sup>,  
Enríque Frías-Martínez<sup>6</sup>, Maxi San Miguel<sup>1</sup>  
and José J. Ramasco<sup>1</sup>

<sup>1</sup>Instituto de Física Interdisciplinar y Sistemas Complejos IFISC (CSIC-UIB), Campus Universitat de les Illes Balears, Palma de Mallorca 07122, Spain

<sup>2</sup>Nommon Solutions and Technologies SL, Calle Cañas 8, Madrid 28043, Spain

<sup>3</sup>Institut de Physique Théorique, CEA-CNRS (URA 2306), Gif-sur-Yvette 91191, France

<sup>4</sup>Géographie-Cités, CNRS-Paris 1-Paris 7 (UMR 8504), 13 rue du four, Paris 75006, France

<sup>5</sup>Centre d'Analyse et de Mathématique Sociales, EHESS-CNRS (UMR 8557), 190-198 avenue de France, Paris 75013, France

<sup>6</sup>Telefónica Research, Madrid 28050, Spain

The advent of geolocated information and communication technologies opens the possibility of exploring how people use space in cities, bringing an important new tool for urban scientists and planners, especially for regions where data are scarce or not available. Here we apply a functional network approach to determine land use patterns from mobile phone records. The versatility of the method allows us to run a systematic comparison between Spanish cities of various sizes. The method detects four major land use types that correspond to different temporal patterns. The proportion of these types, their spatial organization and scaling show a strong similarity between all cities that breaks down at a very local scale, where land use mixing is specific to each urban area. Finally, we introduce a model inspired by Schelling's segregation, able to explain and reproduce these results with simple interaction rules between different land uses.

## 1. Introduction

Land use patterns appear as a natural result of citizens' and planners' interaction with the urban space. However, in a feedback loop, they also play a major role in the experience that residents and visitors have of a city [1]. Land use patterns have

an effect on the liveability of neighbourhoods and even on the health of the local residents [2]. On the other hand, land use and transportation display a well-established relation [3–6]. Transport demand depends on the location of residence and business areas, whereas the presence of new transport lines or facilities such as metro stations can substantially modify the land use mixing in a given area of the city. These ideas lie behind the development of the so-called land use transport interaction models [7,8], which are commonly employed in transport planning around the globe [9].

An important issue regarding land use refers to the methods employed to estimate it. City hall registers, surveys or satellite images have been used in the past to this end [10–15].<sup>1</sup> The emergence of geolocated information and communication technologies introduces extra capabilities to directly measure the use that citizens make of each urban space. The information is exhaustive in terms of spatial and temporal resolution, allowing for the detection of concentrations of people second by second along days, weeks and months. Information from mobile phone call records [16–29], geolocated tweets [14,30–34], credit card use [35] or FourSquare [21] has been considered in the literature. Different data sources have been compared, finding a consistent agreement among the estimations on human concentrations and mobility obtained from different information and communication technology data [26], as well as between information and communication technology data and more traditional techniques [20–23,26,28,36].

Such wealth of information together with the ability to process massive data brought by the Internet era allows the systematic comparison of features across cities. This analysis can lead to the discovery and confirmation of properties that have been hypothesized to be common to all cities, and also to laws providing insights into the way a property scales with city size. Some examples of these properties include number of patents filed, unemployment rates, gross domestic product *per capita*, business diversity, consumption of resources, length of road networks or even crime density [37–43]. The finding of these laws raises the hope of the existence of a coherent framework for city science [38,40–42,44–46].

In this work, we explore land use patterns in the five most populous urban areas of Spain. Land use information is obtained from mobile phone records using a new framework based on network theory and systematic comparisons of land use distribution across the five cities are performed at different scales. Our results reveal common features in the land use types' spatial distributions, which can be understood with a model introduced also here. The similarities break down when the land use type mixing is studied at very short spatial scales, exposing patterns characteristic to each city.

## 2. Material and methods

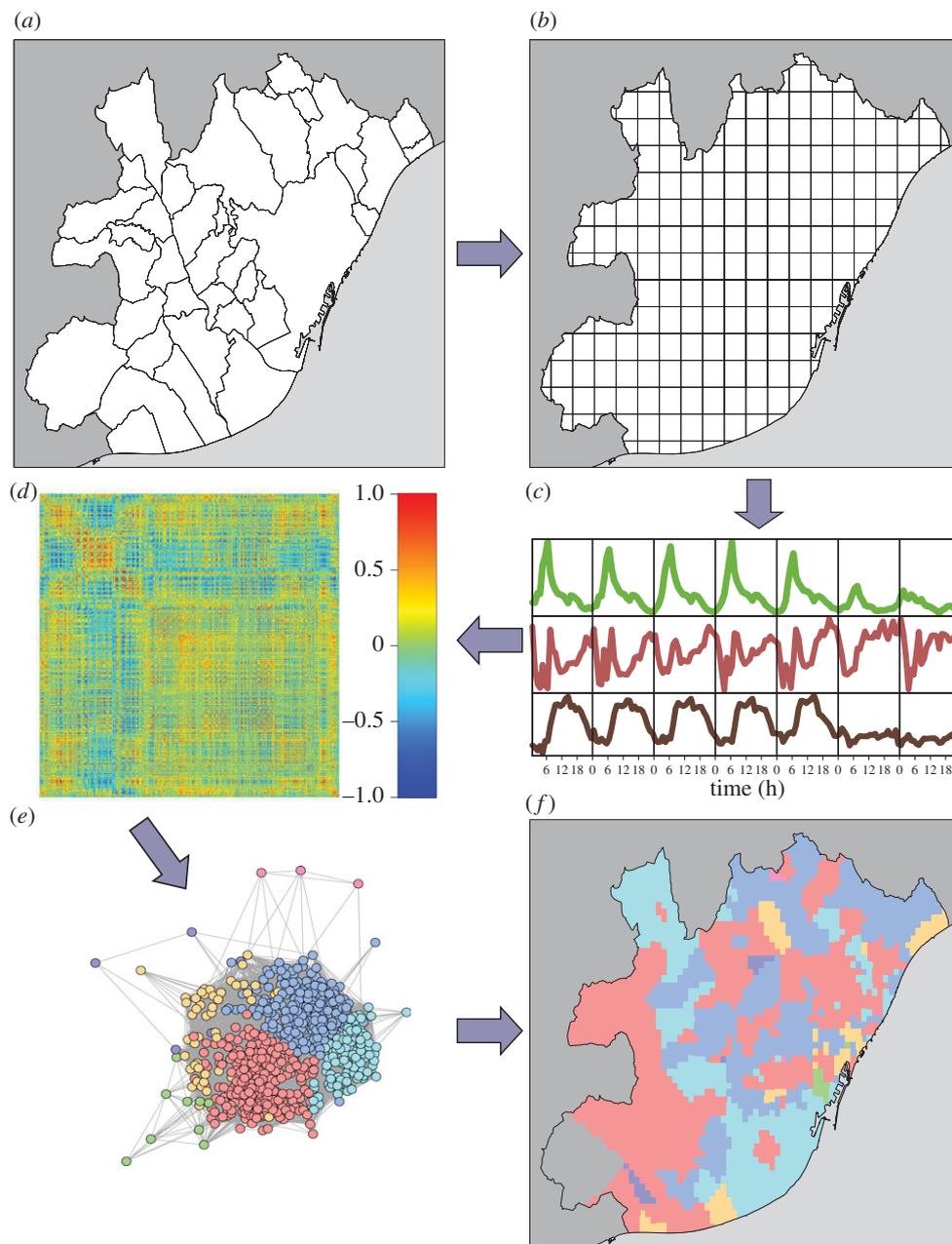
### 2.1. A network approach to detect land use

Our database is composed of aggregated and anonymized call records during 55 days between September and November 2009 in Spain. Every time a user receives or makes a call, the event is registered together with the tower (BTS) providing the service. The positions of the BTSs are georeferenced and so the activity levels of each spatial area can be tracked in time. For this work, we selected the five most populated metropolitan areas of Spain: Madrid (with a population over 5.5 million people), Barcelona (3.2 million), Valencia (1.5 million), Seville (980 000) and Bilbao (900 000). There is no unique definition for the border of an urban area. It may refer, for instance, to official, census or economic delimitation of the cities. Because the focus here is on urban land use, we are interested in identifying the inner zones of each city and, therefore, we use the definition of the metropolitan transportation offices: only areas served by metro or urban buses are considered. This is, nevertheless, an important question, because the selection of borders may influence the scaling analysis when comparing across cities [42,47].

The space of the urban areas is divided following a Voronoi tessellation with the BTS locations as centres. The extension of the areas served by each BTS is very different, because it depends on the expected peaks of demand. To ensure a common geographical framework, the five urban areas are divided into a grid with square cells of  $500 \times 500 \text{ m}^2$  to which the activity is mapped. This should prevent spurious effects owing to the Voronoi area heterogeneity (see electronic supplementary material for a detailed description of the cities and the division process).

The activity (number of users) in each cell is monitored in time and then processed as illustrated in figure 1. Average activity profiles are estimated over each day of the week hour by hour in every cell. These profiles are normalized by the total hourly activity to subtract the trends introduced by the circadian rhythms. A Pearson correlation coefficient is then calculated between the activities of

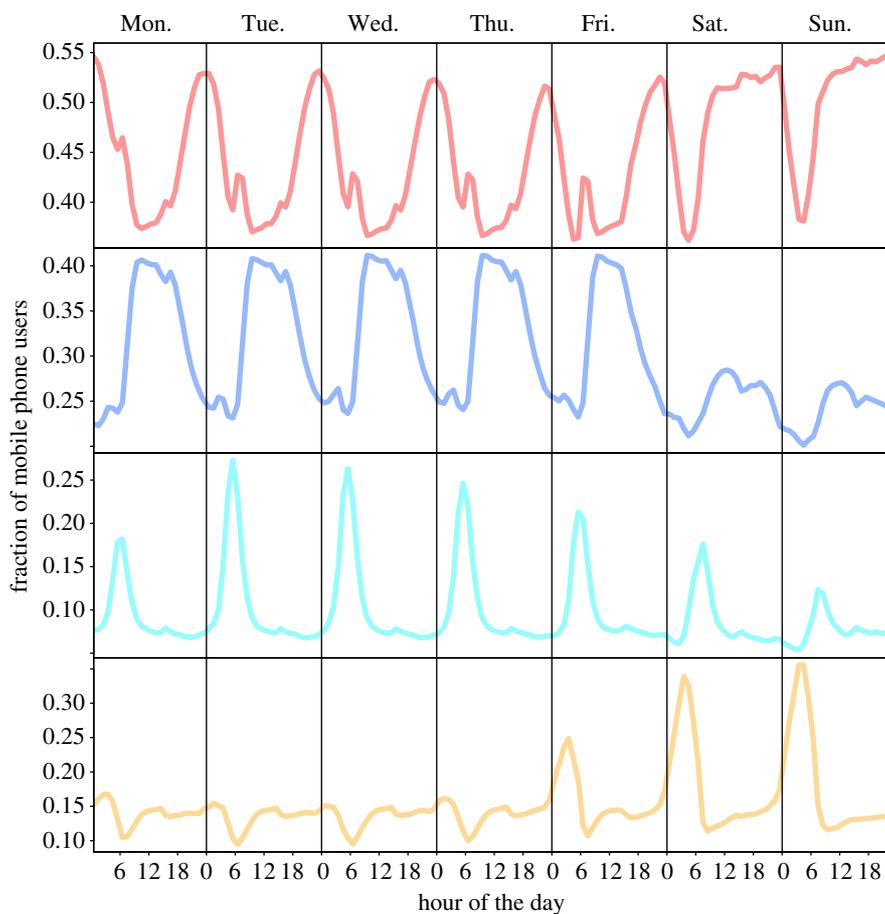
<sup>1</sup>For surveys, see, in Spain, for instance, the Central Office for Cadastral Information (Sede Electrónica de la Dirección General del Catastro (SEC) in Spanish). Accessible at <http://www.sedecatastro.gob.es>.



**Figure 1.** Steps of the method to detect land use. (a,b) The urban area is divided in cells of equal area. (c) For each cell, we calculate an activity profile in terms of phone calls along time during the days of the week. (d) A Pearson correlation matrix between cell activities is computed. Then the matrix formed by correlations over a threshold value  $\delta$  is used to define an undirected weighted network (e), which is clustered using community detection techniques and the results plotted again on the city map (f).

every pair of cells, obtaining a correlation matrix describing the level of similarity between activity profiles. The correlations can take positive and negative values. Distributions of these values are shown in the electronic supplementary material, figure S4. In order to remove non-significant and negative correlations, we consider only Pearson correlation coefficients higher than a threshold  $\delta$ . As a result, we obtain one weighted network per urban area. We first note that variations of the threshold do not produce significant changes in the properties of the resulting network. The results in the main text refer to a value of  $\delta$  equal to the correlation distribution dispersion.

Once the networks are built, their mesoscopic structure is analysed using clustering techniques. The main advantage of community detection algorithms in networks compared with more classical clustering techniques based on dissimilarity matrix is that the number of clusters do not need to be fixed *a priori*. However, it is important to note that different clustering methods can lead to distinct partitions of



**Figure 2.** Temporal patterns associated with the four clusters for the metropolitan area of Madrid. Red: residential cluster; blue: business; cyan: logistics/industry; orange: nightlife.

the networks. We report next results obtained with Infomap [48], whereas a systematic comparison with results obtained with other clustering tools is provided in the electronic supplementary material. As mentioned previously, Infomap does not require the input of a predetermined number of clusters. Therefore, it is interesting to find that in the five cities, between 98% and 100% of the cells are covered with only four groups. Figure 2 shows what the activity looks like for each of these four clusters in Madrid (similar plots for Barcelona, Valencia, Seville and Bilbao are included as electronic supplementary material, figures S9 and S10).

Each of the clusters can be associated with a main land use:

- (1) *Residential (red)*, which is characterized by low activities from 08.00 to 17.00–18.00. For the cells composing this group, the activity peaks around 07.00–08.00 and during the evening. In the weekend, the activity is almost constant except for the night hours.
- (2) *Business (blue)*, where the activity is significantly higher during the weekdays than during the weekends. Furthermore, it concentrates from 09.00 to 18.00–19.00. This land use designation can be related to a wide range of commercial, retail, service and office uses.
- (3) *Logistics/industry (cyan)*, where, as for business, the activity is higher during the weekdays. We observe a large peak between 05.00 and 07.00 followed by a smaller peak around 15.00. This cluster can be related to transport and distribution of goods: for example, ‘Mercamadrid’ (the largest distribution area of Madrid) belongs to this cluster.
- (4) *Nightlife (orange)*, which is characterized by high activity during the night hours (01.00–04.00), especially during the weekends. During the weekdays, these areas show higher activity between 09.00 and 18.00, as for the business cluster, which may be hinting at a certain level of mixing in the land use. Some examples of this category are the ‘Gran Vía’ in Madrid and the ‘Ramblas’ of Barcelona where abound theatres, restaurants and pubs mixed with offices and shops. This is typically the smallest cluster of the four in number of cells.

More systematically, we compared the land use patterns obtained with our algorithm with cadastral data. The dataset contains information about land use for each cadastral parcel of the metropolitan area of Madrid and Barcelona (about 650 000 parcels). In particular, we have for each cadastral parcel the net internal area devoted to residential, business and industrial uses. This information can be used to identify the dominant cadastral land use in each grid cell classified as residential, business and industrial uses by the community detection algorithm. To do so, we need to define a rule to determine what is the dominant land use in a cell. Intuitively, one would tend to identify the dominant land use in a cell as the land use class with the largest area. However, the nature of both land use assignments is very different: the cadastral data are based on the net internal area officially devoted to each activity and not necessarily on the number of people performing it. Therefore, residential use is the land use class with the largest area in most of the cell leading to an over-representation of residential cells in the metropolitan area. To circumvent this limitation, we introduce two thresholds to identify business and logistics cells with cadastral data in order to obtain a distribution of the fraction of cells according to the land use type similar to the one obtained with our algorithm (see electronic supplementary material for details). The overall agreement is high: we find a percentage of correct predictions equal to 65% for Madrid and 60% for Barcelona which is consistent with values obtained in other studies, 54% in [20] and 58% in [22]. Furthermore, for both case studies, almost all land use types have a percentage of correct predictions higher than 50% (see electronic supplementary material for details).

### 3. Results

#### 3.1. Comparison of cities

Once defined the clusters, we can compare the proportion of land use type over the five case studies. Figure 3 shows the fraction of cells and the fraction of mobile phone activity averaged over the time period according to the land use patterns for each metropolitan area. We find similar results for the five case studies: the residential land use type represents on average about 40% of the cells and the mobile phone activity against 30% for the business category and less than 15% for the logistics and nightlife clusters.

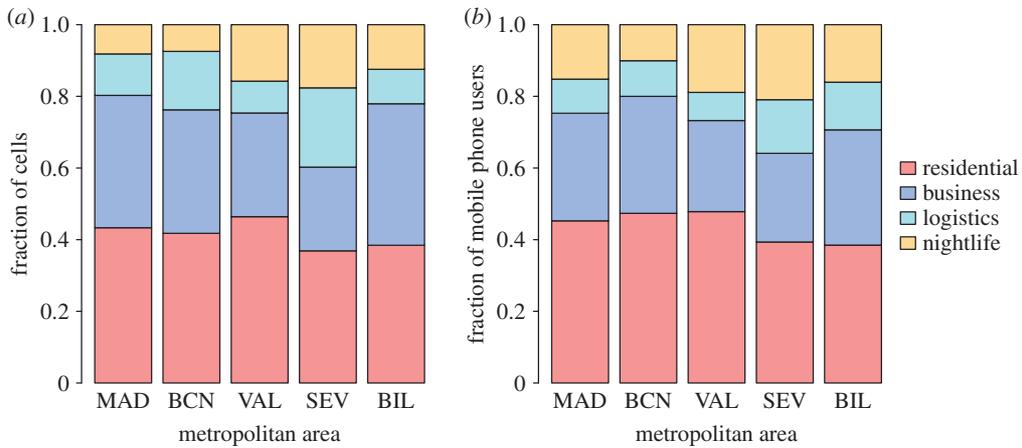
We can also study how the cells in each cluster are organized in the city's space. For the sake of comparison, we arbitrarily consider as city centre the location of the city hall and build a histogram with the number of cells at a certain distance from it. Because each city has a different spatial extension, distances are normalized by dividing by the maximal distance in each city so as to produce comparable results. The distributions are shown in figure 4, where average curves over all cities have been superimposed. It is interesting to note a certain similarity in the distribution of cells for all urban areas. City size acts as a natural cut-off in the distributions, although no simple functional shape is found in any of the clusters. Residential cells are well distributed across the cities but with a maximum not very far from the centre. Business cells appear at a similar distance as residential but peaking a little further. Logistics and industry are preferentially located in the periphery, whereas the nightlife cells are well distributed along the urban areas but slightly more concentrated in the central areas.

In order to quantify land use distribution patterns, we use the Ripley  $K$  [49] defined as

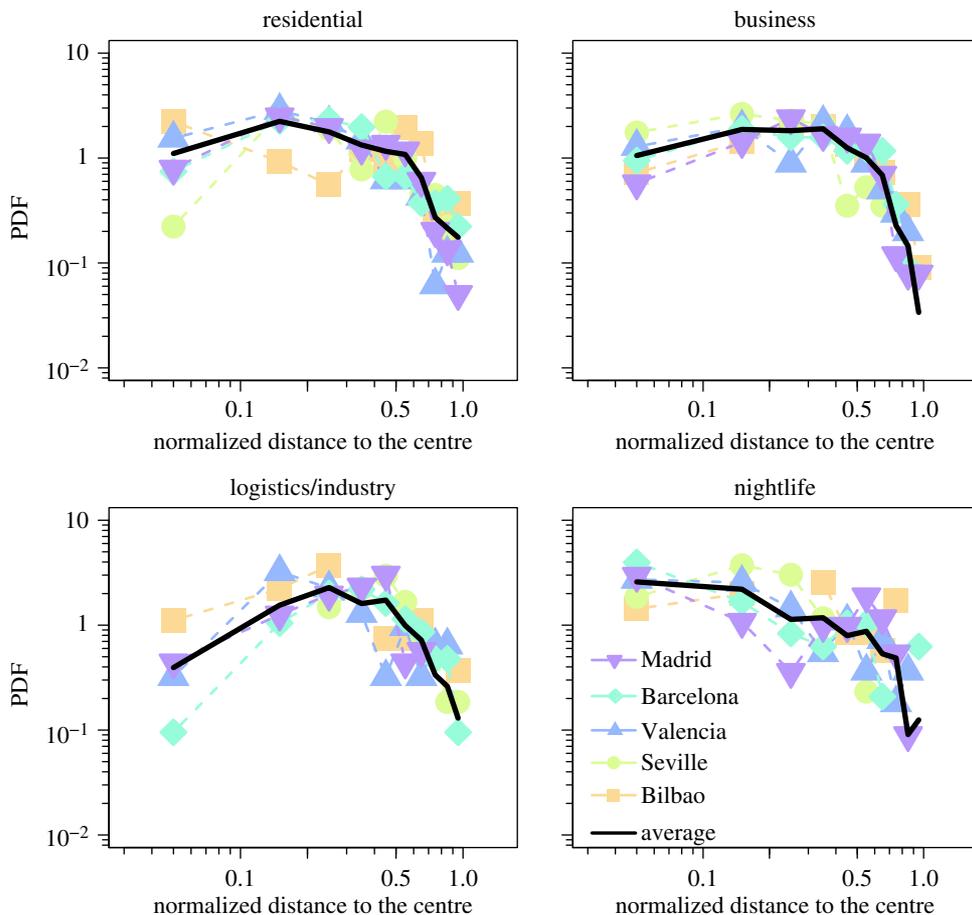
$$K(r) = \frac{A}{n^2} \sum_i^n N_i(r), \quad (3.1)$$

where  $A$  is the city area,  $r$  the search radius (a geographical scale), the index  $i$  runs over the cells in the urban area and  $n$  is the total number of cells.  $N_i(r)$  stands for the number of cells of a given type within a distance  $r$  from the cell  $i$ . This indicator measures the spatial heterogeneity of a given type of cells. The baseline for homogeneous random systems is a growth  $K(r) = \pi r^2$  until reaching  $A$ . If the value of  $K(r)$  is over the random curve for a certain  $r$  it implies that the system is clusterized at that scale. Because cities have different sizes, both  $K(r)$  and the radius must be normalized by their maximum values ( $A$  for  $K(r)$  and the maximum distance for  $r$ ). Curves for the normalized Ripley  $K$  for each city and land use type are displayed in figure 5a as a function of the normalized radius. The  $K(r)/A$  for each city are always above the green curve corresponding to a random distribution of land use types, indicating coarsening of land use. We find a scaling-like curve for all the land use types with most of the cities following well the general trend with some small deviations for nightlife in Seville.

Deepening the analysis, we can also define an entropy index to characterize the land use spatial organization. Let us consider a frame containing the full urban area, which is, in turn, subdivided into a



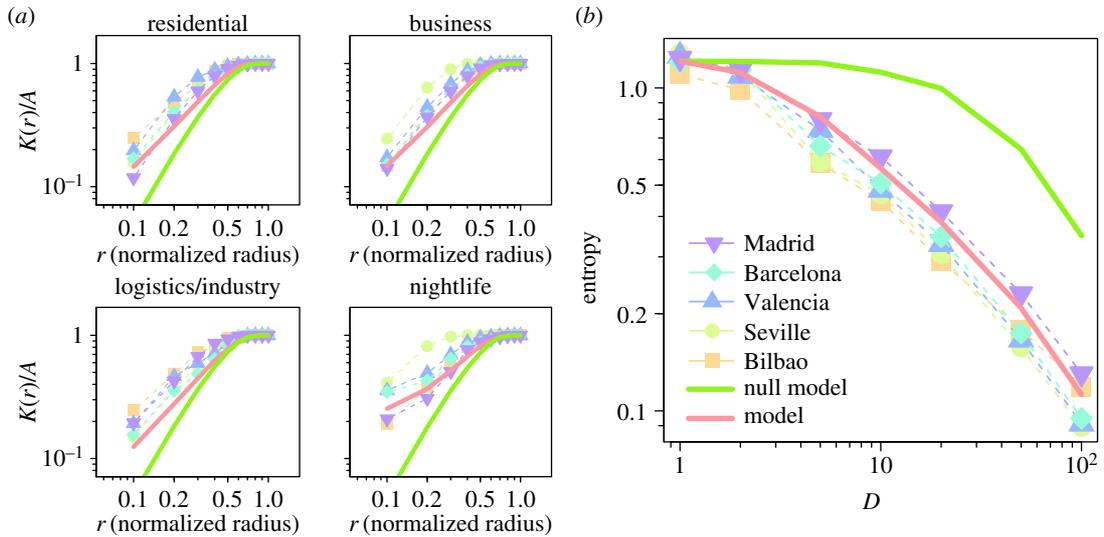
**Figure 3.** Fraction of cells (a) and mobile phone users (b) according to the type of land use for each case study. The fraction of mobile phone users is averaged over the 168 values of the time period.



**Figure 4.** Distribution of the distance between the cells and the city hall according to the type of land use. The distance has been normalized by the maximum distance in each city.

certain number  $D^2$  of equal divisions. Each of these subdivisions,  $B_i$ , intersects the elementary cells so a certain fraction of area falls in each of the land use types:  $f_i^R$  in the residential cluster,  $f_i^B$  in business,  $f_i^L$  in logistics and  $f_i^N$  in nightlife. An entropy index,  $E_i$ , can be defined for  $B_i$  as

$$E_i = - \sum_{\alpha} f_i^{\alpha} \ln(f_i^{\alpha}), \quad (3.2)$$



**Figure 5.** Comparison of the observed and the simulated Ripley's  $K$  and average entropy index. (a) Ripley's  $K$  divided by the city area as a function of the search radius. The radius has been normalized with the maximum value in each urban area. (b) Average entropy index as a function of the lateral number of divisions (inverse scale)  $D$ . The colours and symbols of the curves represent different cities. The red curve corresponds to our model results, and the green curve is the outcome of a random null model. Results for our model were obtained with a calibrated value of  $\gamma = 0.8$ . The red and green curves display the average over 100 realizations.

where  $\alpha$  runs over the four clusters. The entropy  $E_i$  is then averaged over all the divisions to obtain a global metric for the city at a given scale  $E(D)$ .  $E(D)$  tends to zero if the land use within the divisions becomes unique, as occurs for instance at large  $D$  (small spatial scales). On the other extreme, when  $D \rightarrow 1$ ,  $E(D)$  converges to a fixed value describing the full city. Figure 5b shows how the average entropy behaves with  $D$ . The curves are similar across cities, recalling the shape of scaling functions. This is not surprising if the concept of a fractal-like distribution of the city activity applies as has been previously discussed in the literature [37–42].

### 3.2. Modelling land use

Urban land use models in the literature are typically built with relatively elaborate mechanisms [50, 51]. If basic in the rules, the models typically refer to characteristics of cities such as the population or activity distributions [44–46]. The shape of  $E(D)$  can be explained, however, by a simple model inspired by Schelling's segregation [52]. It is important to stress that this model is not intended to reproduce all the processes leading to the land use formation, but to explain the scaling of its spatial distribution patterns.

The basic framework is a lattice in two dimensions representing the urban space. Initially, a variable  $t_i$  with a land use type is assigned to every cell  $i$  at random (residential  $R$ , business  $B$ , logistics  $L$  or nightlife  $N$ ). The global fraction of cells of each type respects the proportions found in the empirical data in such a way that  $E(D = 1)$  coincides with the observations. A *satisfaction* index,  $S_i$ , is then defined per cell taking into account its type and those of its neighbours. Similar to Schelling's model, we assume that the satisfaction increases when a cell is surrounded by cells of its own type. Otherwise,  $S_i$  depends on the particular combinations of types. Some land uses *attract* each other as, for instance, residential and business, whereas others *repel* such as residential and logistics. The existence of rules of attraction and repulsion between different types of land use has already been considered in the past like for example in [53,54]. To be specific, if  $p_t^i$  is the fraction of neighbours of  $i$  of type  $t$ , then  $S_i$  is calculated as

$$\begin{aligned} \text{if } t_i = L, \quad S_i &= \delta_{p_L^i, 1}, \\ \text{if } t_i = N, \quad S_i &= p_N^i \delta_{p_L^i, 0}, \\ \text{if } t_i = R, B, \quad S_i &= \begin{cases} \delta_{p_L^i, 0} & \text{with probability } \gamma, \\ p_{R,B}^i \delta_{p_L^i, 0} & \text{with probability } 1 - \gamma, \end{cases} \end{aligned}$$

where  $\delta_{p,x}$  is the Kronecker delta (equal to one if  $p = x$ , zero otherwise) and  $\gamma$  is the only model parameter. Note that the first condition implies that for logistic cells  $S_i = 1$  only if they are surrounded by cells of

the same type, and that cells of other types have zero satisfaction if surrounded by any logistic one. With this rule, we introduce a tendency to locate industry and logistics out of the core areas of the cities. Residential and business cells have a certain tolerance to the  $R$ ,  $B$  and  $N$  types with  $\gamma$  acting as a mixing control parameter: if  $\gamma = 0$ , mixing is not favoured. A global satisfaction measure is defined as the sum over all the cell satisfaction indices,  $S = \sum_i S_i$ . The model is updated by choosing random pairs of cells and interchanging their land use if the exchange increases  $S$ . This process is repeated until the satisfaction reaches a stationary state.

Calibrating the single parameter  $\gamma$ , we can reproduce the observed  $K(r)/A$  and  $E(D)$  scaling in the real urban areas (see red curves in figure 5). The value of the mixing parameter at which the best average results are obtained is  $\gamma = 0.08$  (electronic supplementary material, figure S8). For comparison, we have included a null model in which the land use types are distributed at random, keeping the real proportions (green curves in figure 5). Unsurprisingly, the null model fails at reproducing the curves obtained with the data, mainly because, generally, areas of a given land use type tend to cluster together to form land use zones. More interestingly, assuming that the satisfaction of a cell increases when it is surrounded by cells of its own type and that logistic cells and cells repel each other (i.e.  $\gamma = 0$ ) is also insufficient to reproduce the properties observed in the data (for more details, see electronic supplementary material, figure S8). Therefore, taking into account the attraction between residential and business areas seems to be crucial for the reproduction of land use spatial organization in cities.

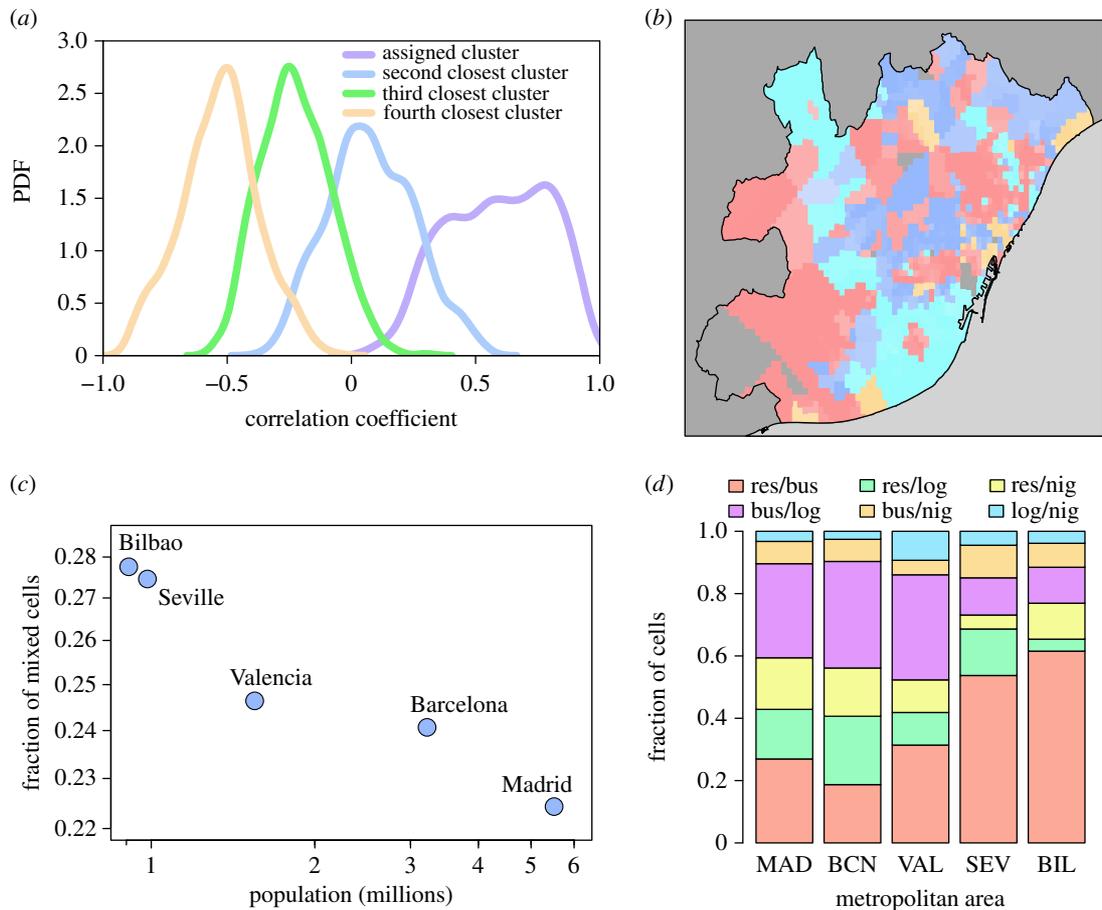
### 3.3. Mixing of land use types

So far, we have considered that each elementary cell has a unique land use type associated. This condition can be easily relaxed. If an average activity profile is defined for each of the four clusters, then a Pearson correlation coefficient between the activity profile in each cell and the clusters' averages can be calculated. The distribution of correlation values is shown in figure 6*a*. The highest correlation value corresponds typically to the cluster at which the cell is assigned. Still, in some cases, positive correlation values are found for other or even two other clusters. For every cell, we can quantify the intensity of its relation with each cluster by summing over these positive correlations and normalizing by the total. A map of the Barcelona metropolitan area with the intensity of each cell relation with its assigned cluster is shown in figure 6*b*. The colours represent the four main type of cells and the colour saturation is related to the correlation: darker if the correlation is high, paler otherwise. Most cells match well with their original assigned cluster, keeping darker colours, whereas some are brighter, implying a higher level of land use mixing.

We arbitrarily define a cell as mixed when the normalized correlations fall within the interval 0.3–0.7 for other clusters besides the assigned one. The fraction of mixed cells as a function of the city population is displayed in figure 6*c*. Larger cities show lower mixing and contain areas devoted to more specific purposes. Figure 6*d* illustrates how the land use types combine in each cell. Business and residential integrate well together as in our model, increasingly so for smaller cities. The mixing proportions are city-dependent and act as a fingerprint to characterize each urban area. This feature may be used to classify cities with similar land use mixing patterns. Besides population size, causal links between level of spatial mixing and city shape, area, age of the city, function, etc., will be explored in future investigations. The mixing proportion can be either related to the organization of cities as monocentric or polycentric [27,55]. Smaller cities display a more monocentric structure, which can be associated with the mixing of land use types given the most restricted space.

## 4. Discussion

In summary, we introduce a method to automatically detect land use from electronic records and apply it to the five largest urban areas of Spain in order to perform a systematic comparison across them on the land use distribution. The urban space is divided into a regular grid to prevent geographic heterogeneity and to maintain the spatial scale under control. The user activity profiles are monitored in each unit cell along time, and then a correlation matrix is established between the profiles of every pair of cells. This correlation matrix encodes the functional network of each city. We analyse them by using network clustering techniques, which ensures that cells showing similar user profiles are grouped together. This method has been applied to the five most populated Spanish cities: Madrid, Barcelona, Valencia, Seville and Bilbao. Because the delimitation of urban areas could affect the results, the definition of the municipal transport offices is employed in each case. Interestingly given that the method is unsupervised, four



**Figure 6.** Land use mixing. (a) Distribution of the Pearson correlations between cell activity and the average cluster profiles. (b) Map of Barcelona displays the four clusters with the colours varying from white to the baseline according to the intensity of the relation with the assigned cluster of each cell. The colour code is red for residential, blue for business, cyan for logistics/industry and orange for nightlife. (c) Fraction of mixed cells as a function of the city population. (d) Fraction of cells classified by the type of land use mixing among those with two types of land use.

groups consistently appear as dominant in all cities. They correspond to activity profiles compatible with main land uses in residence, business, logistics/industry and nightlife. Not only are the types the same across cities, but also the proportions of cells and area devoted to each type are similar.

We also study the distribution of the four land use types at different spatial scales. We define the Ripley  $K$  and the entropy index for each land use type and the behaviour of both metrics is explored as the spatial scale varies from the full city (macroscopic scale) to a single cell (microscopic). The five cities show similar scaling curves for the metrics, implying comparable structures regarding how the four types amalgamate at the urban level. The shape of the scaling curves can be explained by a simple model that has been proposed in this work. The model is based on a Schelling-like segregation in which the different land use types interact to generate a spatial distribution in the city. Cells in a given land use type tend to maximize the number of neighbours undergoing equivalent uses. This rule induces a tendency to coarsening in land use types. The different land uses interact by attracting each other, such as services and residential areas, or by repelling like industry and almost any other type. The calibration of a single parameter regulating the intensity of the attraction between services, residential uses and nightlife is enough to reproduce the scaling curves observed in the real cities. Moreover, we also demonstrate that a model without land use type interactions cannot recreate the empirical scaling.

Similarities across cities break down when one focuses on how the land use types mix microscopically within each unit cell. A characteristic mixing profile is detected for every urban area, providing an individual city fingerprint. Further data on other cities could help to elucidate whether different typologies exist at this microscopic mixing level. In conclusion, despite that further data from other countries and sources could be important to confirm our results, we find that a coherent picture emerges in the land use organization of major urban areas and that its origin can be explained with a basic model.

Data accessibility. The mobile phone dataset used in this study is proprietary and subject to strict privacy regulations. The access to this dataset was granted after reaching a non-disclosure agreement with the proprietor, who anonymized and aggregated the original data before giving access to other authors.

Authors' contributions. M.L. designed the study, analysed the data and wrote the manuscript. O.G.C. and M.P. processed the data. T.L., R.H., M.B. and M.S.M. coordinated the study. E.F.M. obtained and processed the data. J.J.R. designed the study and wrote the manuscript. All authors read, commented on and approved the final version of the manuscript. Competing interests. The authors declare no competing interests.

Funding. Partial financial support was received from the Spanish Ministry of Economy (MINECO) and FEDER (EU) under project INTENSE@COSYP (FIS2012-30634), and from the EU Commission through projects LASAGNE and INSIGHT. The work of M.L. was funded under the PD/004/2013 project, from the Conselleria de Educació, Cultura y Universidades of the Government of the Balearic Islands and from the European Social Fund through the Balearic Islands ESF operational programme for 2013–2017. J.J.R. acknowledges funding from the Ramón y Cajal program of MINECO.

## References

- Humphries C. 2012 The science of cities: life in the concrete jungle. *Nature* **491**, 514–515. (doi:10.1038/491514a)
- Sallis JF, Saelens BE, Frank LD, Conway TL, Slymen DJ, Cain KL, Chapman JE, Kerr J. 2009 Neighborhood built environment and income: examining multiple health outcomes. *Soc. Sci. Med.* **68**, 1285–1293. (doi:10.1016/j.socscimed.2009.01.017)
- Badoo D, Miller E. 2000 Transportation—land-use interactions: empirical findings and implications for modeling. *Transp. Res. D: Transp. Environ.* **5**, 235–263. (doi:10.1016/S1361-9209(99)00036-X)
- Cervero R. 2002 Built environments and mode choice: toward a normative framework. *Transp. Res. D: Transp. Environ.* **7**, 265–284. (doi:10.1016/S1361-9209(01)00024-4)
- Krygsman S, Dijst M, Arentze T. 2004 Multimodal public transport: an analysis of travel time elements and the interconnectivity ratio. *Transp. Policy* **11**, 265–275. (doi:10.1016/j.tranpol.2003.12.001)
- Tsai Y-H. 2005 Quantifying urban form: compactness versus “sprawl”. *Urban Stud.* **42**, 141–161. (doi:10.1080/0042098042000309748)
- Cervero R, Kockelman K. 1997 Travel demand and the 3Ds: density, diversity, and design. *Transp. Res. D: Transp. Environ.* **2**, 199–219. (doi:10.1016/S1361-9209(97)00009-6)
- Waddell P, Ulfarsson GF, Franklin JP, Lobb J. 2007 Incorporating land use in metropolitan transportation planning. *Transp. Res. A: Policy Pract.* **41**, 382–410. (doi:10.1016/j.tra.2006.09.008)
- Bartholomew K. 2007 Land use–transportation scenario planning: promise and reality. *Transportation* **34**, 397–412. (doi:10.1007/s11116-006-9108-2)
- Donnay JP, Unwin D. 2001 Modeling geographical distribution in urban areas. In *Remote sensing and urban analysis* (eds JP Donnay, MJ Barnsley, PA Longley), pp. 205–224. New York, NY: Taylor & Francis.
- Yang X, Lo CP. 2002 Sing a time series of satellite imagery to detect land use and land cover changes in the Atlanta, Georgia metropolitan area. *Int. J. Remote Sens.* **23**, 1775–1798. (doi:10.1080/01431160110075802)
- Geurs KT, van Wee B. 2004 Accessibility evaluation of land-use and transport strategies: review and research directions. *J. Transp. Geogr.* **12**, 127–140. (doi:10.1016/j.jtrangeo.2003.10.005)
- Wu S, Xu B, Wang L. 2006 Urban land use classification using variogram-based analysis with aerial photographs. *Photogramm. Eng. Remote Sens.* **72**, 812–822. (doi:10.14358/PERS.72.7.813)
- Frías-Martínez V, Soto V, Hohwald H, Frías-Martínez E. 2012 Characterizing urban landscapes using geolocated tweets. In *SOCIALCOM-PASSAT '12 Proc. 2012 ASE/IEEE Int. Conf. on Social Computing and 2012 ASE/IEEE Int. Conf. on Privacy, Security, Risk and Trust, Amsterdam, The Netherlands, 3–5 September 2012*, pp. 239–248. Piscataway, NJ: IEEE Computer Society. (doi:10.1109/SocialCom-PASSAT.2012.19)
- Hu S, Wang L. 2013 Automated urban land-use classification with remote sensing. *Int. J. Remote Sens.* **34**, 790–803. (doi:10.1080/01431161.2012.714510)
- Reades J, Calabrese F, Sevtsuk A, Ratti C. 2007 Cellular census: explorations in urban data collection. *IEEE Pervasive Comput.* **6**, 30–38. (doi:10.1109/MPRV.2007.53)
- Gonzalez MC, Hidalgo CA, Barabasi A-L. 2008 Understanding individual human mobility patterns. *Nature* **453**, 779–782. (doi:10.1038/nature06958)
- Reades J, Calabrese F, Ratti C. 2009 Eigenplaces: analysing cities using the space-time structure of the mobile phone network. *Environ. Plan. B, Plan. Des.* **36**, 824–836. (doi:10.1068/b34133t)
- Soto V, Frías-Martínez E. 2011 Automated land use identification using cell-phone records. In *HotPlanet11 Proc. 3rd ACM Int. Workshop on MobiArch*, pp. 17–22. New York, NY: ACM. (doi:10.1145/2000172.2000179)
- Toole JL, Ulm M, González MC, Bauer D. 2012 Inferring land use from mobile phone activity. In *Proc. ACM SIGKDD Int. Workshop on Urban Computing, Beijing, China, 12–16 August 2012*, pp. 1–8. New York, NY: ACM. (doi:10.1145/2346496.2346498)
- Noulas A, Mascol C, Frías-Martínez E. 2013 Exploiting foursquare and cellular data to infer user activity in urban environments. In *Proc. IEEE 14th Int. Conf. on Mobile Data Management, Milan, Italy, 3–6 June 2013*, pp. 167–176. Piscataway, NJ: IEEE Computer Society. (doi:10.1109/MDM.2013.27)
- Pei T, Sobolevsky S, Ratti C, Shaw SL, Zhou C. 2014 A new insight into land use classification based on aggregated mobile phone data. *Int. J. Geogr. Inf. Sci.* **28**, 1988–2007. (doi:10.1080/13658816.2014.913794)
- Grauwín S, Sobolevsky S, Moritz S, Godor I, Ratti C. 2015 Towards a comparative science of cities: using mobile traffic records in New York, London and Hong Kong. In *Computational approaches for urban environments* (eds M Helbich, J Jokar Arsanjani, M Leitner). Geotechnologies and the Environment, vol. 13. Cham, Switzerland: Springer International. (doi:10.1007/978-3-319-11469-9\_15)
- Frías-Martínez V, Soto V, Sanchez A, Frías-Martínez E. 2014 Consensus clustering for urban land use analysis using cell phone network data. *Int. J. Ad Hoc Ubiquitous Comput.* **17**, 39–58. (doi:10.1504/IJAHUC.2014.065157)
- Frías-Martínez V, Frías-Martínez E. 2014 Spectral clustering for sensing urban land use using twitter activity. *Eng. Appl. Artif. Intell.* **35**, 237–245. (doi:10.1016/j.engappai.2014.06.019)
- Lenormand M, Picornell M, Tugores A, Louail T, Herranz R, Barthelemy M, Frías-Martínez E, Ramasco JJ. 2014 Cross-checking different source of mobility information. *PLoS ONE* **9**, e105184. (doi:10.1371/journal.pone.0105184)
- Louail T, Lenormand M, Picornell M, Herranz R, Frías-Martínez E, Ramasco JJ, Barthelemy M. 2014 From mobile phone data to the spatial structure of cities. *Sci. Rep.* **4**, 5276. (doi:10.1038/srep05276)
- Deville P, Linard C, Martin S, Gilbert M, Stevens FR, Gaughan AE, Blondel VD, Tatem AJ. 2014 Dynamics population mapping using mobile phone data. *Proc. Natl Acad. Sci. USA* **111**, 15 888–15 893. (doi:10.1073/pnas.1408439111)
- Louail T, Lenormand M, Picornell M, Herranz R, Frías-Martínez E, Ramasco JJ, Barthelemy M. 2015 Uncovering the spatial structure of mobility networks. *Nat. Commun.* **6**, 6007. (doi:10.1038/ncomms7007)
- Gonçalves B, Sánchez D. 2014 Crowdsourcing dialect characterization through Twitter. *PLoS ONE* **9**, e112074. (doi:10.1371/journal.pone.0112074)
- Llorente A, García-Herranz M, Cebrían M, Moro E. 2014 Social media fingerprints of unemployment. (<http://arxiv.org/abs/1411.3140>)
- Hawelka B, Sitko I, Beinert E, Sobolevsky S, Kazakopoulos P, Ratti C. 2014 Geo-located Twitter as a proxy for global mobility patterns. *Cartogr. Geogr. Inf. Sci.* **41**, 260–271. (doi:10.1080/15230406.2014.890072)
- Lenormand M, Tugores A, Colet P, Ramasco JJ. 2014 Tweets on the road. *PLoS ONE* **9**, e105407. (doi:10.1371/journal.pone.0105407)

34. Lenormand M, Gonçalves B, Tugores A, Ramasco JJ. 2015 Human diffusion and city influence. (<http://arxiv.org/abs/1501.07788>)
35. Lenormand M, Louail T, Cant-Ros OG, Picornell M, Herranz R, Murillo Arias J, Barthelemy M, San Miguel M, Ramasco JJ. 2015 Influence of sociodemographic characteristics on human mobility. *Sci. Rep.* **5**, 10075. (doi:10.1038/srep10075)
36. Tizzoni M, Bajardi P, Decuyper A, Kon Kam King G, Schneider CM, Blondel V, Smoreda Z, Gonzalez MC, Colizza V. 2014 On the use of human mobility proxy for the modeling of epidemics. *PLoS Comput. Biol.* **10**, e1003716. (doi:10.1371/journal.pcbi.1003716)
37. Bettencourt L, Lobo J, Helbing D, Kuhnert C, West G. 2007 Growth, innovation, scaling, and the pace of life in cities. *Proc. Natl Acad. Sci. USA* **104**, 7301–7306. (doi:10.1073/pnas.0610172104)
38. Batty M. 2008 The size, scale, and shape of cities. *Science* **319**, 769–771. (doi:10.1126/science.1151419)
39. Bettencourt L. 2013 The origins of scaling in cities. *Science* **340**, 1438–1441. (doi:10.1126/science.1235823)
40. Bettencourt L, West G. 2010 A unified theory of urban living. *Nature* **467**, 912–913. (doi:10.1038/467912a)
41. Batty M. 2013 *The new science of cities*. Boston, MA: MIT Press.
42. Arcaute E, Hatna E, Ferguson P, Youn H, Johansson A, Batty M. 2015 Constructing cities, deconstructing scaling laws. *J. R. Soc. Interface* **12**, 20140745. (doi:10.1098/rsif.2014.0745)
43. Alves LGA, Ribeiro HV, Lenzi EK, Mendes RS. 2013 Distance to the scaling law: a useful approach for unveiling relationships between crime and urban metrics. *PLoS ONE* **8**, e69580. (doi:10.1371/journal.pone.0069580)
44. Makse HA, Andrade JS, Batty M, Havlin S, Stanley HE. 1998 Modeling urban growth patterns with correlated percolation. *Phys. Rev. E* **58**, 7054–7062. (doi:10.1103/PhysRevE.58.7054)
45. Louf R, Barthelemy M. 2013 Modeling the polycentric transition of cities. *Phys. Rev. Lett.* **111**, 198702. (doi:10.1103/PhysRevLett.111.198702)
46. Louf R, Barthelemy M. 2014 How congestion shapes cities: from mobility patterns to scaling. *Sci. Rep.* **4**, 5561. (doi:10.1038/srep05561)
47. Louf R, Barthelemy M. 2014 Scaling: lost in the smog. *Environ. Plan. B, Plan. Des.* **41**, 767–769. (doi:10.1068/b4105c)
48. Rosvall M, Bergstrom CT. 2008 Maps of random walks on complex networks reveal community structure. *Proc. Natl Acad. Sci. USA* **105**, 1118–1123. (doi:10.1073/pnas.0706851105)
49. Ripley BD. 1976 The second-order analysis of stationary point processes. *J. Appl. Prob.* **13**, 255–266. (doi:10.2307/3212829)
50. Wilson AG. 1969 The use of entropy maximising models, in the theory of trip distribution, mode split and route split. *J. Trans. Econ. Policy* **3**, 108–126. (doi:10.2307/20052128)
51. Decraene J, Monterola C, Lee GKK, Hung TGG, Batty M. 2013 The emergence of urban land use patterns driven by dispersion and aggregation mechanisms. *PLoS ONE* **3**, e80309. (doi:10.1371/journal.pone.0080309)
52. Schelling T. 1971 Dynamic models of segregation. *J. Math. Sociol.* **1**, 143–186. (doi:10.1080/0022250X.1971.9989794)
53. Krugman P. 1996 *The self-organizing economy*. Cambridge, MA: Blackwell.
54. Feng X, Levinson D. 2011 *Evolving transportation networks*. New York, NY: Springer Science & Business Media.
55. Roth C, Kang SM, Batty M, Barthelemy M. 2011 Structure of urban movements: polycentric activity and entangled hierarchical flows. *PLoS ONE* **6**, e15923. (doi:10.1371/journal.pone.0015923)