Appl. Spatial Analysis
https://doi.org/10.1007/s12061-018-9264-8

New Forms of Data for Understanding Urban Activity in Developing Countries

Ed Manley[1] · Adam Dennett[1]

Received: 25 April 2017 / Accepted: 9 July 2018

Abstract Emerging forms of data offer new opportunities for developing a deeper understanding of poorly understood social and spatial processes. This is no more important than in developing countries, where large-scale data collection and processing has been relatively limited. In this paper, we explore how two new datasets can be used to enhance our understanding of human activity and communication interactions in Dakar, Senegal. Starting from a premise of little contextual knowledge about the setting in which we are working, we explore how much these data, combined with novel quantitative methods, are able to inform us about the urban environment in question. Fine-grained infrastructural data are combined with k-means clustering to produce an 11-class land use classification, distinguishing dense and sparse, single and mixed use regions. Using these classifications and over 1.5 billion mobile phone call records, patterns of activity and interaction within and between land use types are analysed. These analyses reveal strong activity associated with high density commercial, governmental and administrative regions. These regions are also identified as relatively strong 'attractors' of communication, and wider patterns show higher interactions between areas with similar land use characteristics. Analyses of dynamic activity and interaction patterns highlight the movement of people from workplace zones to residential areas. Importantly, these analyses reflect the expected patterns of urban activity, providing some validation of the data, methods, and the empirical approach. The paper concludes in addressing the strengths and potential of these approaches, while recognising current limitations and areas for further work.

Keywords Mobile phone data · Urban activity · Africa · Communication · Clustering

✉ Ed Manley
ed.manley@ucl.ac.uk

[1] UCL Centre for Advanced Spatial Analysis, University College London, Gower Street, London, UK

Published online: 30 July 2018

Introduction

The widening availability of fine-grained data provides new opportunities to measure and understand the world in greater detail. Within just the last 10 years, the breadth and variety of datasets available to urban planners, city governments, transport operators, and researchers, has expanded significantly (Batty 2013; Arribas-Bel 2014). This has led to a proliferation of new methods which seek to extract and classify elements of urban activity from singular data sources (Steenbruggen et al. 2013; Salesses et al. 2013; Goel et al. 2018). These new forms of analysis describe cities in greater spatial and temporal detail than has ever previously been possible, offering the prospect of improved governance, decision making and policies in cities (Goldsmith and Crawford 2014).

Much of the research involving new datasets and methods has focused on activity in Western, East Asian and South American countries. This may be an outcome of the interests of the academic community, but is more likely primarily because these are the locations where datasets have been made available to researchers. The result is a growing inequality between our understanding of urban contexts in the developed and developing worlds. While it can be assumed that some transferable (or universal) insights can inform development, we currently lack adequate insight and context around the specific case of African urban activity and dynamics.

Understanding the human dimension of cities – certainly from an aggregate, demographic perspective rather than a more micro-level qualitative perspective – has traditionally been facilitated in Europe and North America by the collection of spatially-referenced population data. Historically these data would have been collected using censuses – full population enumerations taken at a given point in time, usually with a decennial refresh-rate. While countries like the Canada and the UK still collect data in this way, increasingly in Europe in countries like the Netherlands, Finland and Sweden, population registers (which are continually updated) are becoming common place. In addition to centrally collected population data, social surveys, administrative data (for example hospital patient registers or tax records) and privately collected ‘consumer’ data (from loyalty cards, credit checks or customer surveys) are being used to gain even greater insight into the patterns and processes exhibited by humans in cities.

Yet, there are notable limitations associated with the conventional census process. Census datasets can be useful for understanding static demographic profiles, snapshots of daily commuting behaviours or less frequent migration transitions over longer periods. But understanding population dynamics at a much finer temporal resolution (for example the changing commuting patterns at different times of the day or days for the week) has required the use of additional surveys. In addition, within areas of rapid population change, census data can rapidly become obsolete. Censuses are also costly and sophisticated operations to conduct, and may be particularly difficult to conduct within the developing country context. The national government must have the technical infrastructure in place to disseminate, collect, and process census questionnaires, but furthermore maintain technical expertise in sufficient volume and depth to ensure adequate data quality. In Senegal, the setting for this study, censuses have been carried out relatively irregularly (in 1988, 2002, and 2013). The latest census was supported by the United Nations Population Fund with funding for tablet-based collection. Some developing countries will face the additional challenges born from a lack of prior measurement of housing or infrastructure, significant population heterogeneity in literacy levels and languages spoken, the presence and location of nomadic people.

Increasingly, even in countries where Census or register data provide rich demographic information, alternative sources of data are being experimented with. Administrative or data which are the by-product of other processes (such as making telephone calls or interacting on social media) are particularly attractive as, while frequently less rich in detail, they do not incur the costs associated with commissioned data. In recent years, mobile phone transaction data have been suggested as an alternative or proxy measure of population distribution (Frias-Martinez and Virseda 2012). These data record the locations and times at which users of mobile phones undertake activity using their devices. These transactions are served via cell phone towers, with the sum of all activity yielding a distribution of the complete user base of a single mobile phone service provider. By weighting against biases in this user base (for example, while ubiquitous now, in the early years of adoption mobile telephones were predominantly used by younger people), a theoretical proxy for population distribution is achieved. Furthermore, the temporally dynamic nature of this data means that we can capture mobility in addition to location.

The opportunities afforded by this form of data in the spatiotemporal analysis of urban areas have been demonstrated elsewhere in a number of ways. At the urban-wide scale, various projects have shown how mobile phone traces describe the movement of populations over the course of the day. The Mobile Landscapes (Ratti et al. 2006) and Real Time Rome (Calabrese et al. 2011) projects were early runners in this respect. Others have shown how it is possible to differentiate land uses from the temporal signatures of activity at different locations (Toole et al. 2012; Pei et al. 2014), identify origin-destination matrices of population flow (Iqbal et al. 2014), partition space in communities of interaction (Ratti et al. 2010; Expert et al. 2011), and establish centricity in urban structure based on patterns of activity and interaction (Louail et al. 2014), while analysis of fixed-line telephone communications has been shown to be a helpful footprint for identifying sectoral specialisms in industrial activities in the UK (Reades and Smith (2014). A wider review of the extensive volume of work conducted in this area can be found in Blondel et al. (2015).

This form of data furthermore enables the analysis of individual-level mobility patterns. Research in this area has helped demonstrate wide scale patterns of regularity in activity at the individual level (Gonzalez et al. 2008; Song et al. 2010). Other research has extended these findings to demonstrate the predictability of behaviour, building on the past behaviours of individuals and wider population level mobility trends (Calabrese et al. 2010; Song et al. 2010). The analysis of both call activity and mobility at the individual scale have furthermore demonstrated how social ties are manifested in space (Wang et al. 2011).

The demonstration of the capability of mobile phone data has raised the prospect of their use within the data-poor contexts of the developing world. In Africa, in particular, rapid growth in the adoption of mobile phones – 6.1% in users over the last 5 years – has been raised as potentially beneficial in improving inclusion and boosting productivity (Garbellini and Okeleke 2017). Through mobile phone data we are presented with an opportunity to build a detailed spatial and temporal understanding of a city without the need to conduct a full census or commission a survey. Within the developing world context, this potentially allows rapid and low-cost measurement of the urban population, avoiding the cost of a complete population census and leapfrogging stages of development undergone by cities in the developed world.

The release of mobile phone data by Orange under the Data for Development programme in both Cote d'Ivoire and Senegal (used in the paper) have provided impetus for research in developing world context. Previous work using these datasets, encompassed within two volumes (see Data for Development 2013 and Data for Development 2015) demonstrate a focus of research around country-level communication and mobility patterns, communication network community formation, and epidemic networks. Relatively few of these studies have focused on the urban processes specifically however, with this research either addressing urban structure or commuting patterns (Louf, et al. in Data for Development 2015; Naboulsi, et al. in Data for Development 2013).

One of the challenges of deriving meaning from attribute poor but spatially rich interaction data such as the telecommunications interactions used here or by Reades and Smith (2014), taxi journeys (Peng et al. 2012) or flows of students to University (Singleton et al. 2010) is that without other contextual information on the origins and destinations, the interactions remain just flows of information, traffic or people. To overcome this, despite using data from quite different contexts, Reades and Smith (2014), Peng et al. (2012) and Singleton et al. (2010) all made use of different area classifications (derived from alternative data sources) situate the flows they were analysing. The challenge we face in Senegal is that contextual data is scarce and area classifications non-existent, and so developing a new classification based on available infrastructural data should provide a means to contextualise the telecommunication interactions observed. In turn, we might expect communication flows between different types of area to vary and so observed interactions should help us to evaluate any land use classification generated.

In this paper, we seek to expand on these previous studies, by integrating land use classification and mobile phone data, to derive a deeper contextual understanding of urban activity within a developing country. This work focuses on Dakar, the capital city of Senegal, and home to a population of 1.1 m people. The study makes use of the Data for Development dataset, as well as land use data obtained from the Senegalese statistical and mapping agency. The question we are seeking to answer is, in the absence of census or other demographic data, is it possible to derive useful information about the city, its inhabitants and their activities from data relating to the physical features of the city and the mobile phone usage of its inhabitants? If successful, then this could have implications for service planning, transport management and other city governance activities in areas of the world where conventional data are less widely available. At the time of initial analysis, census data for small areas the city had not been published, however in the course of completing this work, data for some administrative units has been published and so offers an opportunity for some comparison.

Following a description of the datasets used in the study in the next section, the methods used to classify land use and measure activity are described. In the fourth section, we describe the nature of the land use classification for Dakar, and what its spatial structure indicates about the nature of the city. Following this, we incorporate our measures of activity and communication interaction, derived from mobile phone data, to analyse how these vary with land use across Dakar. The paper concludes with a discussion of some of the pertinent outcomes of this work and potential avenues for future work.

Data Sources

The study makes use of two newly available datasets covering the whole of Dakar. The first is a spatially granular GIS dataset provided by the Agence Nationale de Statistique et de la Démographie (ANSD), the Senegalese statistical and mapping agency. The dataset, collected by the ANSD as part of the Projet de Numérisation des Districts de Recensement, contains point centroid data for 61 types of infrastructure for precise locations. The data details the locations of all major types of physical infrastructure as of 2014, including water and electricity networks, commercial units, government offices, industry, hotels, schools, markets, religious institutions, health centres, transport infrastructure, and so on. Additional spatial data relating specifically to the road network and the location of military bases (both not covered by the ANSD data) were obtained from OpenStreetMap.[1] These datasets will form the basis of the static analysis of the city.

The second dataset describes mobile phone activity across Dakar over the course of 7 months in 2013. Provided by Orange as part of the Data for Development programme, the dataset describes mobile phone interactions between each of the 435 Orange cell towers in Dakar. In this study, we make use of aggregated Call Detail Record (CDR) data. This dataset contains counts of voice calls, sum call durations, and counts of text messages between all origin and destination pairs, aggregated to each hour of the seven-months period. In total, the dataset contains records of 1,562,327,418 voice calls and 2,287,872,391 text messages between Dakar cell towers. It is estimated that 64% of the population of Senegal is a customer of Orange (de Montjoye et al. 2014), and we may expect higher proportions of the population of Dakar to be represented in this dataset, evidence from Kenya in 2009 suggested three quarters of urban dwellers had mobile phones compared to one third of rural residents (Aker and Mbiti 2010). This dataset will form the basis of an analysis of the dynamic elements of the city.

Methodology

The analysis of Dakar, through these new forms of data, consists of static and dynamic components. The large number or origin/destination pairs in interaction datasets means that analysis of the raw data can be affected by noise and large occurrences of small interactions. To reduce the dimensions of large interaction matrices and smooth out some of the noise, a number of authors have grouped origins and destinations into clusters defined by particular characteristics (Reades and Smith 2014; Dennett and Stillwell 2010). This process also aids the contextualisation of the observed flows as

[1] OpenStreetMap has been shown to provide 100% coverage of the road network in Senegal, relative to established cartographic sources (Mapbox 2016).

described above. Here we make use of the ANSD GIS data to develop a static representation of the city, defining areas of the city by local infrastructure characteristics. This representation is linked to the mobile phone data by spatial unit, as outlined below. The mobile phone transaction data yields indications of activity and interaction between regions over time, and, in reference to the static model of the city, suggests that nature of activity varies at different times of day. Importantly, as we will show, the CDR data both validate the infrastructure classification and supplement it with additional information about activity.

Area Unit Specification

The area units chosen for the basis of the analysis are selected at an aggregation suitable for both the ANSD GIS data and the Orange mobile phone data. Administrative units are available in Dakar, but even the smallest unit can accommodate several cell towers. Mobile phone data are aggregated to spatially located cell towers, and given their sparsity relative to the point-level GIS data, appeared a good basis for the development of an area unit definition. Indeed, many earlier analyses of mobile phone data have been conducted at the level of cell towers (Frias-Martinez and Virseda 2012; Iqbal et al. 2014).

The development of areal units from cell tower locations is, however, less than straight forward. Cell tower reception range is dependent on factors relating to the location, height, and the technology involved. While limitations in range are accounted for by telecom operators, they necessarily result in overlapping service areas to minimise the risk of 'holes' in coverage. A device will seek the cell tower providing the strongest, and therefore usually nearest, signal (Electronic Notes 2017), but connection or handover to the nearest cell tower is not guaranteed. As a result, due to the uncertainty on their proximity to any specific cell tower, the location of a specific user cannot be assured. Nevertheless, at an aggregate level it can be expected that cell tower activity is indicative of nearby activity.

The analysis of mobile phone data has taken a number of approaches, each creating a link between space and a nearby cell tower. Steenbruggen et al. (2015) find that previous research follows one of three approaches – best-serving polygons, rasterisation, and Voronoi tessellation. Best-serving polygons and rasterisation are representations of the areas covered by specific cell towers, and are estimated by the mobile service provider on the basis of tower capabilities and nearby urban form. Voronoi tessellations are a simplified representation of service zones, creating regions associated with cell towers based on nearest proximity alone.

In the case of this study, only cell tower locations were provided, without an indication of the best-service zone, and as such, Voronoi tessellation is the best option. Using the 435 cell towers located within the Dakar region, area unit representations generated through Voronoi tessellation are adopted.

Area Classification

Area classifications in human geographic analysis have been typically carried out using population data. These geodemographic classifications (Harris et al. 2005; Vickers and Rees 2007; Breetzke and Horn 2009) generally use census data as their principal input due to the level of attribute and spatial detail available, however a number of classifications, particularly commercial general use classifications such as Acorn (CACI) and Mosaic (Experian), use commercial and other data to increase discriminatory power. The focus on population data is indicative of the use to which these area classifications are usually put – the segmentation of different population groups and their characteristics for commercial or other purposes. While land-use is a function of human activity and closely related to it, area classifications for land use have typically been the focus of and used in other sub-disciplines of geographic analysis. In the field of remote sensing, for example, there is a long history of area classifications according to land-use (Anderson et al. 1976) with more recent work focusing on differentiation in urban areas (Herold et al. 2002), however these classifications tend to be limited to broad land cover types and poor at differentiating, for example, residential and commercial areas. More nuanced land-use classifications are frequently produced and have been for some time (Shapiro 1959), but they tend to be generated through more qualitative processes. To our knowledge no work has been carried out on defining land-use classifications using infrastructural data and certainly not to explore human activity patterns in urban areas.

The classification phase aims to generate classes representative of area function, enabling the identification of the types of activities on going in each polygon area. The classification scheme is derived from the point centroid GIS dataset provided by ANSD, which contains feature data relating to 61 infrastructure types.

Regional Land Use Attributes

Land use attributes are derived through the calculation of feature densities in each polygon region. These measures of feature density will enable the identification of land use classifications, enabling differentiation in area type. This phase of the methodology takes three stages:

1. **Assignment of features to polygons**. A simple point-in-polygon operation is executed over all 61 feature types and 435 area polygons, yielding counts of features by area.
2. **Consolidation of feature types into categories**. Given the high number of feature types in the original dataset, there was risk of poor cluster formation due to the 'curse of dimensionality' (whereby large sparsity in data results in a reduced ability to identify important trends), and therefore some consolidation of features was required. The process of consolidation was subjective, and combined features according to their functional similarity relative to other groups, with new groups being created where clear distinctions arose. For example, the category 'commercial' was created by combining data for workshops, bakeries, shops, boutiques, and markets. The final category list captures all feature types within 17 categories, aiming to provide a broad classification of all relevant land use types. These categories are as follows – commercial, industrial, business administration, government, leisure, religious, education, medical, transport, hotels, percentage piped water, wells, science and research, diplomacy, roads, military, and nightlife. Feature counts are summed across these categories for each area.

3. **Z-Score Normalisation**. Given significant variance in feature counts across categories, z-scores are calculated by category for each area. This calculation takes the standard form $z = \frac{x-\mu}{\sigma}$, where x is the feature count in a given region, μ is the mean feature count, and σ the standard deviation.

Classification Process

A regional classifier is generated using the normalised feature counts created according to the process above. To create the classifier, a k-means clustering algorithm is applied (Everitt et al. 2001). The k-means method created clusters based on spatial proximity, meaning features with similar measures across all attributes are assigned to the same partition. The number of partitions generated by the algorithm is determined by the parameter k, which is set by the modeller. In addition to selecting k, it must be decided which attributes of the data best contribute towards the clustering, and create meaningful, well-defined clusters.

The parameterisation of k and the selection of attributes is led through metrics measuring the structure and cohesiveness of resulting clusters. The first metric used is the coefficient of variation, $c_v = \frac{\sigma}{\mu}$ where μ is the mean cluster size, and σ represents the standard deviation. This measure yields a relative measure of variance in cluster size, with higher values highlighting misbalancing of cluster sizes. The second metric calculated is the Silhouette score for each clustered feature. The Silhouette score, S, measures how well each feature fits within its designated cluster, relative to other clustered points, where positive scores indicate a well clustered point (limited to +1), and negative scores a poorly allocated point (limited to −1). For evaluation, both individual feature silhouette scores (specifically negative scores), s_{neg}, and the mean silhouette score, S, are used.

The process of choosing k and selecting the attributes is undertaken iteratively, and guided by the modeller, in reference to c_v and Silhouette scores. In cases where a single dominant cluster captures a large proportion of features, a secondary clustering phase is executed upon the largest cluster alone. This secondary level of clustering enables better identification of less distinct regions in land use, such areas of mixed use activity. The two phases of clustering will henceforth be given the terms k_1 and k_2.

Measures of Activity and Interaction

With regional classifications constructed around the Voronoi regions of cell towers, cell phone communication patterns can describe activity and interactions at and between different land use types. Furthermore, the granularity of communication data allows activities and interactions to be analysed between land use types on a fine temporal basis. In defining these metrics, given variation in the size of different land use areas (normalisation on this basis is provisioned).

Regional activity is measured by counts of phone calls initiated in origin region. Activity counts A_i are normalised by both the sum of the area of the region, A_iS, and the sum of the count of road junctions, A_iR, as below. Spatial area will account for differing area size, while road junctions account for density in the built environment, which better accounts for open spaces and is indicative of urban density (Borruso 2003).

$$A_iS = \frac{\sum_{j=1}^{n} C_{ij}}{S_i} \tag{1}$$

$$A_iR = \frac{\sum_{j=1}^{n} C_{ij}}{R_i} \tag{2}$$

Where C_{ij} is the number of phone calls between region i and region j, S_i is the area of region i in square kilometres, and R_i is the number of road junctions in region i. These measures can be presented over all time periods, or disaggregated to hourly measures of activity. By only considering call origin locations, we focus only on instances where users have actively engaged in communication.

Regional interactions are measured by the proportion of calls an origin region makes to any destination region, C_{ij}/A_i. This measure is normalised as above, in this case by alternatively the sum area, S_j, and count of roads within the destination region, R_j, as outlined below. Given the control for origin size through normalisation against all activity, the metric will capture relative attraction to each destination area, and thus only destination size measures (junction counts and area) is used to normalise call activity.

$$A_{ij}S = \frac{{}^{C_{ij}}/_{A_i}}{S_j} \tag{3}$$

$$A_{ij}R = \frac{{}^{C_{ij}}/_{A_i}}{R_j} \tag{4}$$

This measure yields a weighted proportion of calls between an origin and destination region, accounting for both origin call volume and destination area size or road count.

Area Classification Results

The generation of the land use classification, in line with the methodology described above, was carried out within an iterative process. At each iteration the area attributes, assigned across 17 categories, were refined in seeking the best classification fit. Given the relatively few features, the process was modeller-led and non-exhaustive. This approach had the advantage of maintaining features common to land-use classifications, and limiting the double counting of correlated features where identified. Testing proceeded through incremental inclusion or exclusion of singular attributes based on resulting metrics (specifically, c_v, S and s_{neg}). As such, the final attribute selection is not necessarily optimal, but yields classifications based on diverse, uncorrelated features. During each iteration, the c_v, S and s_{neg} metrics described above were extracted, and attribute selection refined according to these measures. Clusters were constructed for values of k_1 between 2 and 14, and k_2 between 2 and 5.

Through attribute selection, clustering was executed on the relative densities of business administration, government, industrial, commercial, science and research, hotels, percentage piped water, wells, and military facilities in each area. Rejected features, such as education, medical, transport, diplomacy, and religious facilities, were shown not to contribute effectively in improving land-use classifications.

Using these attributes, it was found that at its peak $S = 0.27$ when $k_1 = 8$, with $c_v = 1.70$ and $s_{neg} = 152$. By comparison, where $k_1 = 7$, $c_v = 1.61$, $S = 0.25$ and $s_{neg} = 152$, and where $k_1 = 9$, $c_v = 1.56$, $S = 0.18$ and $s_{neg} = 187$. The mean Silhouette scores are relatively lower than we might expect, however, observing the outputs, it is clear that secondary clustering was required on the largest component in all cases. Observing the c_v scores for the secondary clusters where $7 \leq k_1 \geq 9$, low points in c_v are reached where $k_1 = 8$ and $k_2 = 4$ ($c_v = 0.69$) and where $k_1 = 9$ and $k_2 = 4$ ($c_v = 0.63$). In reference to all metrics, generated from both primary and secondary clustering phases, it was decided that the clustering configuration defined where $k_1 = 8$ and $k_2 = 4$ be adopted. This results in 11 cluster types being used going forward.

The 11 land use classes generated through this process can be interpreted by observing the density of each attribute within each class. This is achieved by calculating the mean z-score for each attribute across all Voronoi regions assigned to each cluster. Table 1 provides an indication of the stronger and weaker associations between each attribute with each cluster, and indicates the number of regions associated with each

Table 1 Indicates association of clusters with mean attribute density (calculated using z-scores, where + > 0.25, ++ > 1, +++ > 3, − < −0.25, — < −1 and —— < −3), and cluster size by region count

Attribute	Primary clusters							Secondary clusters			
	1	2	3	4	5	6	7	8	9	10	11
Administrative		+	++		+++		–			–	
Government		+	++	+	+++		–			–	
Industrial			++		+++	+	–	++	–	–	–
Commercial	–		+++	–	–	++		–	+	–	–
Science				+++							
Hotels			+++		+++			+	–		
Piped Water	–	++	++		++	+	–	–	–	–	–
Wells					–		+++	–			++
Military	+++										
Total Regions	13	37	10	20	3	55	17	34	78	135	33
Sum Area in sq. km.	4.54	5.87	1.19	5.07	0.29	9.94	9.95	13.29	21.92	80.14	20.06

cluster. Figure 1 shows the spatial representation of these clusters across the Dakar region.

In interpreting these clusters, it is clear from the results that a large degree of mixed use activity across Dakar, with a number of clusters demonstrating strong association with multiple feature types. It is apparent that the process has successfully distinguished some single use regions, such as clusters 1 (Military), 4 (Science and Research) and, to a lesser extent, cluster 8 (Industrial). The last of these shows indications of strong local spatial autocorrelation in Fig. 1, and thus appears to validate the presence of a single use region.

The cases of clusters 7 (Fringe Residential and Commercial) and 11 (Fringe Residential) are more curious, as these highlight a strong association with the locations of wells alone. Piped water is available across Dakar, and wells therefore may be indicative of informal or recently constructed regions. Given the lack of strong association with other features, and their location at the fringes of the city (as can be seen in Fig. 1), it is likely that these are indicative of informal, residential areas.

The remaining clusters appear indicative of mixed use commercial and administrative regions, with cluster 3 (Dense Commercial and Hotels) and 5 (Central Government and Business District) most strongly linked to government offices and hotel facilities. Again, as shown in Fig. 1, there appears to be strong spatial association between cluster types, with cluster 5 indicative of a Central Business District and Governmental Region, and cluster 3 signifying nearby, dense commercial regions. Cluster 9 appears to represent a lower density demonstration of this trend of mixed use activity, located towards the outer reaches of the city. Finally, cluster 10 (Urban Residential and Open Spaces) appears representative of the remaining regions where no clear land use can be

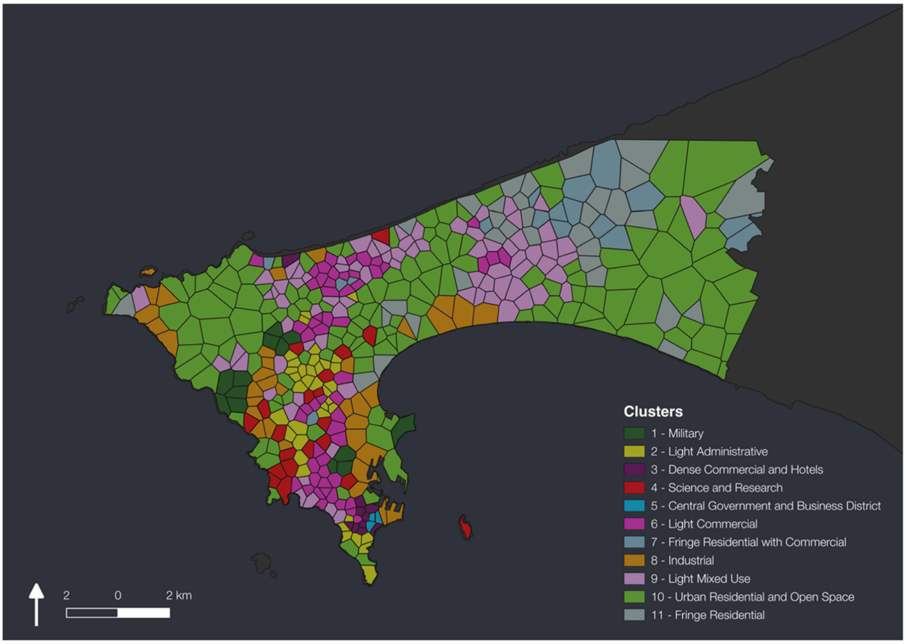
Clusters
1 - Military
2 - Light Administrative
3 - Dense Commercial and Hotels
4 - Science and Research
5 - Central Government and Business District
6 - Light Commercial
7 - Fringe Residential with Commercial
8 - Industrial
9 - Light Mixed Use
10 - Urban Residential and Open Space
11 - Fringe Residential
2
0
2 km

Fig. 1 Map of spatial distribution of land use types

defined. Given the lack of strong association with any attribute, it could be concluded that these areas are representative of residential areas and remaining open spaces.

Using the associations found in Table 1, and the trends observed in Fig. 1, clusters are given names indicative of their function. These names and corresponding cluster numbers are shown in Fig. 1. There are naturally notable limitations to the land use classification approach introduced here. The absence of small area residential data reduces the ability to determine the degree of mixed use activity on going within each area. As shown in the results, some areas appear to suggest a high degree of residential use (e.g. a lack of other infrastructure), but confirming this is difficult.

Recent data from the last Senegal Census offers some opportunity for validation. While only raw counts are available for the smallest areas (which as still far larger than the Voronoi polygons of our classification groups) it is possible to show population densities for the areas with cell tower polygons overlaid on top for comparison (Fig. 2). Observing the cluster groups in Fig. 1, it can be seen that some of the most densely populated areas in Dakar correspond with those areas classified as Light Commercial or Light Mixed Use (clusters 6 and 9). Less densely populated areas appear to correspond with Urban Residential and Open Spaces, Fringe Residential and Industrial areas (clusters 10, 11 and 8). While visual comparison is not in any way scientific, we do at least see population densities corresponding with the kinds of land use we would expect.

There are other difficulties found in relation to the handling of open and restricted spaces within the city. Some areas of Dakar are given away to marshland, open land, or large infrastructure (e.g. airport, large road junctions). In order to maintain a spatially contiguous classification these areas are included within the classification, but like residential areas, data is lacking on their exact extent. Finally, it is inevitable that with

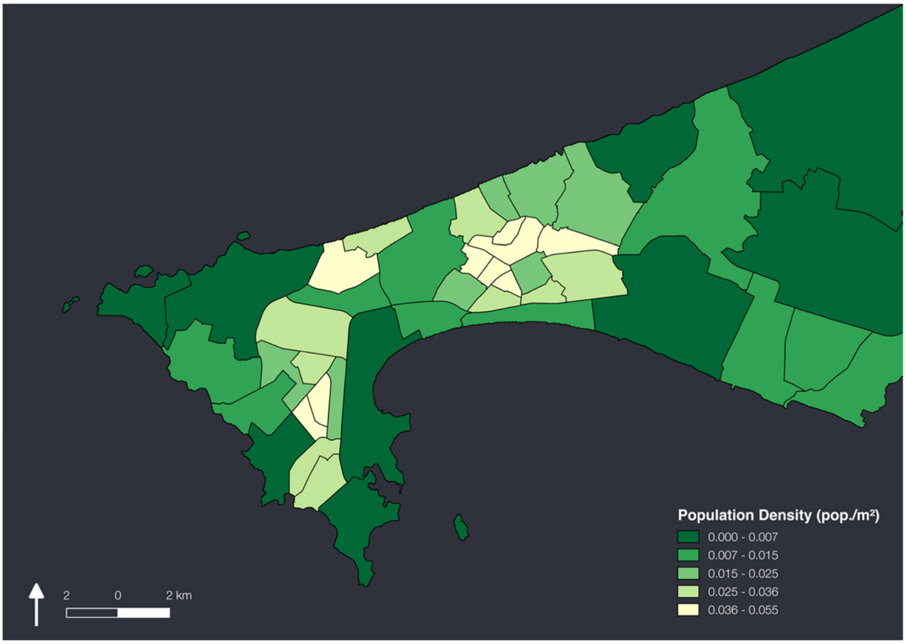

Fig. 2 Census-derived population densities across Dakar

any regionalisation of space that the Modifiable Areal Unit Problem (MAUP – Openshaw 1984) will result in some misclassification. This is noticeable in locations where agglomerations of features do not align with the definitions of the Voronoi polygons. An example of this can be observed on the north coast of city, where around 20 hotels line the coast, but are separated into three regions resulting in only one of these polygons being strongly associated with hotels.

Activity and Interaction

The other source of data we are able to use to verify the land-use cluster groupings is, of course, the Orange mobile telephone activity dataset. The data are mutually reinforcing such that the definition of land use types from cell tower areas, as shown above, also enables the exploration of trends in activity and communication interaction yielded by mobile phone usage.

Land Use Activity

As outlined in the Methodology, activity can be observed both in absolute and relative terms. Activity in these cases is measured by the number of phone calls made or received within each region type during 2013. It is assumed that volumes of calls can be used as a proxy for the location of individuals, and thus indicate how activity varies over space.

Figure 3 shows how call activity varies by land use type, broken down by calls made and calls received. It is firstly clear, and perhaps not too surprising, that there is little difference across each type of region in the numbers of calls made and received. It is also apparent that call volumes tend with the sum area of land use type. The highest

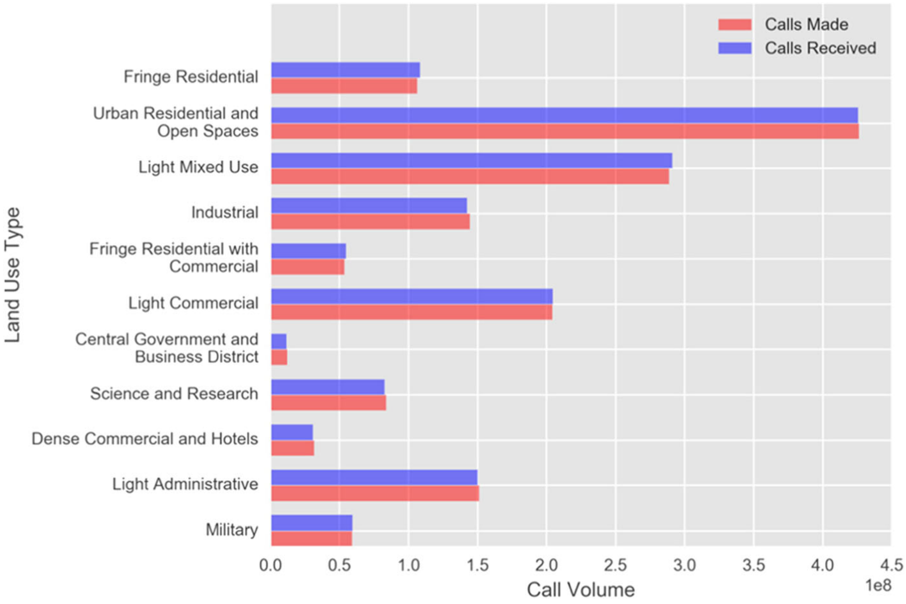

Fig. 3 Chart showing call volumes by land use type

volumes of activity are observed in the Urban Residential and Open Spaces and Light Mixed Use areas, the two largest regions by area (see Table 1), with 426 million and 288 million calls made and received respectively. The lowest levels of activity are associated with the smallest regions, the Central Government and Business District (12 million calls made and received) and the Dense Commercial and Hotels area (31 million).

Greater insight is undoubtedly gained when one observes activity normalised by the sum area size of the region. As can be seen in Fig. 4, this changes the picture considerably, leading to the highest density of activity being observed within the Central Government and Business District, located in Dakar's southern region. Similar regions – such as the Dense Commercial and Hotels and Light Administrative – are also well represented in terms of activity. This would appear to align with expectations, given the likely 'pull' of these regions induced by a greater density of employment opportunities. On the other hand, the larger regions by area exhibit a lower level of activity by density, with the Urban Residential and Open Spaces and Fringe Residential zones ranking lowest on this scale. Importantly, through alignment with expectations, these trends in mobile phone activity validate the land-use classifications derived through clustering.

This form of normalisation, of course, weights against those areas with large open spaces. Within the context of the city constrained by geography – as Dakar is, impacted by marshland, cliffs, and beaches – this could misrepresent the basis of activity. To verify the trends in Fig. 4, we also explore normalisation by road junction density. In doing so, only those locations with purposefully built infrastructure contribute towards our analysis. As shown in Fig. 5, many of the same trends are observed, although with a reduced degree of variance. Many of the regions here exhibit similar properties (varying between around 40,000 to 80,000 calls per road junction). While the lower

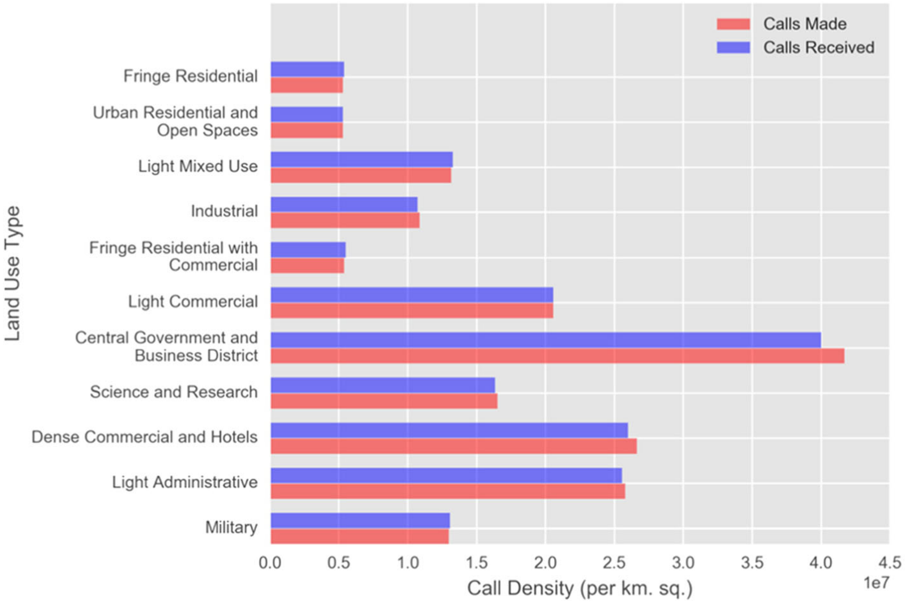

Fig. 4 Chart showing call density (normalised by area size) by land use type

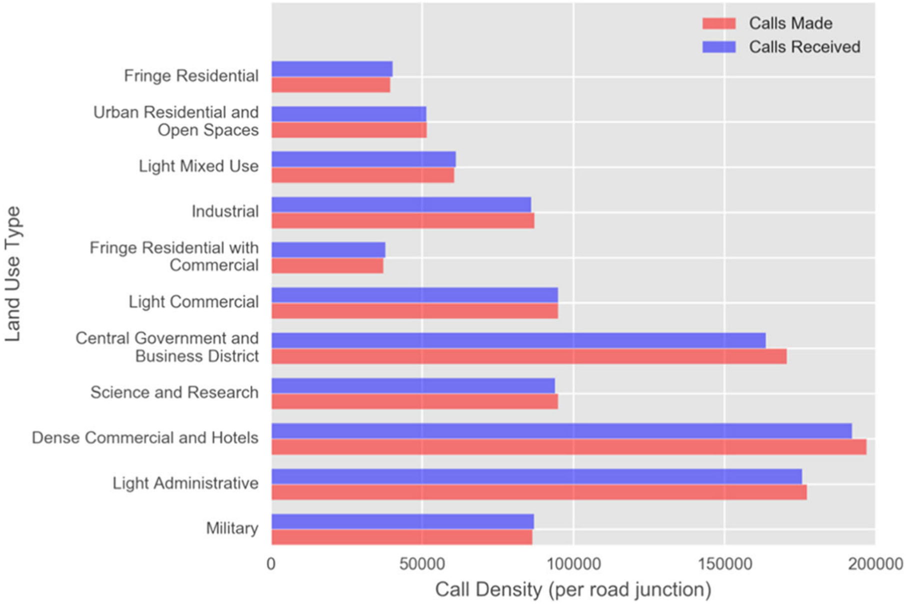

Fig. 5 Chart showing call density (normalised by road junction count) by land use type

density residential areas rank lowest again, the impact of open spaces is mitigated. The three most active areas observed above – the Central Government and Business District, Dense Commercial and Hotels area, and the Light Administrative areas – are once more stand out here, demonstrating significantly greater levels of activity, at between 150,000 and 200,000 calls per road junction. This second view strengthens the case around which are the most and least active regions of Dakar, and signifies the types of functions that may be instigating these levels of activity. These data provide further validation for the specification of the land use classifications.

Interactions between Classes

Further insight can be built through an exploration of the interactions between land type classes. Rather than only measuring the numbers of calls made and received within a region, interactions are measures as calls made between areas (i to j). This will help highlight where there are stronger and weaker interaction relationships between land use types. Given the trends found above, interaction analyses will deal mainly with normalised activity, thus reducing the impact of region size, however we will begin by modelling flow volumes using a spatial interaction model in order to generate a set of call interaction estimates with which to compare our observations. When discussing normalised interactions, normalisation is made according to the properties of the destination region, as outlined in Section "Measures of Activity and Interaction".

Model of Call Interaction

In order to generate estimates of likely flows, we make use of a doubly-constrained spatial interaction model (Wilson 1971). In this model, all estimated call interactions between origins and destinations are constrained to known total call volumes sent from origins and received from destinations. Estimates for the flows are obtained via a bi-proportional fitting algorithm, using the distance between origin and destination cell tower regions as a distribution factor, assuming that call volumes will decrease with distance. This may seem counter-intuitive given that telecommunications were designed to facilitate interaction where physical distance is a barrier, however, empirical observations when fitting the doubly constrained model confirm that distance does indeed have a negative impact on call volumes.

The model takes the following form:

$$C_{ij} = O_i A_i D_j B_j exp\left(-\beta d_{ij}\right) \tag{5}$$

Where C_{ij} are the calls between regions i and j, d_{ij} is the Euclidean distance between them, with total in and out flows to constrain estimates to calculated as follows:

$$O_i = \sum_j C_{ij} \tag{6}$$

$$D_j = \sum_i C_{ij} \tag{7}$$

With balancing factors to insure these constraints hold:

$$A_i = \frac{1}{\sum_j D_j B_j exp\left(-\beta d_{ij}\right)} \tag{8}$$

$$B_j = \frac{1}{\sum_i O_i A_i exp\left(-\beta d_{ij}\right)} \tag{9}$$

β is the distance decay parameter to be estimated, which in this case is achieved through re-formulating Eq. 5 as a log-linear model (see Oshan 2016) and using a negative binomial regression model fitted using the MASS package in R to estimate flows and parameters. By aggregating estimated flows and residuals to land-use-type clusters for origins and destinations, it is possible to see where call volumes are higher or lower than expected between and within clusters across the city of Dakar (Table 2).

In Table 2, where flows are positive, it indicates that more calls are observed than the model predicts; where they are negative, volumes are much smaller. Immediately apparent is that far more calls are occurring within every land-use cluster than would be expected. This perhaps suggests while these clusters have a certain amount of land-use homogeneity, there is also a homogeneity in the activities undertaken by people in these areas – industrial activity communicating with industrial activity, residents with residents, etc. In particular, far more calls are occurring within the central residential areas (Urban Residential and Open Spaces, Light Mixed Use and Light Commercial). If

Table 2 Modelled call volume residuals (in 1000 s) between land-use clusters in Dakar

Origin/Destination residuals	Central government and business district	Dense commercial and hotels	Fringe residential	Fringe residential with commercial	Industrial	Light administrative	Light commercial	Light mixed Use	Military	Science and research	Urban residential and open spaces
Central government and business district	1150	129	−97	−19	−230	39	22	−270	−5	59	−323
Dense commercial and hotels	136	4431	16	−21	−724	−566	1071	−781	−260	−229	−881
Fringe residential	−27	117	16,254	946	−1532	−977	−2666	−3485	−454	−797	−2081
Fringe residential with commercial	14	30	1256	8265	−474	−318	−317	−595	−53	−255	−729
Industrial	−260	−808	−2134	−658	21,300	2296	−494	−2805	425	484	−2048
Light administrative	−1	−662	−1612	−462	2275	21,469	−181	−2561	1658	1858	−3083
Light commercial	48	1073	−3225	−800	94	333	33,706	3009	−203	869	−1457
Light mixed use	−195	−677	−3781	−984	−1823	−1651	4000	44,786	−656	−1731	476
Military	−4	−236	−650	−102	399	1572	−378	−1075	7276	110	−456
Science and research	55	−268	−1138	−336	484	1800	526	−2388	112	11,265	−1733
Urban residential and open spaces	−217	−785	−3192	−1494	−646	−1639	−1217	−1115	134	−965	57,066

we look at the interaction between clusters, far fewer calls than expected are occurring between the Fringe Residential areas and these other residential areas, suggesting a degree of isolation between the fringe (and likely more informal settlements) and more well-established parts of the city.

Observations of Normalised Interactions

The high degree of intraregional interaction is sustained after call volumes are normalised. Taking area normalisation first (Table 3), this can be observed in seven of the 11 regions. For road junction normalisation, as shown in Table 4, again seven of the 11 regions exhibit strongest interaction with areas of the same land use type. These trends demonstrate higher levels of communication interactions between areas of the same land use type, confirming our observations from the model residuals that land use function is a driver of communication. This is not only an important finding, but furthermore provides strong validation of the land use definitions constructed from the ANSD dataset.

Where intraregional communications are ignored, one begins to observe the stronger communication attraction affect imposed by the high activity regions noted earlier. Where area normalisation is used, as one can see in Table 3, the Central Government and Business District is the most used call destination for all regions. This again highlights the importance of this small region as a centre for activity and communication in Dakar. Taking the road junction normalisation measures shown in Table 4, a more mixed picture can be observed. The Central Government and Business District is a strong destination for calls made from Dense Commercial and Light Administrative

Table 3 Strong interactions for each region, including and excluding intraregional communication, calculated using area normalisation

Call origin	Highest interaction ($A_{ij}S$)	Highest interaction excluding intraregional communication
Central Government and Business District	Central Government and Business District	Dense Commercial and Hotels
Dense Commercial and Hotels	Dense Commercial and Hotels	Central Government and Business District
Light Administrative	Light Administrative	Central Government and Business District
Science and Research	Science and Research	Central Government and Business District
Military	Military	Central Government and Business District
Light Commercial	Light Commercial	Central Government and Business District
Industrial	Central Government and Business District	Central Government and Business District
Fringe Residential with Commercial	Fringe Residential with Commercial	Central Government and Business District
Urban Residential and Open Spaces	Central Government and Business District	Central Government and Business District
Fringe Residential	Central Government and Business District	Central Government and Business District
Light Mixed Use	Central Government and Business District	Central Government and Business District

Table 4 Strong interactions for each region, including and excluding intraregional communication, calculated using road junction count normalisation

Call origin	Highest interaction ($A_{ij}R$)	Highest interaction excluding intraregional communication
Central Government and Business District	Central Government and Business District	Dense Commercial and Hotels
Dense Commercial and Hotels	Dense Commercial and Hotels	Central Government and Business District
Light Administrative	Light Administrative	Central Government and Business District
Military	Military	Light Administrative
Science and Research	Science and Research	Light Administrative
Fringe Residential with Commercial	Fringe Residential with Commercial	Dense Commercial and Hotels
Industrial	Industrial	Light Administrative
Light Commercial	Dense Commercial and Hotels	Dense Commercial and Hotels
Light Mixed Use	Dense Commercial and Hotels	Dense Commercial and Hotels
Fringe Residential	Dense Commercial and Hotels	Dense Commercial and Hotels
Urban Residential and Open Spaces	Dense Commercial and Hotels	Dense Commercial and Hotels

areas, which appears to align with the observed relationship between shared function and communication intensity. The majority of regions, however, interact most strongly with the Dense Commercial and Light Administrative areas. Once more these trends align with shared functionality, with commercial districts demonstrating strong interconnection. It is interesting to note that three regions exhibiting single use – namely Military, Industrial, and Science and Research regions – are most strongly connected to the Light Administrative region. This may be explained by the absence of other regions with similar land use characteristics, thus resulting in stronger relative interactions with administrative and mixed use areas.

More widely, a view of all interactions between regions is shown in Fig. 6. This network representation shows the relative normalised flow between all regions, using road junction normalisation. This representation highlights once more the central importance to all regions of the Central Government and Business District and the Dense Commercial and Hotels area. It furthermore indicates that the strongest levels of interaction are found in communication flows between these two dominant regions. The lower importance of the Fringe Residential areas as an attractor of communications is clearly indicated too, in the form of practically no strong connections terminating at these regions.

Activity and Interaction over Time

Using the methods outlined earlier we have shown how we are able to identify specific areas of high activity and interaction. Now we turn to how these patterns vary over

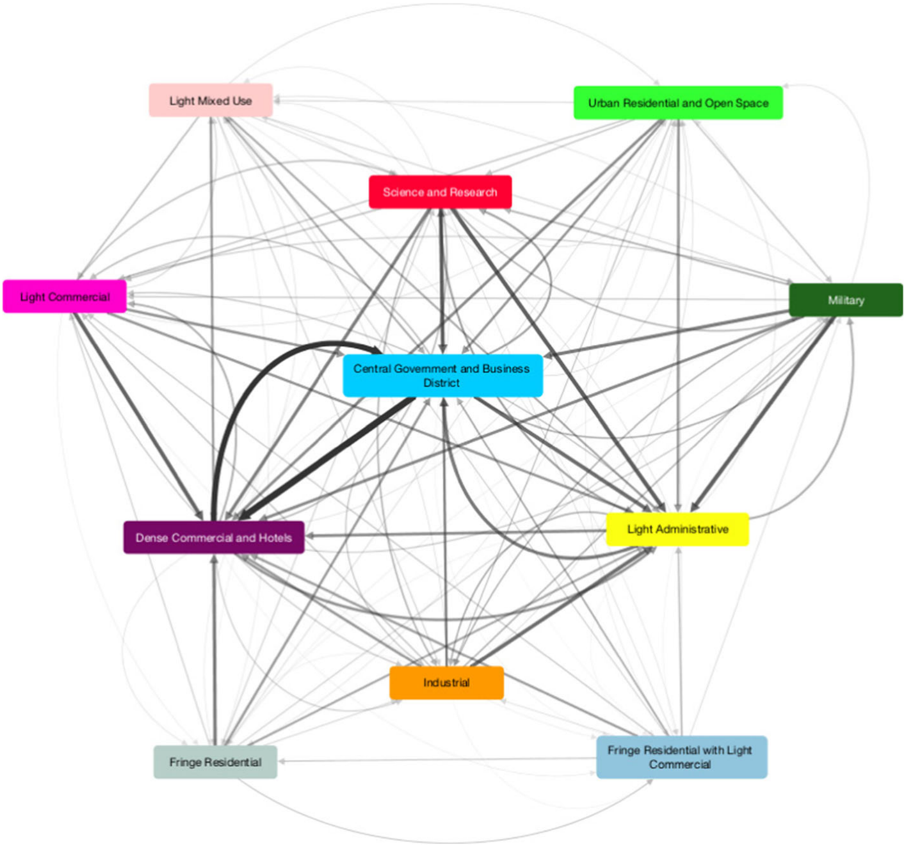
Light Mixed Use
Urban Residential and Open Space
Science and Research
Light Commercial
Military
Central Government and Business District
Dense Commercial and Hotels
Light Administrative
Industrial
Fringe Residential
Fringe Residential with Light Commercial

Fig. 6 Interaction network indicating volume in call traffic flow between land use types

time, potentially highlighting the growing and weakening influence of particular regions during the course of the day.

Measures of activity over time are presented in Fig. 7 – showing call activity normalised by spatial area – and Fig. 8 – presenting call activity normalised by road junctions. The two accounts follow the trends observed above, whereby the Central Government and Business District, Dense Commercial and Hotels, and Light Administrative areas dominating. Again, in line with earlier results, the scale of activity in each region varies with the method adopted. Nevertheless, some important trends emerge from both perspectives.

The most noticeable changes over time are observed in the Central Government, Dense Commercial and Industrial regions. Following a peak in activity at around 12 pm, these regions show sharper reductions in activity after around 6 pm. This is likely reflective of these regions being strongly associated with work activity. Conversely, the remaining areas see general increases in activity from 7 pm onwards, reflecting a return to residential areas. The strongest rise is seen in the Light Administrative area, which rises from around 4 pm, becoming relatively the most active region by road junction normalisation. This is likely due to its mixed-use function and high residential activity. Lower variation in activity is observed in the low activity regions, which show consistent activity from around 11 am to 10 pm.

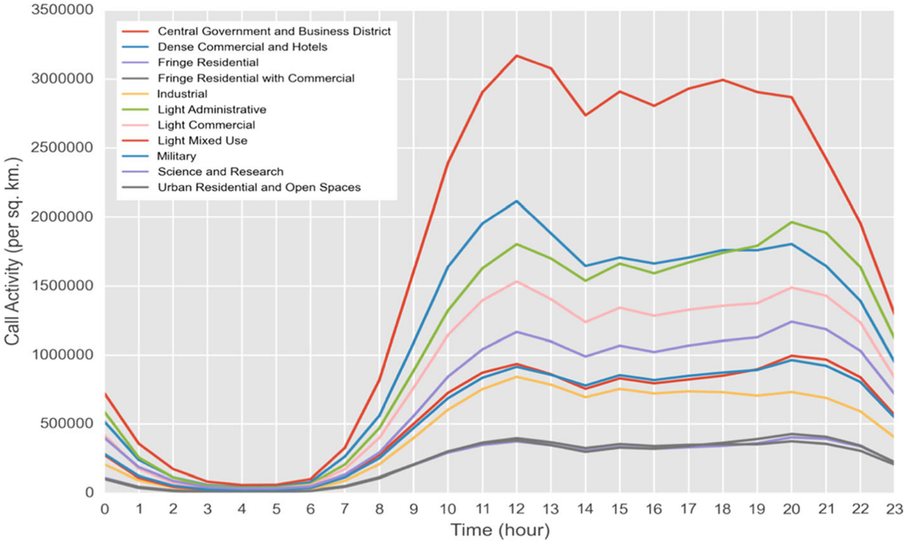

Fig. 7 Variation in call activity (normalised by area) over time by land use type

In view of the above trends, one can explore how interactions vary over time as well. From these data we observe that the relationship between the Central Government and Business and Dense Commercial and Hotels districts dominates throughout most of the day. It is also clear that Light Administrative areas become more attractive as a destination for calls, as well as seeing a growth in activity. This trend is better shown in Fig. 9a (road junction normalisation is used here), where we can observe the changing popularity of the Light Administrative region as a destination for calls over the course of the day. Contrasting this with the Central Government and Business

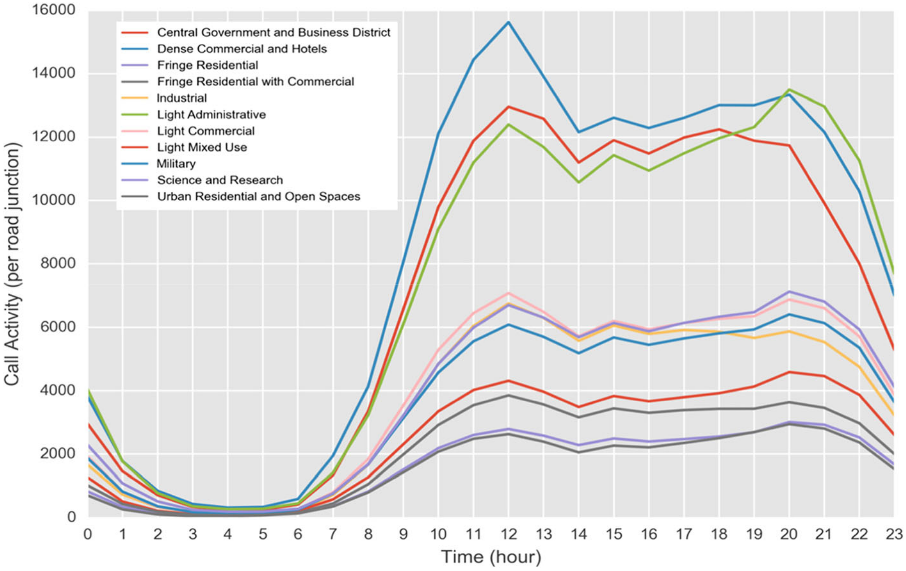

Fig. 8 Variation in call activity (normalised by road junctions) over time by land use type

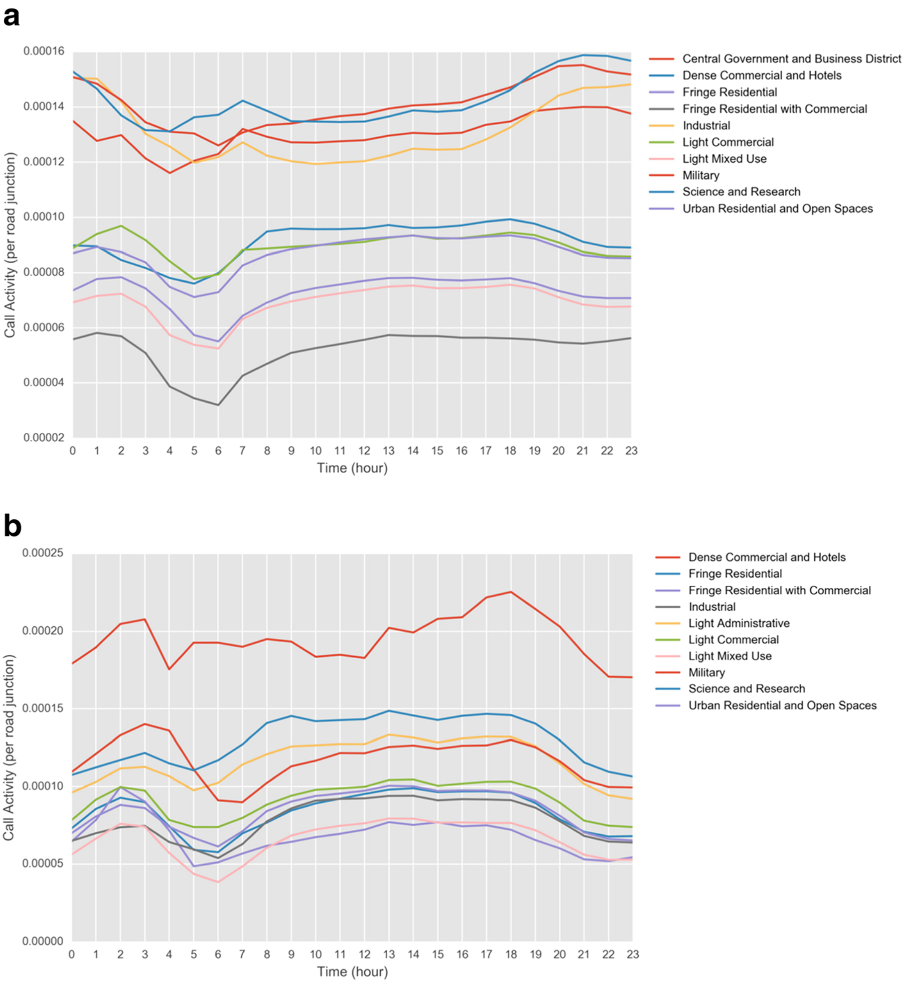

Fig. 9 Call interactions with light administrative (**a**) and the central government and business district (**b**) over time

district, shown in Fig. 9b, we see how there is a relative increase in interaction from business, industrial and military areas from around 4 pm. This is perhaps again indicative of this region's high mixed use function. The Central Government district, conversely, sees a reduction in call activity from 6 pm.

These results add to our understanding of the urban dynamics of Dakar. While the GIS data-derived land use types pointed towards particular types of activity, the analysis of activity and interaction over time add a new dimension to this understanding. While activity and interaction with some areas – such as the Central Government and Business district – aligns with expectations, some – such as the Light Administrative area – demonstrate strong mixed use function (indicative of the presence of some later evening entertainment activities, not indicated through our dataset), and others – such as the Light Mixed Use region – present activity patterns more aligned with expected residential function.

Discussion

At the beginning of this paper, we set out to answer whether, in the absence of census or other demographic data, is it possible to derive useful information about the city, its inhabitants and their activities from data relating to the physical features of the city and the mobile phone usage of its inhabitants. Through a combination of mobile phone transactional data and a fine-grained building use dataset, we have produced analyses of activity and interaction in Dakar, Senegal, and elucidated their variation across land use classification and time. The findings are beneficial in that they provide insight into the behaviour of the city where there was previously little understanding. The provision of these new forms of data have enabled the production of measures of urban activity, without the need to produce conventional forms of population location data through a mechanism such as the census. It is important to recognise that all analyses conducted here have used secondary data only, and have allowed us to achieve reasonably in-depth understanding of activity in Dakar.

On the city itself, the analyses produced an array of useful insights into the city. Fringe residential areas, likely to be newer and potentially more informal, showed clear isolation from other parts of the city with far fewer calls than would be expected being exchanged with other areas. Calls within the area were still relatively high, so this was not purely an artefact of low telephone usage. It was shown how the central government and dense commercial districts drive a disproportionate level of activity and interaction, being areas of call production and reception. The analyses point towards a functionally monocentric structure of the city, directed on this core area of business and governmental activity. Temporal analyses presented dynamic insight into the nature of the city. While areas such as the Central Government and Business district are shown to reduce in activity during the evening hours, others heat up, receiving proportionately more calls during the evening than the day time. These trends align with the classifications of different land use types indicated by building use data. There are additional indications of high levels of intraregional interaction, where areas of the same land use class are observed to exchange higher volumes of calls. Again, this follows our expectations, but adds insight and evidence of the validity of the approach employed here.

The inevitable issue of working within a data poor context is the limited ability to validate the findings derived through this form of secondary data analysis. Yet, through the combination of two distinct datasets within this analysis, a degree of verification of these trends is at least achieved. Across the study, in general, mobile phone call volumes and temporal dynamics of activity in each land-use class align with the expectations we hold from analyses of countless other cities. We observe higher activity in core commercial and administrative areas during the day-time, dropping off as workplaces close in the evening. This reflects the typical activity dynamics of many cities worldwide. Further evidence is added through the strong degree of intraregional interaction between areas of the city classed in the same way. This is a trend one might expect in any city, but provides assurance that the land use definitions are relevant and valid to a degree.

There are natural limitations inherent within an empirical study such as this. Our limited ability to validate findings means we lack strong certainty of the actual activity and interaction dynamics of Dakar. Yet, by analysing the city without additional datasets, we highlight the extent to which these analyses are possible through secondary data alone. The trends we extract, using exploratory, empirical methods, align with our expectations urban activity and interaction. In this sense, the use of mobile phone data, building use data, within the empirical framework outlined, has been positively informative in developing a view of the nature and behaviour of Dakar.

Future work should focus on growing the evidence base for the use of secondary data to explore new geographies and contexts, foregoing the need for conventional methods. There is clear potential for developing countries to leapfrog the evolution of surveying observed in other countries. In this context, mobile phone transaction data, and analyses of the sort presented here, offer significant potential for involvement in future urban planning and policy development. The analyses presented in this paper are merely one approach, and advances on the methods should certainly be explored. We lack residential data within out specification of land use, which means we are unable to be sure of commuting patterns or the cause of activity variation over time. There are unexplored biases within our mobile phone transaction data, and potentially within the GIS dataset too, which require deeper consideration.

References

Aker, J. C., & Mbiti, I. M. (2010). Mobile phones and economic development in Africa. *Journal of Economic Perspectives, 24*(3), 207–232. https://doi.org/10.1257/jep.24.3.207.

Anderson, J.R., Hardy, E., Roach, J.T., Witmer, R. E. (1976). A land use and land cover classification system for use with remote sensor data. U.S. Geological Survey Professional Paper 964, U.S. Geological Survey, Reston, USA.

Arribas-Bel, D. (2014). Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography, The New Urban World, 49*, 45–53. https://doi.org/10.1016/j.apgeog.2013.09.012.

Batty, M. (2013). Big data, smart cities and City planning. *Dialogues in Human Geography, 3*(3), 274–279. https://doi.org/10.1177/2043820613513390.

Blondel, V. D., Decuyper, A., & Krings, G. (2015). A survey of results on mobile phone datasets analysis. *EPJ Data Science, 4*(1), 1.

Borruso, G. (2003). Network density and the delimitation of urban areas. *Transactions in GIS, 7*(2), 177–191.

Breetzke, G. D., & Horn, A. C. (2009). A geodemographic profiler for high offender propensity areas in the City of Tshwane, South Africa. *Environment and Planning A, 41*, 112–127. https://doi.org/10.1068/a40159.

Calabrese, F., Di Lorenzo, G. and Ratti, C. (2010). September. Human mobility prediction based on individual and collective geographical preferences. In *13th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. IEEE Press, New York. 312–317

Calabrese, F., Colonna, M., Lovisolo, P., Parata, D., & Ratti, C. (2011). Real-time urban monitoring using cell phones: A case study in Rome. *IEEE Transactions on Intelligent Transportation Systems, 12*(1), 141–151.

Data for Development. (2013). Analysis of mobile phone datasets for the development of Ivory Coast. Available online: http://perso.uclouvain.be/vincent.blondel/netmob/2013/D4D-book.pdf.

Data for Development. (2015). Data for Development Challenge Senegal. Available online: http://www.netmob.org/www15/assets/img/NetMob%202015_D4D%20Challenge%20Senegal_Sessions_Scientific_Papers.pdf.

de Montjoye, Y.A., Smoreda, Z., Trinquart, R., Ziemlicki, C. and Blondel, V.D., (2014). D4D-Senegal: the second mobile phone data for development challenge. *arXiv preprint arXiv:*1407.4885.

Dennett, A., & Stillwell, J. (2010). Internal migration in Britain, 2000-01, examined through an area classification framework. *Population Space and Place, 16*(6), 517–538. https://doi.org/10.1002/psp.554.

Electronic Notes. (2017). GSM Handover. Available online: https://www.electronics-notes.com/articles/connectivity/2g-gsm/handover-handoff.php.

Everitt, B. S., Landau, S., & Leese, M. (2001). *Cluster analysis* (4th ed.). London: Arnold.

Expert, P., Evans, T. S., Blondel, V. D., & Lambiotte, R. (2011). Uncovering space-independent communities in spatial networks. *Proceedings of the National Academy of Sciences, 108*(19), 7663–7668.

Frias-Martinez, V. and Virseda, J., (2012), March. On the relationship between socio-economic factors and cell phone usage. In *Proceedings of the fifth international conference on information and communication technologies and development*. ACM. New York, pp 76-84

Garbellini, C., Okeleke, K. 2017. The mobile economy sub-Saharan Africa 2017. GSMA Intelligence. Available online: https://www.gsmaintelligence.com/research/2017/07/the-mobile-economy-sub-saharan-africa-2017/631.

Goel, R., Garcia, L.M., Goodman, A., Johnson, R., Aldred, R., Murugesan, M., Brage, S., Bhalla, K. and Woodcock, J., (2018). Estimating city-level travel patterns using street imagery: A case study of using Google Street View in Britain. *PloS one, 13*(5), p.e0196521

Goldsmith, S., & Crawford, S. (2014). *The Responsive City: Engaging communities through data-smart governance*. Hoboken: Wiley.

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A. L. (2008). Understanding individual human mobility patterns. *Nature, 453*(7196), 779–782.

Harris, R., Sleight, P., & Webber, R. (Eds.). (2005). *Geodemographics, GIS and neighbourhood targeting*. Chichister: Wiley.

Herold, M., Scepan, J., & Clarke, K. C. (2002). The use of remote sensing and landscape metrics to describe structures and changes in urban land uses. *Environment and Planning Part A, 34*, 1443–1458. https://doi.org/10.1068/a3496.

Iqbal, M. S., Choudhury, C. F., Wang, P., & González, M. C. (2014). Development of origin–destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies, 40*, 63–74.

Louail, T., Lenormand, M., Cantú, O.G., Picornell, M., Herranz, R., Frias-Martinez, E., Ramasco, J.J. and Barthelemy, M., (2014). From mobile phone data to the spatial structure of cities. *arXiv preprint arXiv:1401.4540*.

Mapbox. (2016). OpenStreetMap Coverage. https://www.mapbox.com/data-platform/country/#senegal.

Openshaw, S. (1984). *The modifiable areal unit problem*. Norwich: Geo Books.

Oshan, T. M. (2016). A primer for working with the Spatial Interaction modeling (SpInt) module in the python spatial analysis library (PySAL). *Region, 1*(3), R11–R23. https://doi.org/10.18335/region.v3i2.175.

Pei, T., Sobolevsky, S., Ratti, C., Shaw, S. L., Li, T., & Zhou, C. (2014). A new insight into land use classification based on aggregated mobile phone data. *International Journal of Geographical Information Science, 28*(9), 1988–2007.

Peng, C., Jin, X., Wong, K. C., Shi, M., & Liò, P. (2012). Collective human mobility pattern from taxi trips in urban area. *PLoS One, 7*(8).

Ratti, C., Frenchman, D., Pulselli, R. M., & Williams, S. (2006). Mobile landscapes: Using location data from cell phones for urban analysis. *Environment and Planning. B, Planning & Design, 33*(5), 727–748.

Ratti, C., Sobolevsky, S., Calabrese, F., Andris, C., Reades, J., Martino, M., Claxton, R., & Strogatz, S. H. (2010). Redrawing the map of Great Britain from a network of human interactions. *PLoS One, 5*(12), e14248.

Reades, J., & Smith, D. A. (2014). Mapping the 'space of flows': The geography of global business telecommunications and employment specialization in the London Mega-City-region. *Regional Studies, 48*(1), 105–126. https://doi.org/10.1080/00343404.2013.856515.

Salesses, P., Schechtner, K., & Hidalgo, C. A. (2013). The collaborative image of the city: Mapping the inequality of urban perception. *PLoS One, 8*(7), e68400. https://doi.org/10.1371/journal.pone.0068400.

Shapiro, I. D. (1959). Urban land use classification. *Land Economics, 35*, 149–155. https://doi.org/10.2307/3144783.

Singleton, A., Wilson, A., & O'Brien, O. (2010). Geodemographics and spatial interaction: An integrated model for higher education. *Journal of Geographical Systems, 14*, 1–19. https://doi.org/10.1007/s10109-010-0141-5.

Song, C., Koren, T., Wang, P., & Barabási, A. L. (2010). Modelling the scaling properties of human mobility. *Nature Physics, 6*(10), 818–823.

Steenbruggen, J., Teresa Borzacchiello, M., Nijkamp, P., & Scholten, H. (2013). 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. https://doi.org/10.1007/s10708-011-9413-y.

Steenbruggen, J., Tranos, E., & Nijkamp, P. (2015). Data from mobile phone operators: A tool for smarter cities? *Telecommunications Policy, 39*(3), 335–346.

Toole, J. L., Ulm, M., González, M. C., & Bauer, D. (2012). Inferring land use from mobile phone activity. In *Proceedings of the ACM SIGKDD international workshop on urban computing*. ACM. Beijing, China. 1–8

Vickers, D., & Rees, P. (2007). Creating the UK National Statistics 2001 output area classification. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 170*, 379–403. https://doi.org/10.1111/j.1467-985X.2007.00466.x.

Wang, D., Pedreschi, D., Song, C., Giannotti, F. and Barabasi, A.L., (2011), August. Human mobility, social ties, and link prediction. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. San Diego, USA. 1100–1108

Wilson, A. (1971). A family of spatial interaction models, and associated developments. *Environment and Planning A, 3*, 1–32.