Estimating modal split using mobile phone location data: A case study of Bamako

Shagun Mittal[a], Satish V. Ukkusuri[a*], Fatima Arroyo Arroyo[b]

[a] Lyles School of Civil Engineering, Purdue University, 550 W Stadium Ave, West Lafayette, IN - 47907, USA
mittal53@purdue.edu, sukkusur@purdue.edu
[b] The World Bank, Washington, DC - 20433, USA
farroyoarroyo@worldbank.org

* Corresponding author

Extended abstract submitted for presentation at the 11th Triennial Symposium on Transportation Analysis conference (TRISTAN XI)
June 19-25, 2022, Mauritius Island

April 4, 22

Keywords: Modal split, K-Means clustering, Data deprivation, Mobile phone location data

1. INTRODUCTION

Transportation analysis typically is performed using surveys especially in the lower income countries which are infrequently conducted. With the boom in use of mobile phones, a new dataset collecting the location of individuals has emerged that can be useful in inferring various transportation aspects. An important transportation feature is the mode of travel. It is selected by people and can assist in comprehending various insights into the characteristics of a region and people’s decisions when traveling from a particular region. This can be used to analyze and plan for various aspects including connectivity of transit services, availability of nearby opportunities, quality of roads, walking infrastructure etc. The statistics on this mode of travel are either very infrequently collected (once in a decade) or reported at a very aggregated scale (overall city). Mobile phone location data collects precise GPS location and timestamp of anonymous users. This enables us to the extract user trips from one disaggregated administrative boundary to other and therefore, analyze the use of modes at such spatial resolutions. A few works have used mobile phone data to infer mode of travel for a trip. But to our knowledge, this has not been performed for a low-income nation where both the availability and quality of datasets is low. Such data sparse conditions require various modifications to the typically implemented algorithms in order to obtain the most accurate results. We conduct this analysis for Bamako, the capital of Mali. This work demonstrates how the algorithms can be modified to perform modal split with satisfactory results. In this extended abstract, we describe our methodology used for assessing modal split, the results obtained and a few policy-based practical applications.

2. METHODOLOGY

We extract the trips using a modified version of methodology described in Batran et al. 2018. The information on trip speed, distance, start and end time, and origin and destination coordinates were extracted. We categorize these extracted trips into walking, transit and driving mode of travel using the following steps.

Data cleaning: Trips with speeds higher than 180 Km/h were considered a nuisance and consequently, filtered out. These constituted a very small fraction (<0.1%) of all extracted trips.

Walking trips: The trips with speed lower than 5 Km/h and with distance of travel shorter than 5 Km were considered as walking trips in compliance with the literature (Batran et al. 2018).

Driving and transit trips: For every OD pair, we categorize the trips in between into driving and transit trips. This is accomplished by initially cleaning the trip data and then performing clustering to classify trips into driving or transit.

Data cleaning: The trips with speed less than 5 Km/h were screened. These trips would generally represent either walking trips or trips where the user location information is collected a while back before the departure from origin or considerably later after arrival at the destination. Moreover, trips with speed greater than 120 Km/h (maximum speed limit in Mali) or twice the free flow speeds were filtered out. Thereafter, we filter out origin-destination (OD) pairs that have only one trip among them, as minimum 2 trips are required to categorize trips into driving and transit.

Classification: Considering all trips between a particular OD pair, we perform K-means (Chen et al. 2016) clustering on speeds to classify transit and driving trips. We observe a high percentage of outliers with driving trips. An outlier is defined as the driving trip speed being both less than 0.4 times the free flow speeds and less than 30.5 Km/h (top 10 percentile transit speed). We perform this classification with multiple thresholds on the minimum number of trips between an OD pair and seek to minimize the overall percent outliers. Figure 1a shows the selection of threshold by minimizing the percent outliers by utilizing heuristic elbow method. Thus, we select a threshold of 3 for the number of trips between an OD pair and screen out accordingly. Figure 1b shows the driving and transit trip classification post-clustering. The driving trip cluster speeds and free flow speeds fit well with a statistically significant slope of 0.83 ($p<0.01$). However, certain outliers still persist. In general, the speeds obtained using mobile phone data are underestimated for both transit and driving.

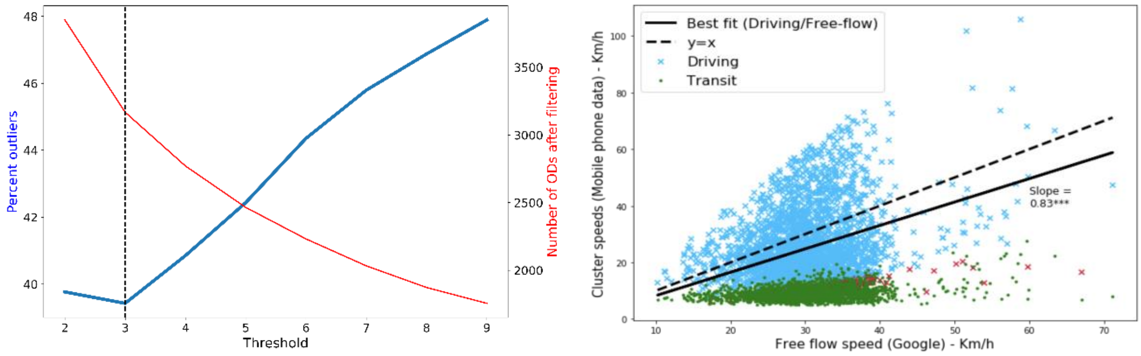

Figure 1: (a) Selecting threshold on number of trips between an OD pair. Following the heuristic elbow method, a threshold of 3 was selected. (b) Driving and transit trips after performing K-means clustering where each marker represents the average driving (or transit) speeds for each OD pair for the classified trips. The best fit line shows that driving cluster speeds, obtained from mobile phone data, are around 0.83 times the free flow speeds. The red colored markers depict an outlier speed for driving as mentioned in earlier. These outliers persist even after the use of the threshold and were eventually eliminated from the data.

3. RESULTS

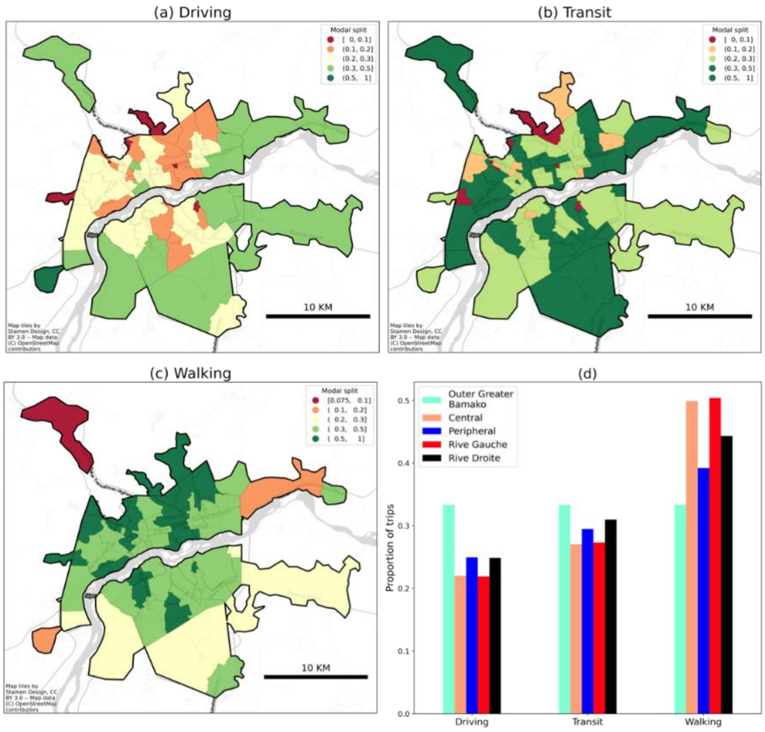

Figure 2: Distribution of proportion of trips generated in Bamako city corresponding to driving, transit and walking modes of travel. Overall the proportion of walking trips are higher in magnitude when compared with transit and driving. The proportion of driving trips is highest in outer Greater Bamako regions.

The distribution of modal split for each quartier located in the city of Bamako is depicted in figure 2. Modal split is the proportion of trips generated from a zone via one particular mode of travel. We observe that from the outer regions (especially in Rive Droite areas), the proportion of driving and transit trips is higher than that of walking trips. However, in almost all other regions, walking tends to be the major mode of travel. Overall the median percentage of walking trips from the regions were 49%, i.e., for the majority of regions the percentage of trips by walk from the region is greater than 49%. The median percentage of driving trips from the regions is 22% while for transit it is 28%. A greater proportion of walking trips implies larger travel times to access facilities. Lower proportion of driving trips can be attributed towards high costs of owning a vehicle; however, lower proportion of transit trips are due to the inadequacy of the transit system in terms of optimally connecting various regions, higher speeds and timely service.

These proportions of modal split are comparable to the results from earlier reports using surveys, where 57% of trips were observed via walking and about 24% via driving. The comparisons are at the overall city scale resolution due to the unavailability of data at a more disaggregate resolution. Table 1 compares the two results.

Table 1: Comparison of modal split using mobile phone data with an earlier report (Olvera et al. 2012)

	Driving	**Transit**	**Walking**
Mobile Phone Data - Median Modal split percentage	22%	28%	49%
Report on Human Settlements (2013) [4] - Modal Split Percentage	24%	17%	57%

4. Discussion

This section presents a few policy based implications of the analysis performed.

Transit is critical for a growing economy like Bamako. It can act as an affordable option for the population unable to buy a vehicle and can be critical to diminish congestion. Understanding the trip generation and attraction trends of transit trips is important to realize the current movement dynamics and to infer regions where improvements are necessary. We observe that during morning peak hours, trips are considerably attracted by regions located in central parts of Rive Gauche area, whereas during evening peak hours, the same regions generate a high number of trips (figure 3). The transit trip generation and attraction distributions during morning and evening peak hours complement one another, indicating work-based trips. Overall, the 25% poorest regions generate and attract low proportions of transit trips in comparison to all zones. On a weekday, about 1.3%

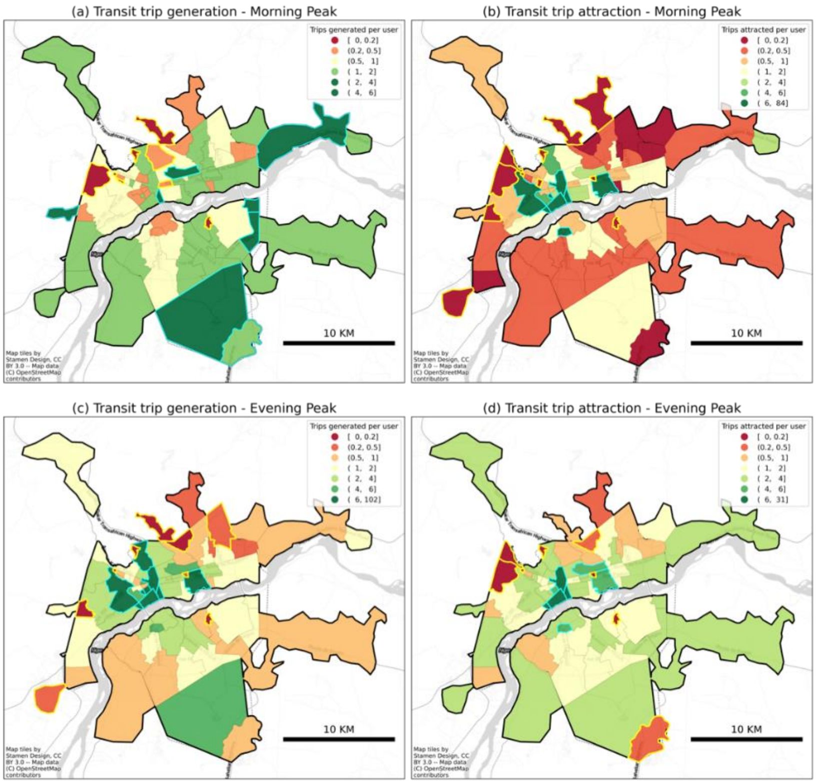

Figure 3: Transit trip generation and attraction for (a-b) morning peak hours (7-10 AM), and (c-d) evening peak hours (4-7 PM). Regions with yellow boundary have least trip generation / attraction while regions with cyan colored boundary have highest trip generation / attraction. Morning and evening peak hours' trends complement each other.

transit trips occur from these poorest to the 25% richest regions while around 0.05% transit trips occur from poorest to poorest regions. While only about 5.5% transit trips are attracted by these poorest regions from the richest ones. Throughout the day, about 18% transit trips occur from rich to rich regions possibly due to more transit stop density in richer zones. People travel from poorer to rich regions to access their jobs but not many people travel from richer to poor regions. Therefore, poorer regions clearly lack in both the job opportunities and the transit facilities. With poorer people not owning vehicles and relying more on transit for longer travels, transit facility improvements are must.

Moreover, we also observe that the poorer regions have higher potential of generating transit trips in comparison to the richer regions. Transit potential is defined as the transit trips generated per user divided by the transit stop density. Intuitively, it is the transit trips generated per transit stop, scaled up to the population of the zone (in comparison to the number of users extracted by the mobile phone data). The figure 4 shows that the transit potential for poorer regions is higher in Bamako. The transit stops located in poorer regions execute a higher number of trips per transit stop when compared to the richer zones. Observing the relation between poverty and transit potential, we highlight a few regions which are extremely poor, have high transit potential and therefore are good candidates for transit improvements (figure 4b). Improving transit in these regions is essential as people from these regions might not be able to afford personal vehicles. Upgrade in transit can enhance the access of these regions leading to their economic growth overall.

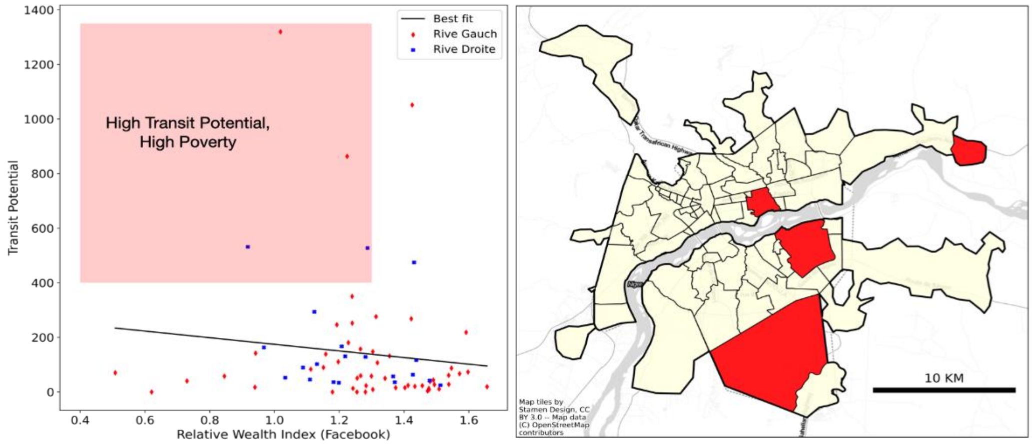

Figure 4: Relation of transit potential and poverty for the Bamako city. Transit potential is higher in regions with higher poverty. The regions with high poverty and high transit potential should be targeted for transit improvement.

In this study, employing passively collected mobile phone data we inferred the modal split of each quarter within Bamako using a modified methodology. The modal split validated well with the previous reported statistics using surveys. From majority of regions, a higher proportion of walking trips is observed when compared to transit and driving, the result urging for improvements in walking infrastructure within the city. We also extended the results to evaluate particularly the transit trips. We observed that very few transit trips are being generated and attracted by the poorer regions. Moreover, the poorer regions have consistently higher trips generated per transit stop when compared to the richer regions, normalizing for the population. These policy-based extensions show the need to invest in transit particularly in the poorer regions. Variations in the representativeness of population exist within the city, however we normalize the extracted trips by the number of users extracted. In addition, concerns like multiple devices carried by a person, and bias in market share have been noted in earlier studies and may induce potential bias. Studies like this are important for understanding the current and monitoring the changes in mobility of people at an extremely disaggregated level.

References

1. Batran, M., Mejia, M.G., Kanasugi, H., Sekimoto, Y. and Shibasaki, R., 2018. Inferencing human spatiotemporal mobility in greater Maputo via mobile phone big data mining. ISPRS International Journal of Geo-Information, 7(7), p.259.
2. Batran, M., Arai, A., Kanasugi, H., Cumbane, S., Grachane, C., Sekimoto, Y. and Shibasaki, R., 2018, July. Urban Travel Time Estimation in Greater Maputo Using Mobile Phone Big Data. In 2018 IEEE 20th Conference on Business Informatics (CBI) (Vol. 2, pp. 122-127). IEEE.

3. Chen, C., Ma, J., Susilo, Y., Liu, Y. and Wang, M., 2016. The promises of big data and small data for travel behavior (aka human mobility) analysis. Transportation research part C: emerging technologies, 68, pp.285-299.
4. Olvera, L., Plat, D. and Pochet. P. (2012). Mobilité et accès à la ville en Afrique Subsaharienne. Godard, X. (2011b). Sustain able urban mobility in "Francophone" Sub-Saharan Africa', regional study prepared for the Global Report on Human Settlements 2013