New Data Sources to Study Airport Competition

Riccardo Gallotti, Marc Fuster and José J. Ramasco
Instituto de Física Interdisciplinar y Sistemas Complejos IFISC (CSIC-UIB)
07122 Palma de Mallorca, Spain
Emails: riccardo@ifisc.uib-csic.es,
jramasco@ifisc.uib-csic.es

***Abstract*—Traditionally, there is a lack of detailed information on passengers' movements from and to the airports. This is due to the limitations in accuracy and coverage of methodologies like local surveys commonly used to obtain data in this context. As a consequence, managers and policy makers must take decisions based on partial information on passengers' transport demands. Recent developments and popularization of the use of Information and Communication Technologies (ICT) provide new alternative data-sources allowing for the precise derivation of individual mobility at different spatial scales. This data may pose some challenges in terms of correcting potential biases, but it overcomes many of the traditional methods limitations. Here, we investigate how the availability of ICT data depicts a new comprehensive perspective on door-to-door air transport mobility. We do this by proposing three case studies involving three new sources of data: i) GPS records of taxi pickups; ii) a database of geolocated tweets including 10 million users tracked for two years in Europe; and iii) the travel-times between the user's home and the alternative airports (provided by Google's API). By integrating this data into simplified discrete choice models, we exemplify how the description of airport catchment areas can be treated in large cities served by more than one airport. This works illustrates how the air transportation system interacts with other transport modes in the passengers decision process. While passengers can still be described within the classical rational choice paradigm, new models must be developed to include the influence of ground transportation aspects in the passenger's travel decisions.**

***Keywords*—data; data analytics; geo-located tweets, travel-times, passenger behaviour; door-to-door mobility**

I. Introduction

The increasing availability of data offered by the explosion of the Information and Communications Technologies, together with the raise in computational power and methodological tools necessary for their elaboration, enables us to study socio-technical systems with unprecedented detail [1]. This possibility propelled a new wave of studies that touched several aspects of human long-range mobility, like seasonal changes in population distribution [2], migration [3], tourism [4], [5], and including air passenger flows [6], [7]. Long-range airline traffic however, interacts with short-range ground transportation. The spreading of epidemics, for instance, is strongly shaped by the interaction of international passenger flows and urban commuting [8]. For this reason, in the effort of expanding our knowledge on the behavior of air passengers, it becomes important to integrate the currently used models with a better understanding of the impact of ground transportation. Also in this case, the recent years have shown a lot of novel data-informed results based on the analysis of cars [9], [10] or taxis [11] GPS traces, mobile phones [12], micro-blogging [13], location based social networks [14], and public transportation timetables [15], [16]. Data by itself is not sufficient and must be supported by adequate methodological foundations to correctly improve our understanding on the evolution of any system and our possibility of forecasting it [17]. In particular, the economical dimension of transportation must be taken into account [18], [19], [20], [21].

In this paper, we propose three case studies to showcase the potentiality of publicly available data sources in describing the effects of the competitions between airports serving the same urban area. We will use the GPS record of taxi pickups in New York City (NYC), geo-referenced tweets in London and Paris, and the trajectories suggested by the Google Maps API [22] for reaching the airports in the same two cities. Our analysis aims at highlighting the role of ground transportation in the choice between alternative airports. In support of this data-driven perspective, we use rational choice theory [23] to model the decision behavior and to point out which aspects are more relevant for the travellers when they face the option of more departure airports.

II. Results

A. Taxi Pickups

As a consequence of the Freedom of Information Law, in 2013 the New York City Taxi and Limousine Commission shared the content of its database to anyone who requested and agreed to physically go to copy the data available in their facilities [24]. For this case study, we use in particular the dataset released by the University of Illinois at Urbana-Champaign [25], but we remark that since then the New York City Taxi and Limousine Commission simplified the access to this type of data which can be now downloaded directly from their website [26]. We limited our analysis to data of the year 2013, which includes over 173 million taxi trips across all the New York state. For each trip, the pickup and dropoff coordinates, and timestamps are recorded, together with the fare charged including tips, tolls, taxes and the surcharges characteristics of the trips to and from the airports. More details can be found in [25].

NYC is served by three airports: John F. Kennedy International Airport (JFK), LaGuardia Airport (LGA) and Newark

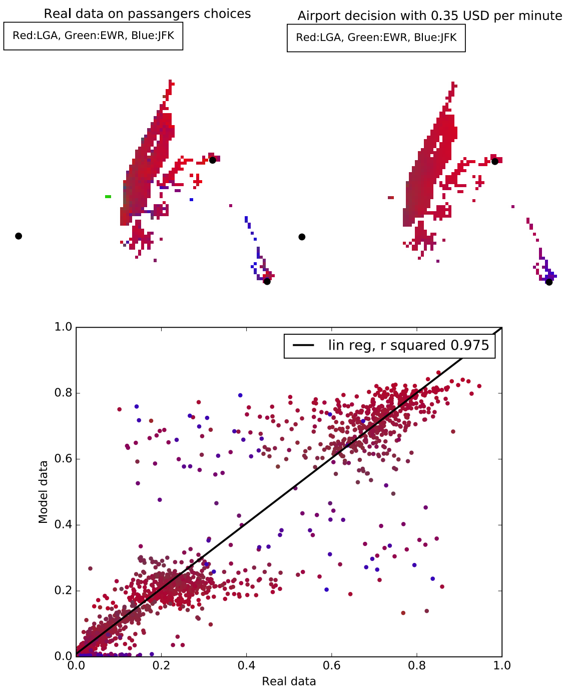

Figure 1. Real (up left) and estimated (up right) fraction of travellers going toward one of the three NYC airports, represented with an RGB scale (Red: La Guardia LGA, Blue: JFK, Green: Newark NWA) for cells with more than 3 journeys. The majority choice is, therefore, the dominant color. Below, a comparison between model and data through a scatter plot of the fraction of users going from a cell to a each given airport.

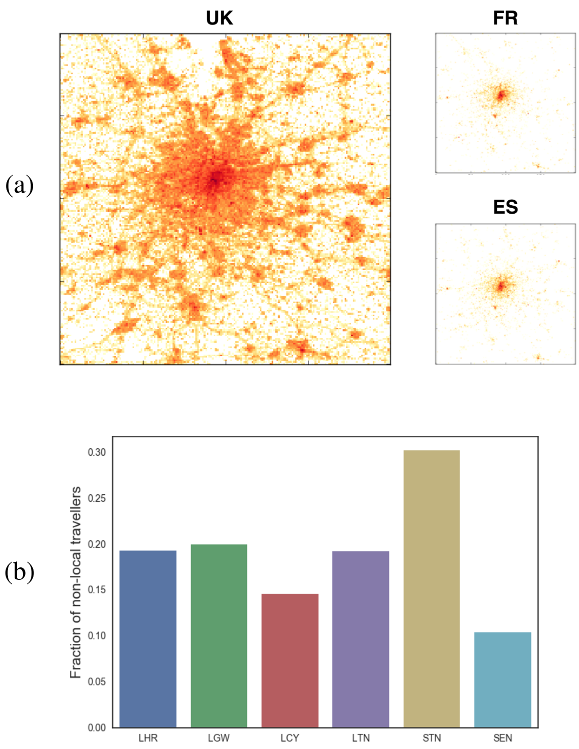

Figure 2. (a) Distribution of tweets of British (UK), French (FR) and Spanish (ES) users around London. (b) Percentage of tourists seen tweeting at each airport.

Liberty International Airport (EWR). JFK is the main international airport and the one with the largest number of passengers (60 million per year), it also has a good public transport connection by train with the island of Manhattan and with Brooklyn. LGA hosts mostly national flights (24 million passengers per year). LGA is also the closest airport to Manhattan but the public transport connection is only by bus. Newark is both national and international, but is has less passengers than JFK (35 million per year). In the data, we identify the trip from and to the airports as if the pickup or dropoff coordinates fall in a box around the airport. Being EWR not in the New York state, only dropoffs are recorded in that airport. For this reasons, the following analysis is limited to dropoffs.

We use here taxi data to model and test the passenger choice selection between airports using a multinomial logit model [23], [27]. We first select an area of analysis comprising the three airports (see Fig. 1 up left) and divide it in bins of approximatively $400m \times 400m$. For each bin i, we evaluate: i) the fraction $F_i(a)$ of pick-ups having as destination one of the airport a; ii) the average travel-time $t_i(a)$ to the airport; and iii) the average cost $c_i(a)$. The travel time and the monetary cost of the taxi travel to the airport allow us to estimate cost associated to the trip as a combination $C_i(a) = c_i(a) + V_T\, t_i(a)$, where V_T is a constant value-of-time. The total utility U associated to a trip should in principle include also the generalized utility gained by performing the trip, and the cost associated to the plane ticket. Here, we simplify of the problem by ignoring these two factors, which are known to be relevant in transport analysis and planning, implicitly inducing the naive assumption that all three airports offer similar flights at similar times, with similar quality and costs of the trips, in order to focus on the effect of ground transportation. Under these assumptions, we can model the probability of choosing the airport a from i as:

$$P_i(a) = \frac{\exp(-C_i(a)/k)}{\sum_i \exp(-C_i(a)/k)}$$

where k is a free parameter representing uncertainty of information [27]. These probabilities can be compared with the observed fractions $F_i(a)$. By minimizing the total error $\sum_{i,a}(P_i(a) - F_i(a))^2$, we identify the optimal values for k and for the value of time $V_T = 0.35$ USD/minute ($R^2 = 0.975$). These value allow us to reproduce the observed catchment areas for taxi users with surprisingly good precision (see Fig. 1 up right and the flow comparison below).

B. Geo-referenced tweets

Geo-located tweets have been continuously recorded by querying the Twitter API [28]. The system we implemented allows us to capture a good part of the entire streaming of geo-located tweets [29]. For this work, we filter only the countries where air traffic is handled by the European Civil Aviation Conference (ECAC). For this selection, we find a total of 9.8

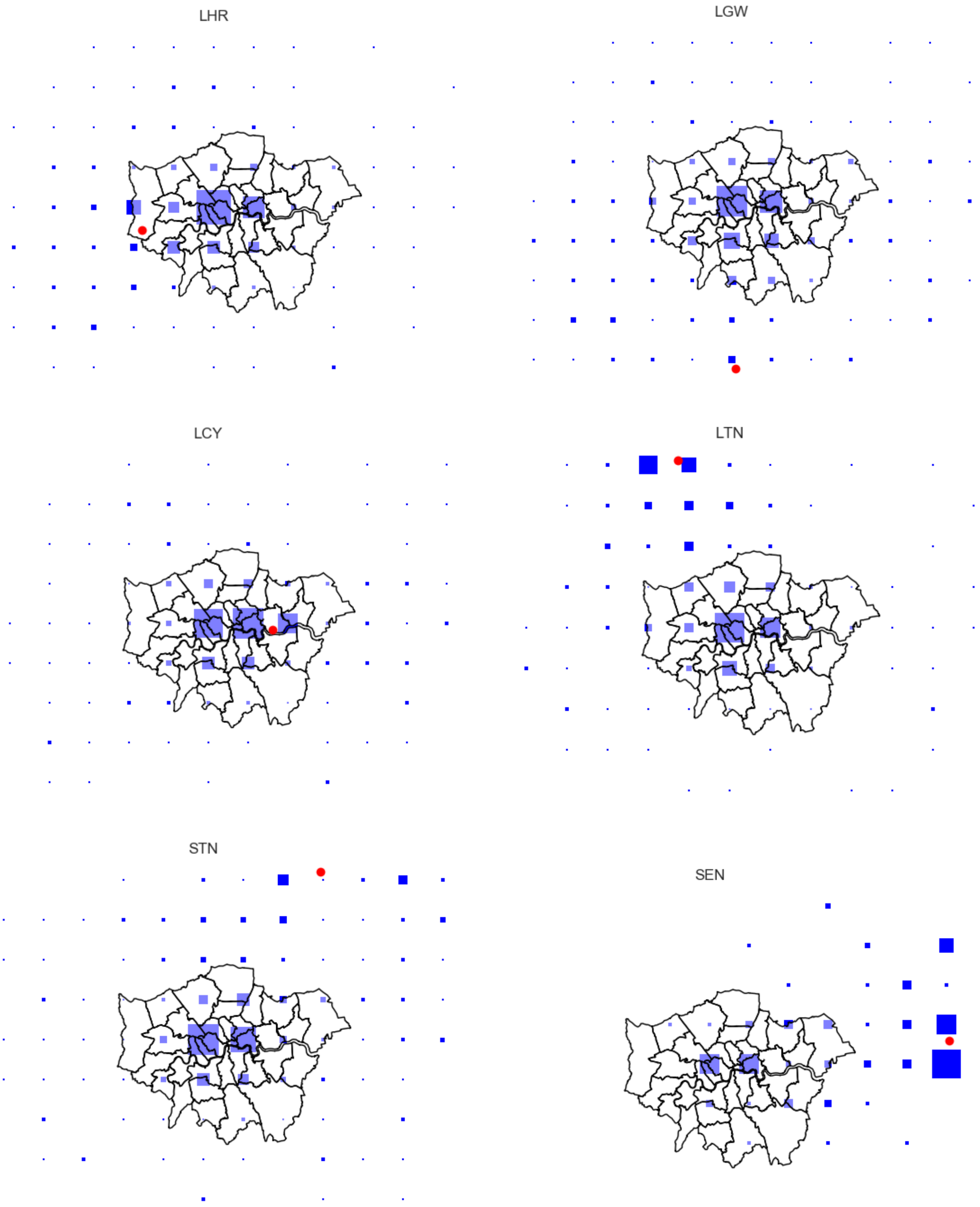

Figure 4. Empirical distribution of the approximate home location for the local passengers observed in the six London airports.

Million users observed during the two-years period of analysis considered (2015-2016).

By tracking the movements of the individual (anonymized) users, it is possible to approximatively reconstruct their home country as the country where they have been observed for the longer time. This allow us to distinguish between tweets of locals and tourists in the same area. In Fig. 2 (a), we can see, for instance, the distribution of tweets of locals (UK) and tourists of two different nationality (France (FR) and Spain (ES)) within the metropolitan area of London.

London is served by six airports: Heathrow (LHR), Gatwick (LGW), City (LCY), Luton (LTN), Stansted (STN), and Southend (SEN). These airports are characterized by a different user base. For studying this difference, we define for each airport a polygon describing its contour (see Fig. 3 (a)) and isolate the tweets performed inside this polygon (see as example the tweets distribution in Heathrow in Fig. 3 (b)). This permits us to assign a set of users to the airport they

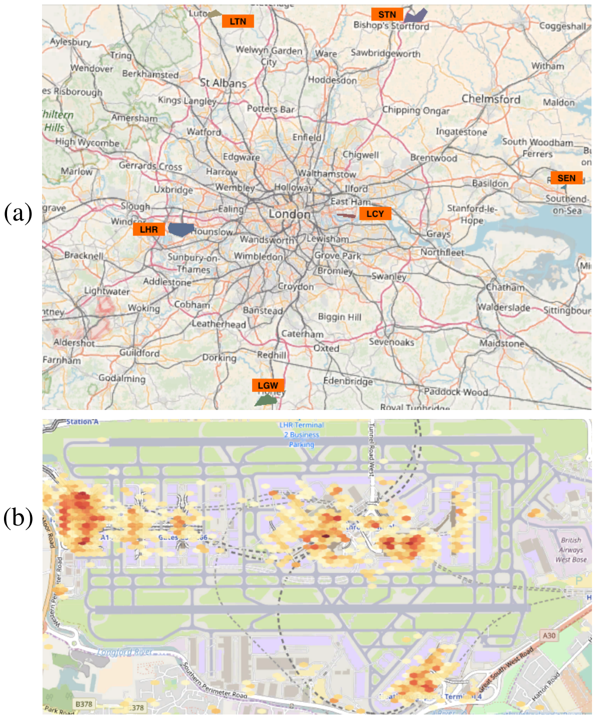

Figure 3. (a) Map of the 6 London airports with the shapes of the polygon used for filtering the tweets. (b) Distribution of tweets observed in the area of the Heathrow airport.

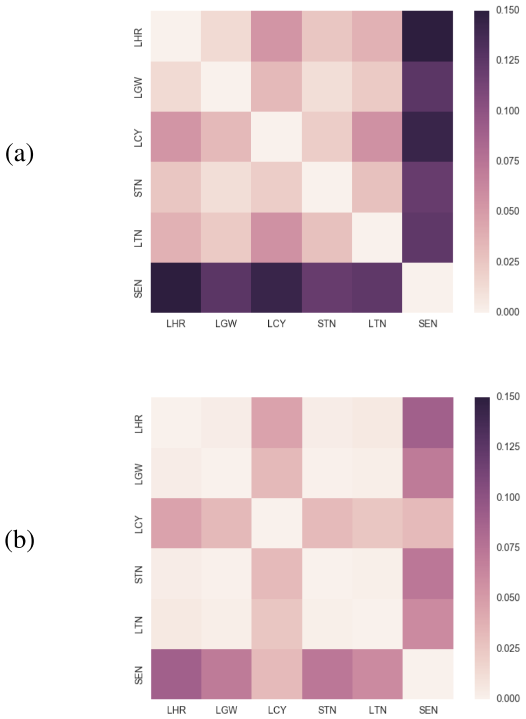

Figure 5. L2 distance between the distribution of home location of local passengers (a) and destination of tourists (b).

tweeted from. It is important to note that we filter out users tweeting frequently in different days from the same airport in order to exclude workers and population living around airports. The first thing we can quantify is what proportion of users observed in the airport are local (residents in the London area) or non-local travelers (having their residence within the rest of the ECAC area) (see Fig. 2 (b)). The airport most used by tourist is Stansted, which is indeed a base for a number of major European low-cost carriers, while the one least used in proportion (and also in total) is the Southend airport.

As one could notice in Fig. 2, locals and tourists are distributed differently in London's metropolitan area, with the tourist mostly concentrated in the central districts. This is reflected also by the subset of users we observed tweeting from within the airports. For each of these users, we can approximate the position of the home-place (for locals) or the final urban destination (for tourists) as the area from where the user tweets the most. In this approximation, we first divide the area of the analysis (Fig. 2) in a number N of sub-areas commensurate to the total number n of tweets recorded (using the rule $N \approx \sqrt{n}$), and identify the home/destination as the center of mass of the points in such sub-area. This procedure has of course some limits, but assures that the approximated location is in an area that has been largely visited, which is not true if one uses the alternative option of computing directly the center of mass of the tweets.

The spatial distribution of the final urban destination of the tourist observed in the different airports show very small differences with the exception of the City and Southend airport from where most often the travelers go to 'The City' of London. A clearer evidence of a spatial optimization in the choice of the airport can be observed in Fig. 4 for local travelers. In these maps, it appears evident that the residence of the travelers observed in an airport are typically closer to that airport than to others. In the choice between the alternative airports, local travelers are therefore minimizing the travel-time (and cost) between home-place and airport. We will build on this observation in the following section, integrating Twitter data with travel-time information extracted from Google maps in two case studies: London and Paris.

The observed difference in the catchment area of the six airport between can be quantified using the L2-distance between the distribution of Fig. 4. The L2-distance is computed as the sum of the square differences between the value of each cell: $D_{L2}(p1, p2) = \sum_{i,j} (p1(i,j) - p2(i,j))^2$, where $p1$ and $p2$ are two probability density distribution, and i and j respectively the row and column index of the cell. In Fig. 5, we represent the L2-distances with a color-scale. In the (a) panel, we can observe the distances for locals and in the (b) panel for tourists. We observe that in both cases the two airports with the most peculiar catchment areas are City and Southend, and that for tourists in the (b) panel all differences are less pronounced than for locals in (a).

We can finally associate to each sector of the area analyzed in Fig. 4 the most common airport used. This permits us to outline the empirical catchment areas of each airport. These are

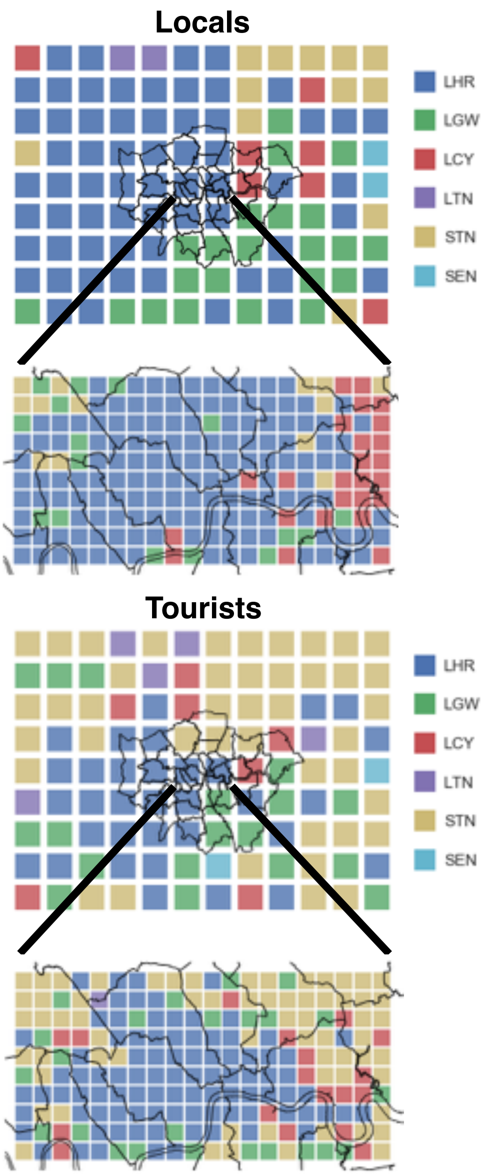

Figure 6. Empirical catchment areas in London for locals (up) and tourists (bottom). The black line represents the London borough. In the top panels each square represents an area of 10x10 km^2. In the bottom panels we zoom in the two most central sections and each square is of 1x1 km^2.

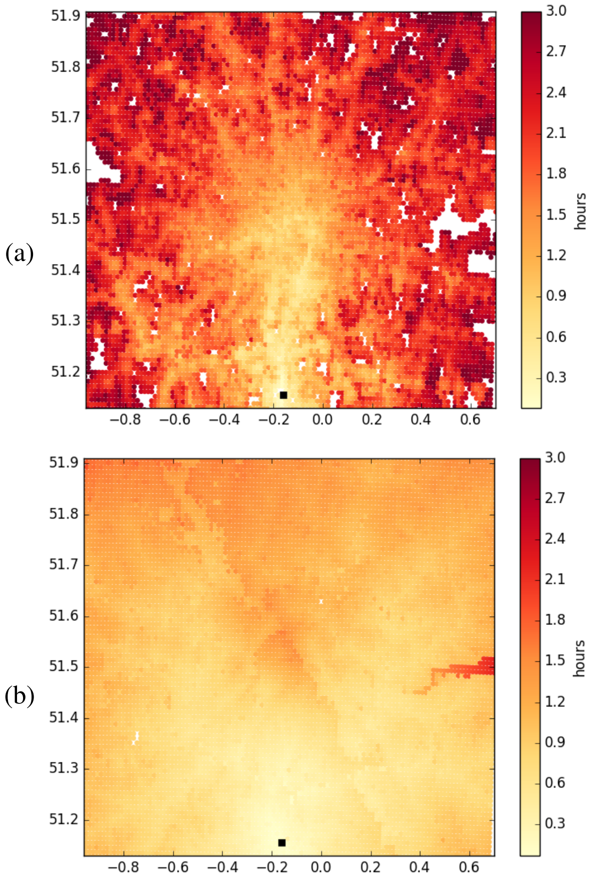

Figure 7. Transit time to Gatwick airport (black square). We remark that for many areas in the centre of London transit (left) is faster than driving (right). In white areas the API does not provide any possible trajectory for reaching the airport from the centre of the box of 1x1 km.

shaped differently for locals and tourists (see Fig. 6). In both cases, the largest and relatively central airport of Heathrow is the dominant option in the center of London. It is noteworthy the effect of low cost companies, which are most used by tourists. Consequently the yellow area representing Stansted is wider for tourist than for local travelers. Conversely, the central and more expensive City airport, in red, is more widely used by locals.

C. Google Maps API

The results illustrated in the preceding section show how the choice between alternative airports is dictated by their accessibility, together naturally with the offer of flights and their cost. In this section, we want to investigate how information on the travel-time of ground transportation acquired through the google maps API [22], integrated with spatial information on the distribution of population in a city (Fig. 2) and ticket costs extracted by the Sabre market intelligence database [30] allows for the modeling the catchment areas of cities with more than one airport such as London and Paris.

We divide the area including the six airports of London (see previous section) and the three airports of Paris: Charles de Gaulle CDG, Orly ORY, and the low cost Beauvais BVA (situated in the far north) in cells of a square kilometer, each identified by a couple of indexes for row and columns i, j. We approximate the real destinations distribution Pop_{ij} with Twitter data by associating each user to the cell where he/she tweets the most, and the travel-times $t_{a,m}(i, j)$ to the airport a and a mode of transport m as the time given by Google Maps for a trip from the cell to the airport at 8am of a Monday (see Fig. 7 for an illustration of trips to London Gatwick). The mode of transport we considered in this analysis are cars ('driving' option in Google API), and public transport ('transit' option in Google API). From the Sabre dataset we obtain the average price of a ticket $c_{a,b}$ between the origin

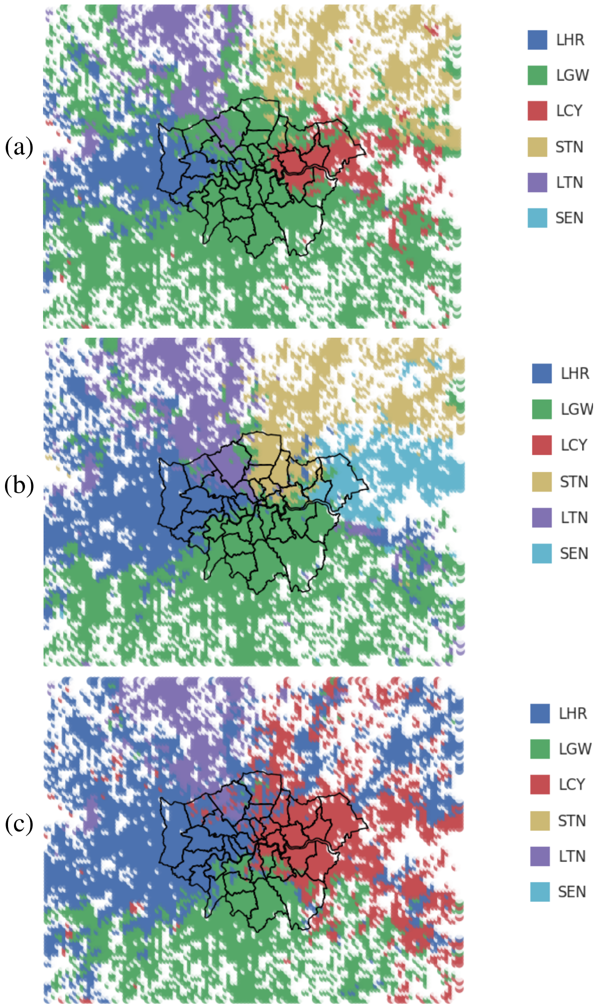

Figure 8. Map of the modeled catchment areas in London for three possible destinations: Madrid (a), Palma de Mallorca (b) and Zurich (c). The black lines represent the London borough.

airport a in London or Paris and a destination b in the year 2014. For this study, we considered a set of $\approx$200 destinations $b \in B(a)$ within the ECAC area and for which more than an origin/destination a was available in the city.

Similarly to what proposed for taxi, we use a multinomial logit to model the decision between alternative airports. We define a generalized cost function as $C_{ij}(a, b, m) = c(a, b) + V_T\, t_{ij}(a, m)$. This cost function would predict that for the travelers departing from cell (i, j) having final destination b and using the mode of transport m for getting to the airport the probability of using airport a is

$$P_{ij}(a; b, m) = \frac{\exp(-C_{ij}(a, b, m)/k)}{\sum_i \exp(-C_{ij}(a, b, m)/k)}$$

where k is a free parameter. From this, we can estimate the fraction of passengers that would choose to go to the airport a for reaching their destination b in the area of analysis as

$$F(a, b, m) = \frac{\sum_{i,j} P_{ij}(a; b, m) Pop_{ij}}{\sum_{i,j} Pop_{ij}}$$

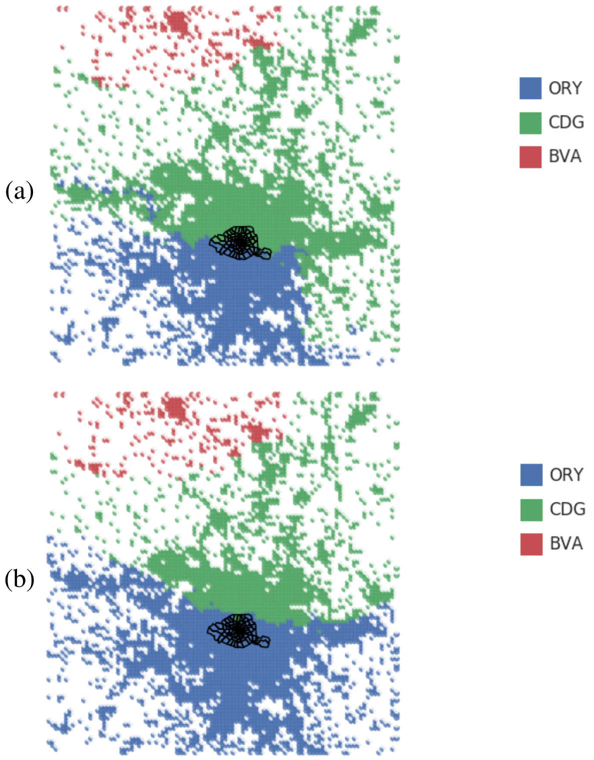

Figure 9. Map of the modelled catchment areas in Paris for two possible destinations: Madrid (a), Palma de Mallorca (b). The black lines represent the Paris arrondissements.

and compare them with the empirical fractions $F^*(a, b)$ that can be extracted from Sabre data, from where we can obtain number of passengers $pass(a, b)$ that have flown from a to b in 2014. This approach is naturally based on a series of simplifying assumptions: i) the cost of ground transportation is totally represented by travel-time, ignoring the monetary cost of the trip to the airport; ii) travel-times are set as constant regarding the days and hours of possible departure; iii) the price of the air ticket is constant in time (notice also that our price data does not include promotions); iv) the choice of the travel destination b and the mode of transport m is independent on where the traveller starts the trip (cell i, j); v) the value of time V_T and the parameter k are constant across the population. vi) we do not consider the option of choosing an alternative destination b (e.g., another airport in the same city) or another mode of transport alternative to the flight.

A last strong assumption we are making at this stage is that we propose here two different reconstruction of the catchment areas for the two alternative means of transport m (cars, or public transport) that can be used for reaching the airport. A more parsimonious way of modeling, for instance, would be for instance to separate the population Pop_{ij} in two sub-groups using different transportation means, but this would require further input information not available at the moment, or alternatively to couple the modeling of the decision between the airports with a second modal-split model describing the decision among the available modes of transport. This alternative modeling option would be in our opinion too refined at this point, since the current model is already based on the aforementioned list of very important assumptions and finds most of his strength in its relatively simple interpretation. Therefore, we compare directly the empirical fraction $F^*(a,b) = pass(a,b)/\sum_a pass(a,b)$ with the theoretical fractions $F(a,b,m)$ obtained from our models, that explicitly depends upon the mode of transport chosen for reaching the airport m.

For each departure city, we obtain the values of V_T (and k) that better approximate the passenger's behavior by minimizing the mean square deviation

$$err(m) = \sum_{a\in A(b), b\in B} (F^*(a,b) - F(a,b,m))^2$$

for a set of destination airports B where the set of possible departure airports $A(b)$ includes more than a single airports.

The key factor to interpret, with this model, the passenger's choice behavior and the differences between cities is the value of time V_T. We show here two case studies describing the London airports reached via public transportation (Fig. 8) and the Paris airports reached via private transportation 9. In both cases, the value of time we found is remarkably high: 150 USD/h for London and 190 USD/h for Paris. This high value suggests us that the more central airports offer probably some further advantages that exceeds simple accessibility, such for example a better choice of flighting time. In reality the collectivity of passenger studies would be better characterized by a distribution of value of times: for some passengers money is more an issue than time, while for others, like for instance business flyers, time is more relevant. This variability, neglected by assumption v) is probably a very important aspect that requires further investigation.

The different availability of departure airports and difference in ticket costs induce a particular outline of the airport catchment areas for every destination b. In Fig. 8 (a) and (b) we can appreciate how, as a consequence of the topology of the underlying transit network and of the radial distribution of the airports, the visual representation of the most frequently used airport seems to cut the center of London "as slices of a cake". More in detail, we see that the area of influence low cost airport of Stansted (purple) for passengers using public transport is expected to expand for trips to the touristic destination of Palma de Mallorca (PMI, panel b) as compared to the state capital of the same country Madrid (MAD). Comparing these with the catchment ares for trips to Zurich (ZRH) displayed in panel (c), we observe how for this in general more expensive destination the role of the central City airport is expected to become more important. In Fig. 8 (a) and (b) we propose two catchment areas for the Paris airports, if reached by car, proposing a similar comparison between an important touristic destination (PMI) and a state capital (MAD) as destination. In this second case, the visual representation of the most frequently used airport represented in Figure 9a shows the two main airports (CDG and ORY) splitting horizontally in half the municipality area of Paris. For the touristic destination PMI (Fig. 9b), the low cost options in ORY expands its area of influence to the whole Paris municipality. The influence of the low-cost airport of Beauvais seems to be very marginal under this perspective. We expect that the same study focused on transit travel-times would have differed from this picture, but we discovered that information on the shuttle service from the center of Paris to the BVA airport is not provided by the Google Maps API, practically reducing the study of the French capital to its main two airports.

III. Conclusion

In this paper, we assessed a series of novel modelling opportunities provided by three new sources of ICT data with application to the description of the catchment areas in large metropolitan areas served by many airports. For these cities, we have shown that the way air transport system interacts with other transport modes can be an important factor in the decision process of air passengers. To include the influence of ground transportation, we modeled the passengers choices with an utility dominated by a cost function that includes the travel-times necessary to reach the airport. Our modeling approach involved series of strong simplifying assumptions in the choice modeling, but it still allows us to illustrate how new types of data can be used to study mobility and travelling behaviour in air transport.

From the three case-studies proposed, we can already reach some preliminary conclusions at a relatively coarse level of granularity. The fact that it has been possible to successfully reconstruct the proportion of taxi passengers going the different NYC airports using only information on the ground transportation, without the need of introducing a different cost associated to each airports, suggests that the advantages and dis-advantages associated to the different airport might be very balanced once one aggregates over all possible destination offered (as it is implicitly done in our study).

Detailed insight on the passenger's behavior is now available thanks to open-access individual data such as geo-located tweets. For instance, we observe a clear difference between locals and tourists behavior: London citizens more often choose the closest airports, while this behavior is less remarkable when tracking tourists. However, the quantity of Twitter records available after selecting the passengers observed within the airport is often insufficient for high resolution statistics. For this reason, in our third case study we relied on the Sabre dataset for validation, while Twitter data were used to reconstruct the final destinations within the city.

As extremely rich spatial data such as those offered by mobile phone records [31] are becoming progressively more accessible, the individual trajectory data can successfully be integrated with other rich and high definition sources of data on ground transportation such as public transport timetables [16], open street map [32], or the Google Maps API [22]. Finally, all these sources of mobility and transportation data can be enhanced thanks to more traditional models such as rational theory. Indeed, as we have shown in our last case-study, even within the clear assumption shortcomings of the chosen model, this approach is useful to identify patterns in the data that can inform on the systems's characteristics. This data assimilation is naturally limited of the assumptions behind the model we chose. For instance, the fact that in our analysis some airports have clearer catchment areas than others is probably generated by the fact that the assumption of similar service offerings is violated: whilst some airports may compete for passengers with a similar service offering, others do not and are unique.

Acknowledgment

The project leading to these results has received funding from the SESAR Joint Undertaking under grant agreement No 699260 under European Union's Horizon 2020 research and innovation programme.

References

[1] A. Vespignani, "Modelling dynamical processes in complex socio-technical systems," *Nature Physics*, vol. 8, p. 32, 2012.

[2] P. Deville, C. Linard, S. Martin, M. Gilbert, F. R. Stevens, A. E. Gaughan, V. D. Blondel, and A. J. Tatem, "Dynamic population mapping using mobile phone data," *Proceedings of the National Academy of Sciences USA*, vol. 111, pp. 15 888–15 893, 2014.

[3] F. Simini, M. C. González, A. Maritan, and A.-L. Barabási, "A universal model for mobility and migration patterns," *Nature*, vol. 484, p. 5, 2012.

[4] M. Lenormand, B. Gonçalves, A. Tugores, and J. J. Ramasco, "Human diffusion and city influence," *Journal of The Royal Society Interface*, vol. 12, p. 20150473, 2015.

[5] A. Bassolas, M. Lenormand, A. Tugores, B. Gonçalves, and J. J. Ramasco, "Touristic site attractiveness seen through twitter," *EPJ Data Science*, vol. 5, p. 12, 2016.

[6] B. Hawelka, I. Sitko, E. Beinat, S. Sobolevsky, P. Kazakopoulos, and C. Ratti, "Geo-located twitter as proxy for global mobility patterns," *Cartography and Geographic Information Science*, vol. 41, pp. 260–271, 2014.

[7] M. G. Beiró, A. Panisson, M. Tizzoni, and C. Cattuto, "Predicting human mobility through the assimilation of social media traces into mobility models," *EPJ Data Science*, vol. 5, p. 30, 2016.

[8] D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J. J. Ramasco, and A. Vespignani, "Multiscale mobility networks and the spatial spreading of infectious diseases," *Proceedings of the National Academy of Sciences USA*, vol. 106, pp. 21 484–21 489, 2009.

[9] R. Gallotti, A. Bazzani, and S. Rambaldi, "Towards a statistical physics of human mobility," *Int. J. Mod. Phys. C*, vol. 23, p. 1250061, 2012.

[10] R. Gallotti, A. Bazzani, S. Rambaldi, and M. Barthelemy, "A stochastic model of randomly accelerated walkers for human mobility," *Nature Communications*, vol. 7, p. 12600, 2016.

[11] J. Yuan, Y. Zheng, and X. Xie, "Discovering regions of different functions in a city using human mobility and pois," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012, pp. 186–194.

[12] M. S. Iqbal, C. F. Choudhury, P. Wang, and M. C. González, "Development of origin–destination matrices using mobile phone call data," *Transportation Research Part C: Emerging Technologies*, vol. 40, pp. 63–74, 2014.

[13] M. Lenormand, M. Picornell, O. G. Cantú-Ros, A. Tugores, T. Louail, R. Herranz, M. Barthelemy, E. Frias-Martinez, and J. J. Ramasco, "Cross-checking different sources of mobility information," *PLoS ONE*, vol. 9, p. e105184, 2014.

[14] A. Noulas, S. Scellato, R. Lambiotte, M. Pontil, and C. Mascolo, "A tale of many cities: universal patterns in human urban mobility," *PloS ONE*, vol. 7, p. e37027, 2012.

[15] R. Gallotti and M. Barthelemy, "Anatomy and efficiency of urban multimodal mobility," *Scientific Reports*, vol. 4, 2014.

[16] ——, "The multilayer temporal network of public transport in great britain," *Scientific Data*, vol. 2, 2015.

[17] H. Hosni and A. Vulpiani, "Forecasting in light of big data," *Philosophy & Technology*, pp. 1–13, 2017.

[18] M. Lenormand, T. Louail, O. G. Cantú-Ros, M. Picornell, R. Herranz, J. M. Arias, M. Barthelemy, M. San Miguel, and J. J. Ramasco, "Influence of sociodemographics on human mobility," *Scientific Reports*, vol. 5, 2015.

[19] L. Lotero, A. Cardillo, R. Hurtado, and J. Gómez-Gardeñes, "Several multiplexes in the same city: the role of socioeconomic differences in urban mobility," in *Interconnected Networks*. Springer, 2016, pp. 149–164.

[20] L. Lotero, R. G. Hurtado, L. M. Floría, and J. Gómez-Gardeñes, "Rich do not rise early: spatio-temporal patterns in the mobility networks of different socio-economic classes," *Royal Society Open Science*, vol. 3, p. 150654, 2016.

[21] M. A. Florez, S. Jiang, R. Li, C. H. Mojica, R. A. Rios, and M. C. González, "Measuring the impacts of economic well being in commuting networks-a case study of bogota, colombia," 2017.

[22] "Google maps API," https://developers.google.com/maps/, accessed: 2017-09-14.

[23] D. MCFADDEN, "Conditional logit analysis of qualitative choice behavior," *Frontiers in Econometrics*, pp. 105–142, 1974.

[24] "Chris whong blog," http://chriswhong.com/open-data/foil_nyc_taxi/, accessed: 2017-09-04.

[25] B. Donovan and D. Work, "New york city taxi trip data (2010-2013) ," https://doi.org/10.13012/J8PN93H8, 2016.

[26] "NYC taxi & limousine commission," http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml, accessed: 2017-09-04.

[27] D. Helbing, *Quantitative sociodynamics: stochastic methods and models of social interaction processes*. Springer Science & Business Media, 2010.

[28] "Twitter API," https://dev.twitter.com/overview/api, accessed: 2017-09-20.

[29] A. Tugores and P. Colet, "Mining online social networks with python to study urban mobility," *arXiv preprint arXiv:1404.6966*, 2014.

[30] "Sabre market intelligence tool," http://www.sabre.com, accessed: 2015.

[31] P. García, J. J. Ramasco, G. Andrienko, N. Adler, C. Ciruelos, and R. Herranz, "Big data analytics for a passenger-centric atm system: A case study of door-to-door intermodal passenger journey inferred from mobile phone data," in *Proceedings of the SESAR Innovation Days 2016*. EUROCONTROL, 2016.

[32] "Open street map API," http://wiki.openstreetmap.org/wiki/API, accessed: 2017-09-22.