Big Data Analytics for a Passenger-Centric Air Traffic Management System

A Case Study of Door-to-Door Intermodal Passenger Journey Inferred from Mobile Phone Data

Pedro García, Ricardo Herranz
Nommon Solutions and Technologies
Madrid, Spain

José Javier Ramasco
IFISC (CSIC-UIB)
Palma de Mallorca, Spain

Gennady Andrienko
Fraunhofer IAIS
Sankt Augustin, Germany

Nicole Adler
Hebrew University of Jerusalem
Jerusalem, Israel

Carla Ciruelos
ISDEFE
Madrid, Spain

***Abstract*—BigData4ATM is a SESAR 2020 Exploratory Research project that investigates how new sources of passenger-centric data coming from smart personal devices can be analysed to extract relevant information about passengers' behaviour. In this paper, we introduce the project and present a case study focused on the analysis of door-to-door passenger journeys from mobile phone data. Anonymised call detail records (CDRs) for several million users in Spain are used to infer door-to-door trips for the Madrid-Barcelona corridor and identify the long-distance transportation mode (air, rail, road) chosen by each user. These door-to-door itineraries are upscaled to the total population using demographic data and used to estimate modal split, airport catchment areas and door-to-door travel times. Estimated modal shares are compared to official statistics to validate the results. We finish by outlining future research directions and discussing how the information extracted from mobile phone records can be exploited to inform decision making in air transport and ATM.**

Keywords-big data; data analytics; mobile phone records, passenger behaviour; door-to-door mobility

I. INTRODUCTION

The long-term vision for the European aviation sector outlined in the report 'Flightpath 2050 - Europe's Vision for Aviation' [1] identifies five challenges that aviation will have to face at the 2050 horizon: Meeting Societal and Market Needs; Maintaining and Extending Industrial Leadership; Protecting the Environment and the Energy Supply; Ensuring Safety and Security; and Prioritising Research, Testing Capabilities and Education. The report envisages a passenger-centric air transport system thoroughly integrated with other transport modes, with the ultimate goal of taking travellers and their baggage from door to door predictably and efficiently while enhancing passenger experience and rendering the transport system more resilient against disruptive events, drawing attention to the key role of the ATM system in realising this vision.

In contrast with this high-level vision, ATM operations have so far lacked a passenger-oriented perspective, with performance objectives and decision criteria (e.g., flight prioritisation rules) not necessarily taking into account the ultimate consequences for the passenger [2]. To overcome this problem, further research is needed on the interactions between the ATM system and passengers' needs, choices and behaviour. However, current methods to collect data on passengers' activities are limited in accuracy and validity. Traditional methods based on observations and surveys are important national planning instruments, but they present intrinsic limitations (e.g., incorrect and imprecise answers, dependence on the availability and willingness to answer of the interviewed persons, etc.), and they are also expensive and time-consuming. Useful data can also be collected from other sources such as air traffic databases, travel reservation systems and market intelligence data services, but these data typically fail to capture important information, such as door-to-door origin-destination pairs and travel times.

The generalised use of geolocated devices in our daily activities opens new opportunities to collect rich data and overcome many of the limitations of traditional methods. The very same tools that are enabling new forms of bidirectional communication with the passenger are making it possible to gather permanently updated information on passengers' activity-mobility patterns with an unprecedented level of detail. BigData4ATM (www.bigdata4atm-sesar.eu) investigates how to exploit these new opportunities to inform air transport and ATM decision making processes. In this paper, we present a case study that illustrates such opportunities. Section II provides an overview of these new data and depicts the objectives of BigData4ATM. Section III reviews the state-of-the-art in the analysis of travel behaviour from mobile phone data and discusses the opportunities and challenges for using these data in the context of air transport and ATM. Section IV presents the case study, which explores how anonymised mobile phone records can be analysed to study door-to-door passenger journeys. Section V describes the approach that has been followed and the methodology and algorithms used to analyse the mobile phone data. Section VI summarises the main results of the case study. Section VII concludes and outlines future research directions.

II. New Big Data Sources for Socioeconomic and Behavioural Research in ATM

A. *Passenger-Centric Data for Travel Behaviour Analysis*

Along the last decade, numerous studies have addressed the analysis of human mobility and travel behaviour based on geolocated data coming from mobile personal devices, including mobile phone records ([3], [4]), Twitter data [5], credit card records [6] and public transport smart card data [7], among other examples. Additionally, some studies have addressed the validation of the mobility statistics obtained from these new, non-conventional data by comparing them with those obtained from travel surveys [8], evidencing the potential of these new big data sources to complement or even replace traditional data collection methods.

B. *The BigData4ATM Project*

With society being one of the stakeholders of the ATM Performance Partnership [9], air transport users' requirements need to be accommodated to enable the execution of flights as close as possible to the passengers' intention. Planning and managing the future ATM system will thus require systematic socioeconomic monitoring to understand the factors influencing changes in demand and flights patterns [10]. A variety of studies can be found on factors influencing future air traffic on the economic dimension. However, the societal dimension, essential in the paradigm shift from a flight-centric to a passenger-centric air transport system, has been little explored [11]. A comprehensive assessment of the impact of social and behavioural factors (behavioural attitudes, mobility needs, tolerance to waiting times, transport mode preferences, etc.), as well as of the net impact that SESAR and other ongoing ATM modernisation programmes will have on the passenger, remains to be done.

BigData4ATM aims to fill this gap by taking advantage of the opportunities opened by new sources of passenger-centric geolocated data. The project pursues the following objectives:

1) To develop a set of methodologies and algorithms to acquire, integrate and analyse multiple distributed sources of non-conventional ICT-based spatio-temporal data — including mobile phone records, data from geolocation technologies, credit card records and data from Internet social networks, among others — with the aim of eliciting passengers' behavioural patterns.

2) To develop new theoretical models translating these behavioural patterns into relevant and actionable indicators for the planning and management of the ATM system.

3) To evaluate the potential applications of the new data sources, data analytics techniques and theoretical models through a number of case studies relevant for the European ATM system.

A detailed description of the project can be found in [12].

III. Inferring Door-to-Door Intermodal Passenger Journey from Mobile Phone Data

A. *Analysis of Human Mobility from Mobile Phone Data: State-of-the-Art*

The pervasive penetration of mobile phones allows the collection of a vast amount of spatio-temporal data about human activity, which is attracting the interest of a growing community of researchers. In particular, a significant amount of research has been conducted using the so-called call detail records (CDRs). A CDR is a data record produced every time a mobile phone interacts with the network through a voice call, a text message or an Internet data connection. Among other information, each record contains an identifier of the user, a timestamp and a location. Location is in most cases given by the position of the cell phone tower to which the user is connected, but it can sometimes consist in the geographical coordinates of the user estimated through triangulation. Temporal resolution depends on the frequency with which the mobile phone interacts with the network. A few years ago, when CDRs mainly included voice calls and text messages, the temporal resolution for most users was relatively low. This situation has changed thanks to the ubiquity of smartphones: with apps connecting to the network without user interaction and the way data sessions are managed by the mobile phone or the network operator, if the mobile phone has a data session opened it will typically produce a register every hour/half an hour. Therefore, thanks to smartphones the temporal granularity of CDRs is significantly higher.

Although CDRs are originally collected for billing purposes, the digital footprints they contain provide useful insights into population movement. When compared to traditional travel surveys, mobile phone records provide important advantages:

- Large samples, typically several orders of magnitude higher than the samples of traditional surveys.
- Data is collected passively, eliminating the interaction with the interviewed person, which often leads to incorrect or imprecise answers.
- Cost savings, as data are already being recorded by mobile networks operators.
- Information can be permanently updated, since data collection and analysis can be performed in days.

On the other hand, CDRs also have a number of limitations:

- Noise and false positions: when working at cell-based level, the assumption is typically that the mobile phone connects to the closest tower, which is not always true, leading to inaccuracies in the positioning of the users.
- Spatial resolution: unless triangulated position data are available, spatial resolution depends on cell size. This resolution is relatively high in dense urban areas (a few

hundred meters), but drops in rural/less populated areas (up to several kilometers).

- Geographical scope: as CDRs are obtained from a mobile network operator, they are limited to the area (usually one country) where the operator has presence.
- Sociodemographic information about the user is limited to the information available to the mobile network operator, which is less rich than the information typically available from travel surveys.

These limitations need to be carefully taken into account when using CDRs to analyse human mobility, making it necessary to develop sophisticated data analysis and data fusion algorithms able to translate raw mobile phone data into relevant activity-travel information. It has been demonstrated that the analysis of CDRs allows the identification of user's home and workplace/place of study in a very reliable manner. Other common locations, as well as less frequent activities, can also be detected, allowing the extraction of activity-travel diaries that are consistent with those obtained from household travel surveys [13]. The resulting origin-destination matrices are in many aspects (particularly regarding trip distribution) superior to those obtained from surveys, and are already being successfully used in transport planning projects [14]. However, significant work remains to be done to generate activity-travel diaries that include transportation mode, route, and travel purpose. The fusion with land use information, combined with the improvement of location accuracy through triangulated data, is a promising avenue for the determination of activity types other than home and work [13]. There are also examples in the literature where trip mode is inferred from CDRs, typically based on trip time [15], registers produced during the trip [16], or a combination thereof, but the spatial and temporal resolution typically provided by CDRs restricts the use of these techniques to long-distance trips. Another approach, arguably more promising for determining transport mode in urban areas, is to combine CDRs with other complementary data, such as public transport smart card data [17].

B. *Applications to Air Transport and ATM: Analysis of Door-to-Door Mobility*

When it comes to air transport and ATM, one of the main advantages of mobile phone data with respect to traditional air transport databases is the ability to provide information about the whole door-to-kerb and kerb-to-door trip segments. These segments are of paramount importance for the achievement of the European 4-hour door-to-door target, according to which, by 2050, 90% of travellers within Europe should be able to complete their journey, door-to-door within 4 hours [1]. The connection of airports with other transport modes is also an important factor for addressing other important questions, such as air traffic forecasting along certain routes, the evaluation of the resilience of the air transport system against disruptive events, and the understanding of the role of ATM in the context of the intermodal transport system. However, door-to-kerb and kerb-to-door trip segments are often the most difficult segments to characterise, not only due to the existence of multiple airport access/egress modes, but also because passengers' origin/destination can be virtually anywhere. Either alone or in combination with other data sources, mobile phone records offer a great potential to characterise the origin/destination of passengers within airport catchment areas, measure travel times to/from airport for different access/egress modes, and ultimately evaluate the level of achievement of the 4-hour door-to-door target.

IV. Case Study: Modal Split and Airport Catchment Area in the Madrid-Barcelona Corridor

To exemplify the potential of mobile phone data, we analyse the competition between transport modes in the Madrid-Barcelona corridor. For this preliminary study, we have chosen to focus on one specific day, Tuesday 8th March 2016, as an example of a typical working day.

A. *The Madrid-Barcelona Corridor*

Madrid and Barcelona are the two largest cities in Spain, and they form the main passenger transport corridor in the country. They are separated by a distance of around 600 km, with different alternatives connecting the two cities.

The air route connecting Madrid and Barcelona is one of the most used flight routes in Europe. There are more than 20 flights per day, starting from 6:45 AM, with a frequency of 15 minutes during peak hours. The flight duration is less than 90 minutes. Before 2008, the rail connection was very poor, with a duration higher than 5 hours, and was no real competition for the air route. In 2008 a high-speed train (HST) line was inaugurated that connects Madrid and Barcelona in less than 3h, and started to attract passengers from air transport. Currently HST has a market share of around 62% compared to air transport [18]. There are around 26 direct HST connections per day, starting from 5:50 until 21:25.

In terms of competition between modes, the aircraft and the HST compete for the same type of passengers: business passengers and other passengers for which the value of time is high. Although flight duration is half the HST trip duration, the airport processes (check-in, security, etc.), at the moment not required for HST, reduce this difference significantly. In terms of accessibility to airports and train stations, the HST stations (Madrid-Atocha and Barcelona-Sants) are much more central than the airports, so the access and egress legs are likely to influence modal choice. Recent studies have concluded that the relative accessibility levels to terminals play an important role in the overall competitiveness of each transport modes, and have shown that access by private car favours the relative competitiveness of air transport, while public transport favours HST competitiveness [19].

Finally, the road connection connects the two cities in 6 to 9 hours, depending on the mode (bus trips are typically around 8h long) and the number of intermediate stops.

B. Objectives of the Case Study

The case study explores the potential of mobile phone data to obtain the following information: (1) modal split; (2) airport and HST catchment areas; (3) total door-to-door travel times.

V. APPROACH AND METHODOLOGY

A. Dataset

The mobile phone data used for this study consist of a set of anonymised CDRs provided by Orange Spain, covering several months of 2016. Orange is currently the second largest mobile network operator in Spain in terms of number of clients, with a market share of around 27%. The dataset includes anonymous user identifiers of Orange's clients, as well as of roamers that connect to Orange's network. In this study, we have only worked with the data corresponding to the residents in Spain, which constitute the majority of users of the Madrid-Barcelona corridor. For the day of study, the dataset provides a sample of active users of around 20% of the total population. The CDRs include: (1) sociodemographic data for each user (age and gender); and (2) spatio-temporal data: time and cell tower to which the user is connected every time an event occurs (a phone call starts/ends, an SMS is sent, a data transfer takes place, etc.). Triangulated data were not available in this dataset.

B. Sample Construction

We select a sample of users according to a set of criteria related with phone activity, such that CDRs are granular enough to determine the user's activities and trips along the whole day in a reliable manner (e.g., if a user switches off his mobile phone or has no activity during a long period of time, it is filtered out). After applying these criteria, we end up with a sample of around 10% of the total Spanish adult population. Here it is important to highlight that the size of this sample is influenced by the fact that we focus on one specific day. The sample would be larger if we looked, for example, at the average number of trips during a working day of March 2016.

C. Identification of User's Home

By using CDRs corresponding to several weeks of the period February-March 2016, we analyse the user's daily patterns to identify the user's home location, by means of an algorithm along the lines of that described in [13]. The accuracy with which this location can be determined depends on the density of antennas. Initially, we assume that the user's home is located within the Voronoi area corresponding to the tower to which the user is connected while he/she is at home. The next step is the assignment of the user to the corresponding census tract (the smallest territorial unit for which population data are available in the Spanish Census), since population expansion will be later on performed based on the census data at census tract level. In some cases, one Voronoi area may cover one or more census tracts, while in other cases several Voronoi polygons may be needed to cover a single census tract. Assuming that the market share of the mobile network operator is relatively homogeneous between adjacent tracts, the assignation method minimises the differences in the ratio between the sample and the real population in contiguous cells.

D. Generation of Activity-Travel Diaries

By analysing the CDRs, we determine the locations of the activities performed by the users during the day. When there are several consecutive registers in the same area during at least 20 minutes, this stay is identified as an activity; otherwise these registers are classified as in-transit registers that occur during trips between activities. It is to be noted that this criterion alone does not allow us to distinguish between real activities and intermediate stops between different segments of a multimodal trip if such stops have a significant duration, as it is usually the case of the stay at an airport and, to a lesser extent, at a train station.

E. Expansion of the Sample to the Total Population

In order to obtain meaningful results, the sample of mobile phone users is expanded to the total adult population. Based on the user's home location, in this study we simply apply an expansion factor at census tract level.

F. Determination of Corridor Users

We identify those users that performed an activity both in Madrid and Catalonia within the day of study and discard the rest, leading to a total of 20,393 people that travelled between the two regions during the day of study. This is of course an arbitrary criterion, chosen in this preliminary study mainly for the sake of simplicity, but it seems a reasonable choice considering that the HST connecting Madrid and Barcelona stops in Lleida, Tarragona and Gerona, the three other capital cities in Catalonia, all of them having small airports with no direct flights from/to Madrid.

Figure 1: Origin and destination areas considered for the analysis

A corridor trip is considered to be the displacement between the last activity in Madrid/Catalonia and the next activity in the other end of the corridor. As already explained, due to the algorithm used for activity identification, the stays at the airport and at the train station are likely to be labelled as activities. Since we are interested in door-to-door itineraries, when an activity is identified at these locations we consider the trip between the activity that is previous to the stay at the origin airport/train station and/or the next activity after the stay at the destination airport/train station. In the case of road trips, intermediate activities detected by the algorithm are treated as in-transit footprints, assuming that users travelling by bus or by car are likely to perform at least one intermediate stop.

G. Determination of Transport Mode

The next step is to identify the transport mode used for each of the corridor trips. The approach followed in this preliminary study is based on using the locations of the intermediate mobile phone registers that appeared during the trip. Another approach, based on using timestamps to estimate durations and from that modes, was discarded because it would not accurately detect mode during disruptions (where trip durations for HST and plane may be similar) or for users with a low rate of mobile phone registers. First, the most probable mobile phone cells to which a user would connect during a trip, for each one of the transport modes considered (road, HST and plane) are calculated (see Figure 2). These are the cells that give service to the roads and HST railways. Green areas, corresponding to air transport, were included because it was noticed that some intermediate mobile phone registers were detected for a few users that clearly travel by plane (as they appear in Madrid-Barajas and Barcelona-El Prat cells within an interval of less than 2 hours) along the take-off and landing trajectories, corresponding to connections made during the flight by users that didn't switched off their mobile phones.

Figure 2: Most probable intermediate cells for each transport mode (red: road, blue: HST, green: air)

To determine transport mode, in-transit registers are compared with the most probable cells for each transport mode. The transport mode assigned is the one that shares the highest number of these probable cells with the in-transit positions of the user. Despite being a rather simplistic approach, the detailed inspection of some individual trajectories and the modal split obtained (see section VI) suggests that, at least for this particular case, this approach performs reasonably well. During the next stages of the BigData4ATM project, several improvements to this method will be explored, such as introducing trip duration criteria and/or combining mobile phone data with flights, train and bus schedule data.

VI. RESULTS

A. Number or Trips

Using the definition of corridor trip explained in section V, a total of 24,463 trips are detected between Madrid and Catalonia for the day of study. Since we had detected 20,393 people that travelled between the two regions, this means that around 4,000 people are detected to make a return trip during the same day. No user was detected doing more than 2 corridor trips during the same day.

B. Modal Split

The modal shares estimated from mobile phone data are presented in Figure 3.

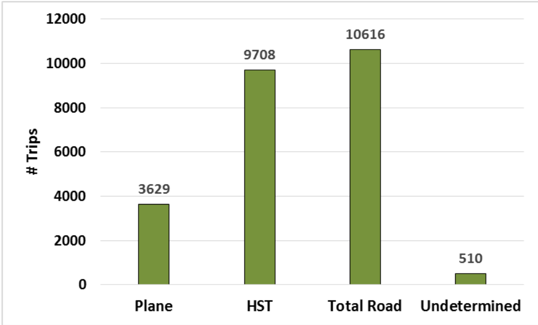

Figure 3: Modal split in the corridor estimated from mobile phone data

According to these results, if only plane and HST are considered, around 72% of passengers choose HST over plane. According to official data, in the Madrid-Barcelona corridor HST has a 62% market share when compared to plane [18], which is consistent with the results obtained here.

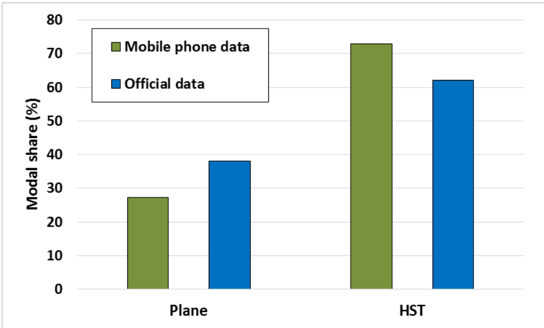

Figure 4: Comparison of the plane vs HST market data (green: estimation from mobile phone data for Tue 8th March 2016; blue: statistics provided by the Barcelona City Council [18] for March 2016)

The observed differences may be due to several reasons:

- We have analysed the modal share for a single day, while official data is aggregated on a monthly basis.
- We have considered trips not only between the cities of Madrid and Barcelona, but between the region of Madrid and Catalonia. Since the HST connects also with Lleida, Tarragona and Girona, which are not connected by flight with Madrid, the results shown in Figure 3 are not directly comparable to official data, which only takes into account the trips between Madrid and Barcelona.
- Transport mode was not detected for 510 trips due to an insufficient number of intermediate registers, which would be typically expected from air trips. The future improvements of the mode detection algorithm described in section V are expected to solve this issue.

As for the number of trips by road, there is no official source of information for the year 2016, but the overall figures are consistent with the road traffic count data available from the Spanish Ministry of Public Works and Transport [20]. Here again, the improvement of the algorithm to consider also the duration of the trip is expected to help better distinguish road trips from other transport modes.

C. Door-to-Door Origin and Destination

Figures 5 and 6 show the origins and destinations for the identified trips. In the case of Catalonia, it must be recalled that the HST line from Madrid to Barcelona also stops in Lleida, Tarragona and Girona, which explains the pattern observed in Figure 7. Since we have only retained those trips between the regions of Madrid and Catalonia, which is a purely administrative criterion, this leaves out other door-to-door origins and destinations (e.g., users travelling to Catalonia from Madrid surrounding provinces). This is particularly relevant in the case of Madrid, where the catchment areas of both the airport and the HST go well beyond the limits of the province.

Figures 7 and 8 show, for the metropolitan areas of Madrid and Barcelona, the trips originated or attracted that are made by HST and by plane, and the difference between them. It is visible that, the more central the location, the stronger the preference for the HST, with the exception of some areas located close to the airport, which shows the important role of terminal accessibility in demand captured by each mode.

Figures 5 to 8 illustrate the potential of mobile phone data for conducting a fine-grained characterisation of airport catchment areas. Such detailed characterisation will be carried out in subsequent steps of the project, extending the analysis to the whole range of origins and destinations of the corridor trips.

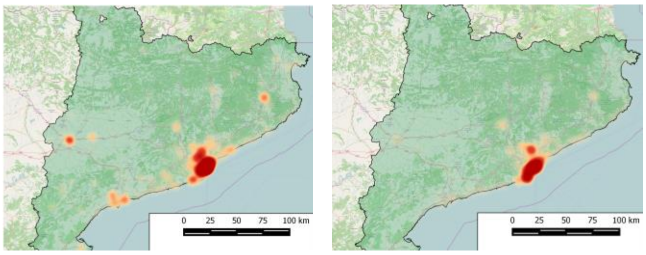

Figure 7: Catchment areas of Barcelona, Lleida, Tarragona and Girona HST stations (left) and Barcelona-El Prat airport (right)

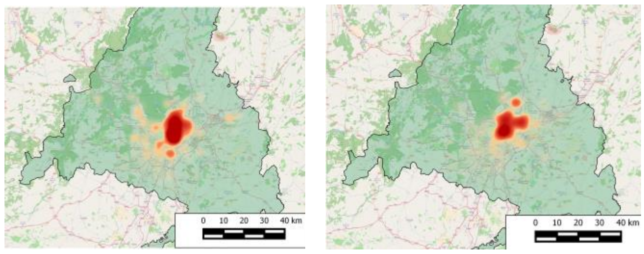

Figure 8: Catchment areas of Madrid-Atocha HST station (left) and Madrid-Barajas airport (right)

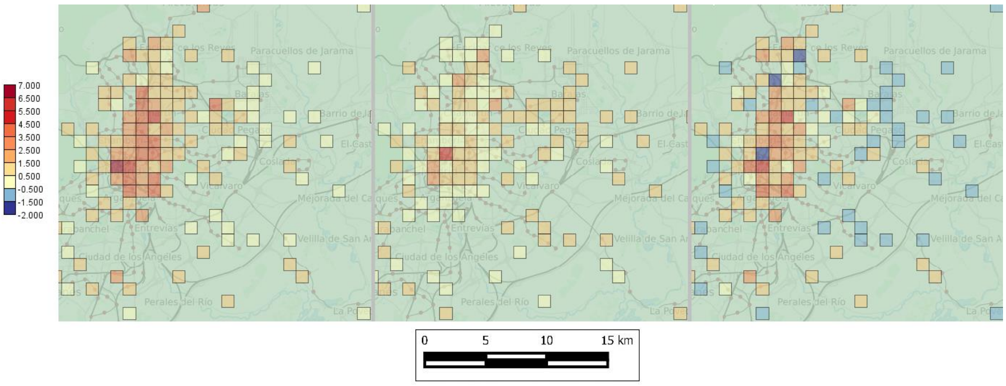

Figure 5: No of trips by HST (left), No of trips by plane (centre) and [No of trips by HST - No of trips by plane] (right) in Madrid metropolitan area

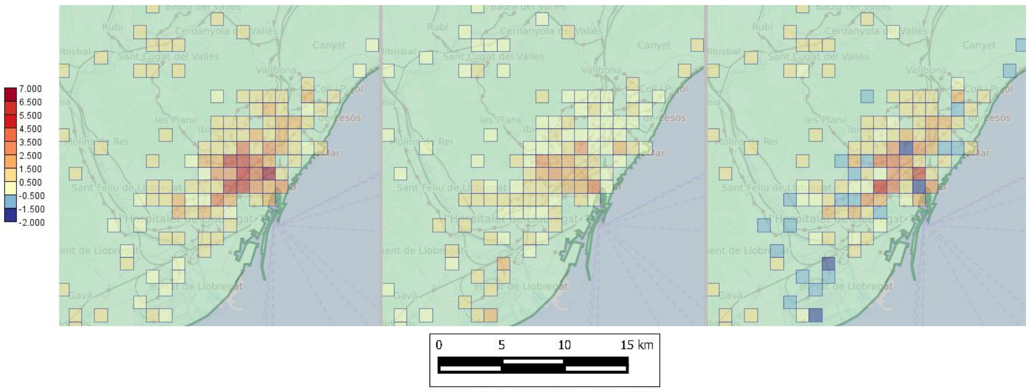

Figure 6: No of trips by HST (left), No of trips by plane (centre) and [No of trips by HST - No of trips by plane] (right) in Barcelona metropolitan area

D. *Door-to-Door Travel Times*

Using the door-to-door origin and destination activities, total door-to-door travel times for plane trips have been estimated. In Figure 9, the estimated door-to-door travel times are compared with the duration estimated for the air segment.

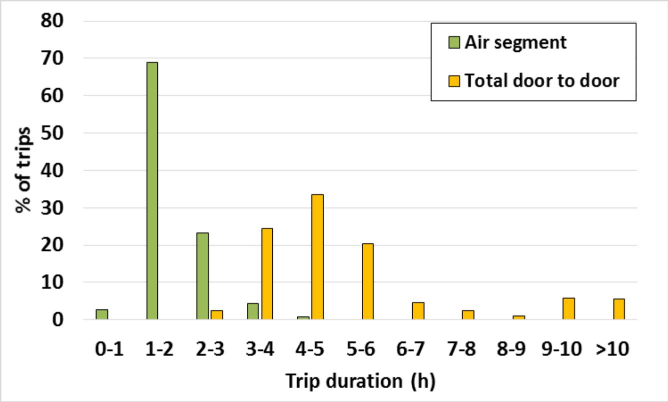

Figure 9: Distribution of air segment and door-to-door trip duration

It can be observed that the most part of the air segment durations are between 1 and 2 hours. This is consistent with the typical duration of the Madrid Barcelona flight, which is about 75 minutes. Longer durations correspond to cases where the mobile phone registers are away from flight departure/arrival (for example, when passengers switch off/on their mobile phones long before boarding/after landing).

With respect to the total door-to-door duration, it can be seen that all trips have a duration longer than 2 hours. Also, the distribution of durations is centred in the 4 to 5 hours interval. This means that, on average, the door-to-kerb, buffer times, airport processes and kerb-to-door times sum up to 3 hours. Taking into account that Madrid and Barcelona airports are not close to the city centre, these values look plausible.

It is also noticeable the presence of a few door-to-door durations above 8 hours. Looking at the individual sequence of mobile phone registers, it becomes evident that they correspond to passengers whose final destination was the airport (for example, the case of a person from Madrid who has a meeting at the Barcelona airport facilities). Therefore, in these cases the approach of considering the next activity after the airport as the door-to-door destination is not correct.

In the next stages of the project, the algorithm for door-to door trip determination will be revised to properly take into account these particular cases, and the analysis will be extended to conduct a comprehensive comparison of door-to-door travel times for the three modes and for different origin-destination pairs.

VII. Conclusions and Future Directions

The preliminary results of the study presented in this paper show the potential of mobile phone records as a source of detailed passenger centric information for air transport and ATM. By analysing a dataset of anonymised CDRs, we are able not only to replicate official aggregated data in a satisfactory manner, but also to provide new insights that cannot be obtained from traditional data sources. At the same time, these results show that there is significant room for improvement in the methodologies applied and provide useful hints on future research avenues that are to be developed in order to make the most of this type of data:

- The activity detection algorithm will be improved to better characterise the purpose of the trips (business, leisure, etc.). In addition to the algorithms already available in the literature for the identification of users' workplace/place of study, we will explore new algorithms based on both the information extracted from mobile phone records (e.g., duration of the stay at the destination) and the fusion with other sources (e.g., land use data).
- The characterisation of trips will be improved in several aspects. The mode detection algorithm for long-distance trips will be enhanced by taking into account the estimated duration of the trip at different moments of the day, using information about scheduled transport services as well as information from Google Maps API on the duration of car trips. In parallel, we will investigate the identification of the short-distance modes used for airport and HST stations access/egress, by blending the mobile phone data with other sources (e.g., data collected through the public transport smart card data, taxi GPS trajectories, etc.). This information will allow us to break down door-to-door travel times into the duration of the different trip segments.
- The sociodemographic information available from the mobile network operator will be used to merge the mobile phone records with census data and other official statistics, in order to refine the expansion methodology and build a synthetic population that replicates the sociodemographic characteristics of the real population.

The previous developments are expected to enable the investigation of a number of questions of paramount importance for European air transport and ATM. Some examples of these questions are presented below.

- The information available from mobile phone records can help develop better airport accessibility indicators and passenger-centric door-to-door delay indicators. These indicators could support the evaluation of the level of achievement of the 4-hour-door-to-door target and the assessment of new infrastructure, services and policies aimed at enhancing the connection of air transport with other transport modes.
- Mobile phone data can also provide valuable insights on the impact of ATM delays and disruptions on total passenger travel time, which is currently very difficult to measure due to the fact that conventional air

transport databases provide information on delays at flight level, without taking into account how these delays affect connections from the airport to the passenger´s final destination. This information could be a valuable input to Airport Collaborative Decision Making (A-CDM) for traffic flow and airport departure management, helping develop an extended A-CDM concept able to better account for passenger door-to-door journey, along the lines suggested by recent European research projects [21].

- Currently, air traffic forecasts typically combine flight statistics with econometric models that relate air traffic to observed variables whose correlation with the traffic values is plausible, such as demographic, macroeconomic and tourism variables. The segmentation of passengers according to their sociodemographic characteristics and trip purpose, together with airport accessibility indicators, could help develop improved traffic forecasting methodologies, thanks to larger, more detailed, cheaper and permanently updated behavioural data samples including multimodal information. The newly developed door-to-door delay indicators could also help investigate the relationship between demand and travel time reliability.

These and other related research questions will be addressed in the subsequent stages of the BigData4ATM project through different case studies.

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] European Commission, "Flightpath 2050. Europe´s Vision for Aviation. Report of the High Level Group on Aviation Research", 2011.

[2] A. Cook, G. Tanner, and M. Zanin, "Towards superior air transport performance metrics – imperatives", Journal of Aerospace Operations, vol. 2, nº 1-2, pp. 3-19, 2013.

[3] J. L. Toole, S. Colak, B. Sturt, L. Alexandre, A. Evsukoff, and M.C. González, "The path most travelled: Travel demand estimation using Big Data resources", Transportation Research C: Emerging Technologies, vol. 58 part B, pp. 162-177, 2015.

[4] M. Picornell, T. Ruiz, M. Lenormand, J. J. Ramasco, T. Dubernet, and E. Frías-Martínez, "Exploring the potential of phone call data to characterize the relationship between social network and travel behavior", Transportation, vol. 42, nº4, pp. 647-668, 2015.

[5] M. Lenormand, A. Tugores, P. Colet, and J. J. Ramasco, "Tweets on the Road", PLoS ONE 9(8): e105407. doi:10.1371/journal.pone.0105407, 2014.

[6] S. Sobolevsky, I. Sitjo, R. Tachet Des Combes, B. Hawelka, J. Murillo Arias, and C. Ratti, "Money on the move: Big Data of bank card transactions as the new proxy for human mobility patterns and regional delineation. The case of residents and foreign visitors in Spain", IEEE International Congress on Big Data, 2014.

[7] H. Samiul, C. M. Schneider, S. V. Ukkusuri, and M.C. González, "Spatiotemporal patterns of human mobility", Journal of Statistical Physics, vol. 151, nº 1-2, pp. 304-318, 2013.

[8] M. Lenormand, M. Picornell, O.G. Cantú-Ros, A. Tugores, T. Louail, R. Herranz, M. Barthelemy, E. Frías-Martínez, and J.J. Ramasco, "Cross-checking different sources of mobility information", PLoS One 9, no. 8, 2014.

[9] SESAR, "Definition Phase D3: The ATM target concept", 2007.

[10] EUROCONTROL, "Challenges of Air Transport 2030: Survey of Experts´Views", 2009.

[11] EUROCONTROL, "Synthesis of strategic and socio-economic studies at EUROCONTROL 2003-2010", 2010.

[12] BigData4ATM Consortium, "Big Data Analytics for Socioeconomic and Behavioural Research in ATM", 2016.

[13] L. Alexander, S. Jiang, M. Murga, and M.C. González, "Origin–destination trips by purpose and time of day inferred from mobile phone data", Transportation Research Part C: Emerging Technologies vol. 58 pp. 240-250, 2015.

[14] M. Picornell and L. Willumsen, "Transport Models and Big Data Fusion: Lessons from Experience", Proceedings of the European Transport Conference, 2016.

[15] H. Wang, F. Calabrese, G. Di Lorenzo, and C. Ratti, "Transportation mode inference from anonymized and aggregated mobile phone call detail records", Intelligent Transportation Systems, Funchal, 2010.

[16] J. Doyle, P. Hung, D. Kelly, S. McLoone and R. Farrell, "Utilising mobile phone billing records for travel mode discovery", 22nd IET Irish signals and Systems Conference, Dublin, 2011.

[17] T. Holleczek, L. Yu, J. Kang Lee, O. Senn, C. Ratti, and P. Jaillet, "Detecting weak public transport connections from cellphone and public transport data", Proceedings of the 2014 International Conference on Big Data Science and Computing, 2014.

[18] Barcelona City Council Statistics, available online at: http://www.bcn.cat/estadistica/angles/dades/economia/transport

[19] J. C. Martín, C. Román, J. C. García-Parlomares and J. Gutiérrez, "Spatial analysis of the competitiveness of the high-speed train and air transport: the role of access to terminals in the Madrid-Barcelona corridor", Transportation Research Part A: Policy and Practice, vol. 69, pp. 392-408, 2014.

[20] Spanish Ministry of Public Works and Transport Traffic Statistics:: http://www.fomento.gob.es/MFOM/LANG_CASTELLANO/DIRECCIONES_GENERALES/CARRETERAS/TRAFICO_VELOCIDADES/

[21] I. Laplace, A. Marzuoli, and E. Feron, "META-CDM: Multimodal, Efficient Transportation in Airports and Collaborative Decision Making", Airports in Urban Networks, 2014.