Data fusion for the analysis of air travel behavior: Application to Palma de Mallorca Airport

J. Blasco-Puyuelo, C. Blasco, R. Jordá-Muñoz, J. Burrieza, O. G. Cantú Ros, D. Mocholí
Nommon Solutions and Technologies
Madrid, Spain

***Abstract*— The European transport policy envisages a multimodal, passenger-centric transport system that allows travelers to reach their destination by the most efficient and sustainable combination of modes. Achieving this vision calls for an in-depth understanding of air passengers' door-to-door travel behavior. This article presents a set of data analysis and fusion methodologies for the door-to-gate and gate-to-door reconstruction of the passenger journey through a coherent combination of anonymized mobile network data with a wide range of heterogeneous data sources. The proposed approach is demonstrated and evaluated through a case study conducted in the Palma de Mallorca International airport.**

Keywords- air travel behaviour, machine learning, data analysis, data fusion, mobile network data, door-to-door journey

I. Introduction

A. Background and motivation

The European high-level vision on transport foresees a multimodal, passenger-centric system that allows passengers to travel from their origin to their destination in a seamless, efficient, predictable, environmentally-friendly and resilient manner [1]. In the context of an air transport system thoroughly integrated with other transport modes, airports are expected to become multimodal connection platforms, creating the conditions for travelers to reach their destination by the most efficient and sustainable combination of modes [2].

An essential enabler of this vision is the ability to characterize door-to-door air travel demand behavior, as a necessary condition to design a system able to satisfy passenger preferences, needs and constraints. Passenger access/egress legs and terminal legs have usually been addressed separately, thus preventing a comprehensive view of the passenger flows and the impact of different events on the complete transport network.

The analysis of air passenger behavior has traditionally relied on surveys. Passenger surveys provide a detailed characterization of the respondent, but are expensive and time-consuming, which limits the sample size and the frequency of update. In recent years, different studies have explored how the digital traces generated by personal mobile devices can be exploited to study passenger behavior, measure door-to-door travel times and characterize airport catchment areas. Mobile network data has been identified as particularly suitable for this purpose, thanks to the possibility of working with large, well-distributed population samples with high temporal and spatial resolution ([3], [4]).

The use of mobile network data enables the continuous monitoring of air travel demand, which is particularly valuable at the present moment, when the volatile environment created by digitalization and exacerbated by the COVID-19 pandemic is modifying mobility patterns and passenger preferences at timescales that are often not observable through traditional surveys. From mobile network data, passenger travel diaries covering all the steps of the door-to-door journey can be derived ([3], [4]). However, mobile network data do not provide a number of key features about the profile of the users and the characteristics of the identified trips, such as the trip purpose (business vs leisure), which can have a strong influence on passenger behavior. Additionally, while the spatio-temporal resolution of the data makes it possible the detailed characterization of long-distance trips, including the identification of the transport mode through map-matching techniques, this is not often the case for the modal choices in the access/egress legs to/from the airport (which is crucial information for the design of multimodal strategies), as those trips usually take place in urban areas [5]. Lastly, although the use of mobile network data allows the identification of relevant characteristics of the surface leg such as the origin of the trip and the arrival time to the airport, its contribution to the description of the passenger mobility patterns within the airport terminal is notably more limited due to the spatial granularity of the data.

B. Previous work

During the last decade, mobile network data have gained recognition in transport planning and are now commonly used to derive mobility indicators. These data register the interactions between mobile devices and the network antennas. The spatio-temporal sequence of registers can be analyzed to derive activities and trips, producing a mobility diary for each user; these diaries can then be extrapolated to the whole population using census data and other sociodemographic statistics, in a similar process to the sample expansion of a traditional travel survey.

The use of mobile network data has come a long way. The first studies focused on dynamic population mapping [6], deriving the number of people at a given location and at a specific time. Subsequent studies were able to estimate trips and build origin-destination matrices ([7], [8]). Recent advancements have addressed aspects such as the identification

The project leading to these results has received funding from the SESAR Joint Undertaking under grant agreement No 891287 under European Union's Horizon 2020 research and innovation program

of transport mode and route choice in long-distance trips by using map-matching techniques [9] and travel times [10].

A number of studies have explored the use of mobile network data to analyze the air travel behavior. The work developed in [11] describes a methodology for the analysis of the trips performed in Madrid-Barajas airport, including the expansion of the mobile network data sample using daily passenger counts and the inference of unknown passenger characteristics (e.g., trip purpose) using heuristic rules based on passenger surveys. The work done in [12] addressed the estimation of unknown characteristics, such as trip purpose and airport access mode, using machine learning models previously calibrated with passenger surveys.

C. *Objectives of the study*

This study aims at enhancing the work performed in [11] by developing and validating a methodology to enrich the activity-travel diaries (hereinafter, ATD) obtained from mobile network data with additional information extracted from a wide range of data sources. This enrichment comprehends: (i) the adjustment of the trips detected with mobile network data to the actual number of airport passengers using airport flight schedules and ticketing data, (ii) the addition of trip characteristics (trip purpose and access mode), based on the development of machine learning models trained with passenger survey data, following the approach described in [12], and (iii) the enhancement of the characterization of the terminal leg by using boarding card reader and mobile apps geolocation data. The ultimate goal is to enable a more detailed and comprehensive characterization of door-to-gate and gate-to-door passenger behavior. This paper describes the proposed approach and its application to the Palma de Mallorca International airport (hereinafter PMI).

II. DATA AND METHODOLOGY

A. *Datasets*

The data used for this study are the following:

- Mobile network data (MND) from July and August 2019. This data consists of a set of anonymized mobile phone records obtained through a collaboration agreement with one of the main three mobile network operators (MNOs) in Spain, with a market share over 20%. The mobile phone records used for the study do not provide the exact location of the users, but the cell to which the user is connected when the interaction with the network takes place. Therefore, the accuracy of the location depends on the density of the mobile network, which typically ranges from dozens to hundreds of meters in urban environments, to several kilometers in rural areas. In addition to the geolocated events, the data also include some basic sociodemographic information linked to each anonymized user, such as age, gender and nationality.
- Flight schedules for July and August 2019. Airport flight schedules include flight information such as the airport of destination/origin, terminal gate, flight type (regular, charter, training, ambulance, etc.), Scheduled In/Off-Block Time (SIBT/SOBT), Estimated In/Off-Block Time (EIBT/EOBT), etc., as well as aggregated demand information (actual number of passengers per flight).
- Passenger surveys conducted at PMI between the 18th and the 24th of July 2018. The surveys include information on different trip characteristics (destination, purpose of the trip, duration of the stay, access mode, etc.) as well as sociodemographic details (age, gender, nationality, etc.)
- Public transport ticketing data for July and August 2019 provided by the Palma de Mallorca municipal transport company, EMT Palma, which operates the public bus services that connect the airport with the city. The data includes hourly demand on the two bus lines serving the PMI airport
- Boarding card reader (BCR) data from October 2019 to March 2020. These data capture the timestamp of each passenger that scans his/her boarding pass at the BCR located before the security area, including flight-related information (flight number, SOBT, EOBT, etc.).
- Mobile apps data from August 2019, obtained from a provider that aggregates and commercializes the geolocation data collected by a variety of mobile apps. The dataset includes anonymized geolocated data generated when mobile phones make use of certain apps. Location data is based on GPS and/or Wi-Fi sensors, thus providing a more precise location than the mobile network data.

B. *Methodology*

The approach followed for reconstructing the ATD relies on the coherent fusion of the different data sources. The process takes as a starting point the ATD built from anonymized mobile network data, by applying the data processing pipeline described in [11].

The steps that have been followed to obtain the complete ATD are shown in Fig. 1 and described in the following subsections.

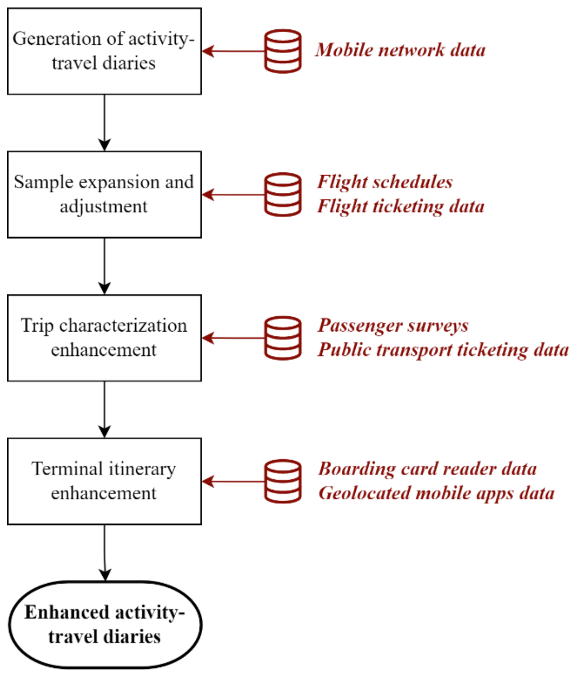


Figure 1. Methodology for the enhancement of the activity-travel diaries

1) Sample expansion and adjustment

The passenger trips detected using MND represent a sample of the total trips performed at the airport. Table I presents the sample size of passengers detected using MND for July and August 2019. For this study, only regular and charter flights have been considered, representing the 99.85% of the total passengers of the airport (cargo, ambulance and other residual categories have been discarded).

Previous weighting processes have been performed using census data and official tourism statistics to expand the samples of national travelers and foreign visitors, respectively, to the total population under study. However, further adjustment is required so the total number of users match the actual number of passengers using PMI. To do so, a statistical matching process has been performed using the airport flight schedules in order to assign each detected trip to an existing flight:

- A set of potential flights is assigned to each of the trips detected with MND based on two main features: the departure time and the origin/destination.

  From MND, the last user register at the airport terminal is available. This time is compared with the Estimated Off-Block Time (EOBT) of the available flights according to the flight schedules and "compatible" flights are initially selected.

TABLE I. SAMPLE SIZE OF THE PMI PASSENGERS DETECTED USING MND

Month	*Passengers according to flight schedules*	*Passengers detected using MND*	*Sample size*
July	4,199,919	461,802	10.99%
August	4,275,063	470.471	11.01%

The origin/destination of those flights is then compared with the actual origin/destination information available from the MND. The level of detail of this information depends on the trip and user characteristics. For domestic flights, the airport of origin/destination can be derived from MND by searching the nearest airport to the last/first user register location at their origin/destination. For international flights, information regarding the country of arrival/departure is available for national passengers (thanks to roaming-out registers), but no information is available for foreign (roaming-in) users. In this latter case, it is assumed that passengers are flying to/from their country of residence; if there are no flights that meet these conditions, flights to the other abroad destination are allowed.

- An iterative flight selection process is performed to assign each trip to one of the available flights. This process considers that all the initially selected flights are equally likely. The selection process is based on the estimated occupation of the flights, so as to first fill emptier flights. The probability of the passengers (p_p) to be assigned to a single flight (f_i) is proportional to:

$$p_{p\to f_i} = P^2_{real_i} / P_{assign_i} \tag{1}$$

  where P_{real_i} is the actual number of passengers on the flight i (according to the flight schedules) and P_{assign_i} is the number of passengers already assigned to flight i during the process. Note that the actual number of passengers (P_{real_i}) is squared so, in the case of similar occupancies, the algorithm tends to first fill busier flights.

Once all the MND detected passengers have been assigned to a specific flight, passengers' expansion weights (previously calculated using census and tourism data) are adjusted so the sum of the weights of all the passengers attending each flight coincides with the actual number of passengers.

The objective of this process is twofold: first, to adjust the number of passenger trips estimated with MND to the actual number of passengers registered at the airport; second, to assign each passenger to a specific flight, in order to integrate all the information attached to that flight (flight number, departure time, origin/destination airport, airline, gate, etc.). This information will be later used to infer different passenger characteristics and reconstruct the passenger itinerary within the terminal.

2) Enhancement of trip characterization

PMI passenger surveys have been used to calibrate machine learning models able to estimate the purpose of the trips (business or leisure) and the transport mode used by the passengers to access and egress the airport terminal (the categories available from the survey are: private bus from tour operators, public bus, rental car, private car and taxi/ride-sharing services), following the approach described in [12].

Due to the unavailability of some relevant features for some users (e.g., the lack of information on the sociodemographic profile of roaming-in users, as they are not customers of the MNO), and the fact that passenger surveys are limited to departing passengers, different machine learning models are developed for each specific user type.

The same features used to calibrate the models with the passenger surveys are then extracted from MND so the models can be applied to the reconstructed ATD to estimate the unknown properties, i.e., trip purpose and transport mode. Since the passenger surveys are only conducted for departing passengers, it has been assumed that the behavior and characteristics of passengers regarding the trip purpose and the surface access modal choices remain the same for both arriving and departing trips.

In the case of transport mode classification, a mixed approach using machine learning models and public transport ticketing data has been implemented to validate and adjust the number of public transport users detected using the models. The public transport data on the hourly demand of the lines serving the airport is used to validate that the hourly amount of public transport users detected using the machine learning models does not exceed the actual demand. Further adjustment is not possible as the public transport is also used by airport workers — while the ATD of airport workers is also available, it is not possible to use the previous machine learning models as they have been trained exclusively with passenger data. In the future, this limitation could be overcome by using data from other travel surveys (e.g., a traditional household survey on daily mobility in the metropolitan area could be used to train similar models for airport workers and other visitors).

3) Enhancement of passenger terminal itineraries

The objective of this step is to analyze the passengers' trajectory within the airport terminal. From MND, and after the previous adjustment process, information on the passengers' arrival time to the airport and departure time is available; however, passenger information between these two milestones is absent.

A mixed approach that analyses and fuses MND, BCR data and geolocated mobile app data has been developed to characterize the passenger behavior within the terminal. BCR data is used to estimate the passengers' arrival to the security control, while mobile app data is used to characterize the passenger stay times at the different airport areas and facilities.

Regarding the estimation of the passengers' arrival to security control, in an initial calibration phase BCR data is used to analyze the presentation curves for different passenger flows, aiming at extracting the main factors that determine the behavior of passengers when accessing the security control, which has been proved to be the time of the day and the destination (see Fig. 2).

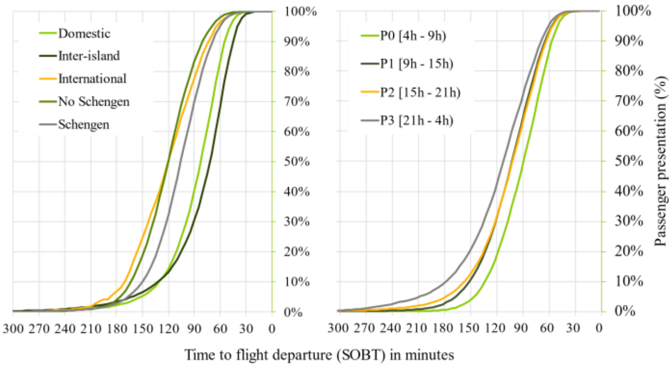


Figure 2. Passenger presentation profiles at BCR depending on the passenger final destination (left) and the period of the day (right).

These curves are then approximated by probability distributions. Gamma and normal distributions have been explored in order to assess which one provides a better approximation. Subsequently, in an assignment phase, these probability distributions are applied to the trips detected with MND to estimate the arrival to the security control.

Mobile apps data has been used to reconstruct the passenger terminal itineraries between the security check and the gate. The check-in area has not been considered for this study due to the PMI layout, as the check-in desk area is located underneath the security control area, complicating the analysis of the registers as they do not contain altitude information. The following study areas have been defined: security control, commercial area, passport control and modules (A, B, C and D).

As the mobile app registers can be located anywhere inside the airport (see Fig. 3), feasible travel paths have to be defined so that passengers are only able to travel along such paths. To do so, the airport area has been translated into a squared network. This network connects every square with its neighbors, allowing the flow of people between one square and the surrounding squares. The airport areas defined and the 5x5 meter grid constructed are presented in Fig. 4. The construction of the airport travel paths comprises the following steps: (i) the assignment of the mobile app registers' location to the centroid of the square they fall into, (ii) the connection of each register to the immediately next register following the shortest path, and (iii) the calculation of the timestamp in each square of the shortest path, assuming constant speed between the two points.

From these trajectories, the passenger stay times in the different airport areas are derived, allowing the calculation of passenger stay time distributions in those areas. These distributions are fitted into gamma probability distributions and are assigned to the previously calculated ATD based on different relevant features available in both the mobile apps data and the MND, namely the period of the day and the module the passenger is accessing to (A, B, C or D).

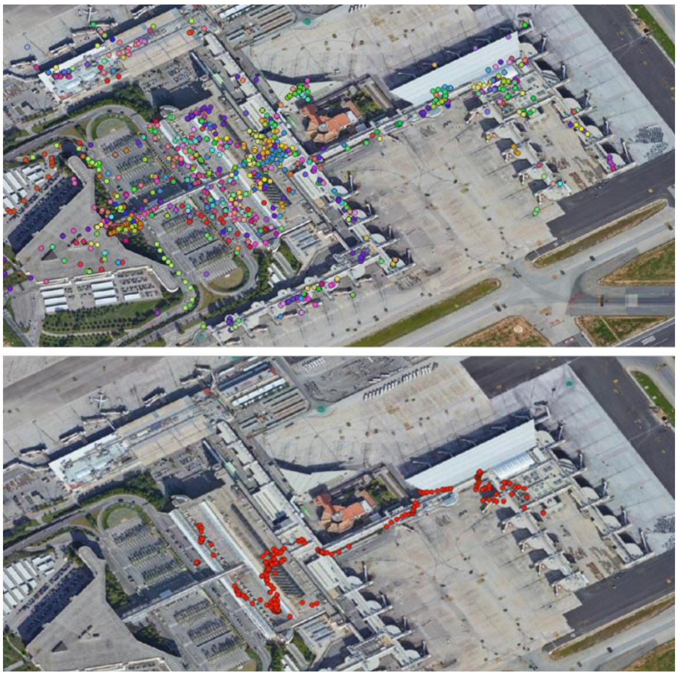


Figure 3. Sample of mobile app registers for a single day categorised by user (top) and example of single user trajectory (bottom).

Figure 4. PMI study areas and 5x5m network grid.

III. RESULTS

A. *Sample expansion and adjustment of the ATD*

The described methodology has allowed the expansion of the sample of trips detected with MND to the actual number of PMI users in an accurate way. A set of validations have been conducted based on the comparison of indicators extracted from the ATD with additional data sources in order to verify that the information extracted from both sources is coherent.

The data used for this validation experiment are the PMI passenger surveys. Fig. 5 shows the correlation between the main nationalities detected in the airport according to the surveys and the ATD.

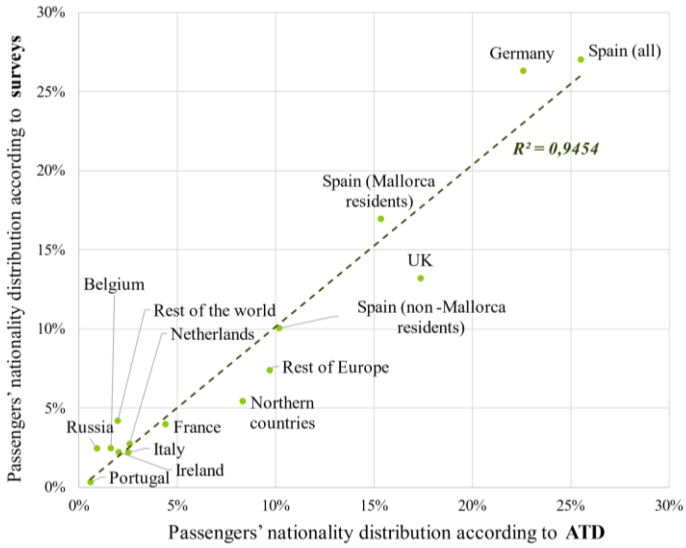


Figure 5. Passengers' nationality correlation between distributions observed in the surveys and the ATD

Fig. 6 shows a comparative analysis between the airport catchment areas obtained using the MND-based ATDs and the passenger surveys. Both data sources display a strong correlation. A relevant advantage of using MND is that the larger sample size allows the obtention of more disaggregated data than that obtained with the passenger surveys: in the municipality of Palma, for example, the origin of the passengers has been disaggregated by districts; this information is not available with the passenger surveys, which only provide information at municipality level.

B. *Enhancement of trip characterization*

Table II presents the performance of the machine learning models developed for each specific passenger type. The performance has been measured using the F1-Score metric, which computes the harmonic average between precision (number of elements correctly classified with respect to the total elements the model predicts to be in the positive class) and recall (number of elements correctly classified into the positive class out of the total elements that are labelled with the positive class).

TABLE II. PERFORMANCE OF THE MACHINE LEARNING MODELS

User type		*Trip purpose model F1-Score*	*Transport mode model F1-Score*
Residence	*Direction*		
National	Outbound	0.75	0.42
National	Inbound	0.75	0.39
Foreign	Outbound	0.70	0.39
Foreign	Inbound	0.69	0.38

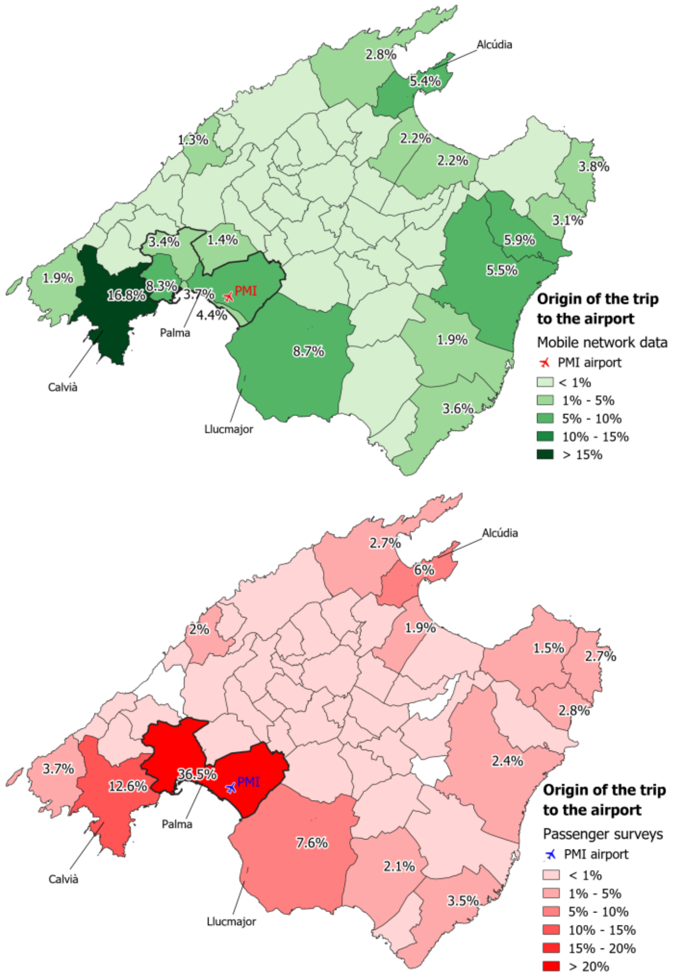


Figure 6. Airport catchment area according to mobile network data (top) and passenger surveys (bottom)

The results are promising. The trip purpose model performance ranges from around 0.70 to 0.75 for two classes, which implies a notable improvement compared with the use of a random model. In Table III, a more detailed summary of the results achieved for the 'National-Outbound' trip purpose model is presented. Especially relevant is the ability of the model to accurately distinguish business passengers, as their presence on the training sample is almost negligible (around 5%). As for the transport mode model, average performance is around 0.40, which is twice the F1-Score provided by a random model (see Table IV), achieving a notable performance for all the classes despite the fact that some classes are less represented (e.g., 'Public bus').

TABLE III. PERFORMANCE OF 'NATIONAL-OUTBOUND' TRIP PURPOSE MODEL

Class	*Precision*	*Recall*	*F1-Score*	*Training sample size*
Leisure	0.98	0.97	0.97	3,906
Business	0.49	0.57	0.53	210
Average	**0.73**	**0.77**	**0.75**	**4,116**

TABLE IV. PERFORMANCE OF 'NATIONAL-OUTBOUND' MODE CHOICE MODEL

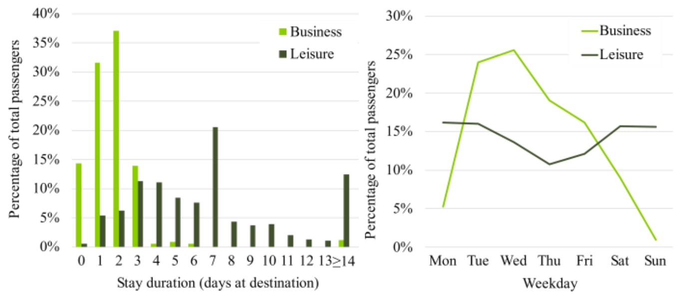
Class	*Precision*	*Recall*	*F1-Score*	*Training sample size*
Private bus	0.42	0.51	0.46	983
Public bus	0.17	0.39	0.24	344
Rental car	0.45	0.41	0.43	956
Private car	0.71	0.60	0.66	774
Taxi/ride-sharing services	0.42	0.26	0.32	1,052
Average	**0.44**	**0.43**	**0.42**	**4,109**

The results of assigning trip purpose with the proposed models are presented in Table V, where the business/leisure distribution for the period of study is compared with the distribution provided by the surveys used for training the models. The proportion of business passengers estimated is slightly higher than the values observed in the survey. A potential explanation could be a slightly higher reluctance of business passengers to answer airport surveys.

Fig. 7 shows the distribution of business and leisure passengers based on the passenger stay duration (a) and the weekday of the trip (b). It can be observed that the models developed tend to assign business purpose to passengers with shorter stay durations (passengers with trip durations of more than 3 days are rarely categorized as 'business') while longer trips are usually assigned a leisure purpose; also, trips performed during the weekdays rather than on weekends have higher probability of being classified as business trips.

TABLE V. BUSINESS/LEISURE ESTIMATED DISTRIBUTIONS

Period	*Business passengers*	*Leisure passengers*
July	6.8%	93.2%
August	6.0%	94.0%
Survey	**5.3%**	**94.7%**



(a) (b)

Figure 7. Business/Leisure passengers distribution according to the duration of the stay at the trip destination (a) and the weekday when the trip is performed (b)

A similar analysis has been performed for the transport mode classification. In Table VI, the modal share obtained for the period of study is compared with the modal share observed in the passenger surveys. The behavior observed in both cases is significantly similar.

In Fig. 8, the transport mode distribution is presented depending on the most relevant feature: the place of residence of the user. The models tend to assign ‘Private car’ rather than any other transport option to Mallorca residents, which is understandable as this kind of users will have access to a private car on the island (unlike visitors from outside the island). ‘Public bus’ and ‘Taxi’ categories are also assigned. ‘Rental car’ and ‘Private bus’ are discarded by the models, which is logical as it is not usual that island residents use these modes to access the airport. Palma de Mallorca visitors’ modal choice is more regular, being ‘Private car’ the less used option, which is understandable for the reasons presented above. This behavior explains the performance presented in Table VI, in which ‘Private car’ class performance is the highest due to the fact that its use is mainly restricted to national users, followed by ‘Rental car’ and ‘Private bus’, which are practically restricted to visitors. On the other hand, ‘Public bus’ and ‘Taxi’ are the categories that are more difficult to be distinguished by the model as both national users and visitors use such modes. The distributions obtained present a high correlation with the patterns observed in the surveys, which suggest a proper translation of the passenger behaviors to the classification model.

An analysis of the hourly distribution of passengers using public transport modes has also been performed, comparing the information extracted from the MND-based ATDs and the information available from public transport demand data. The results are presented in Fig. 9, where it can be observed that the correspondence between both data sources is very high.

TABLE VI. MODAL SHARE ESTIMATED DISTRIBUTIONS

Period	*Private bus*	*Private car*	*Public bus*	*Rental car*	*Taxi*
July 2019	26.2%	20.0%	7.8%	21.5%	24.6%
August 2019	25.5%	23.2%	6.6%	18.3%	26.4%
Survey (2018)	**19.9%**	**21.4%**	**8.2%**	**23.9%**	**26.3%**

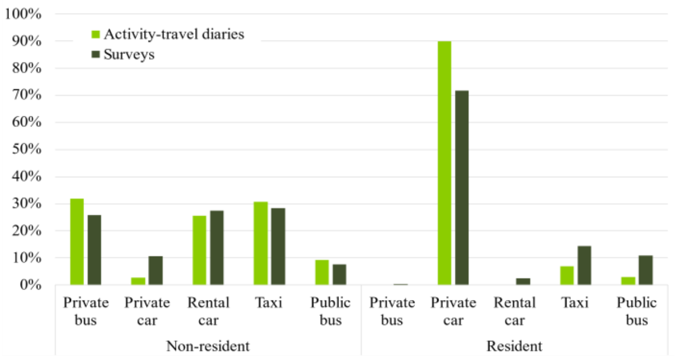


Figure 8. Transport mode distribution according to the passengers’ place of residence: non-residents vs residents in the island of Mallorca

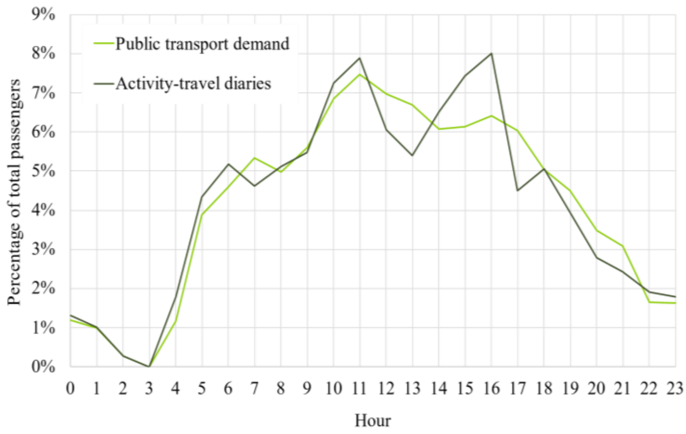


Figure 9. Hourly distribution of public transport users according to the ATD and the public transport demand data

C. Enhancement of passenger terminal itineraries

In Table VII, the performance of the BCR arrival estimation using normal and gamma probability distributions is presented. This performance is measured by calculating the average and maximum errors between the actual curve and the estimated one. As the gamma probability distribution is the one providing a better approximation, it has been the one selected for approximating the passenger profiles. In Fig. 10, a comparison between the actual curve and the estimation using a gamma distribution is presented for the worst-case scenario (worst approximation), which are Schengen flights performed during period P3 ([21-4h)). It can be observed that the consistency between the actual and the estimated curves is notably high.

TABLE VII. PASSENGER ARRIVAL TO BCR ESTIMATION ERRORS

Probability distribution	*Average error*	*Maximum error*
Normal distribution	1.4%	12.1%
Gamma distribution	0.8%	6.2%

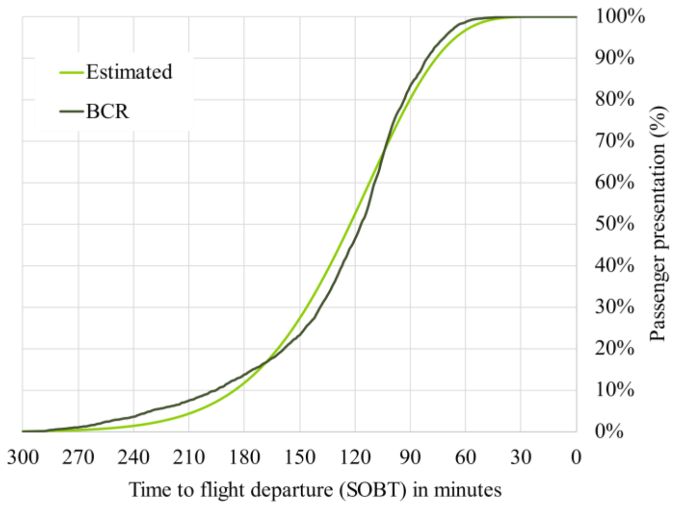


Figure 10. Actual vs estimated BCR arrival distribution

Fig. 11 presents an example of the trajectories extracted using the proposed methodology for the trips detected in a one week sample of mobile apps data. Fig. 12 shows an example of the passengers' stay time distributions at the passport control (a) and the commercial area located before entering the different modules (b). The quality of the data limited the possibilities of this approach due to its low sample size and temporal granularity (in many cases we have one or few registers in each area, which cannot be used to reliably infer the stay times), so the temporal segmentation of the periods has been limited to peak/off-peak periods, as a highly detailed segmentation (e.g., hourly) would have significantly reduced the sample in each group.

The application of the time distributions extracted using both the BCR data and the mobile apps data to the travel diaries allows the identification of the subsequent points of the passenger terminal itinerary: the security control entry/exit times, the commercial area entry/exit times, the passport control entry/exit times (if applicable) and the passenger module entry/exit times. This makes it possible to extract relevant indicators such as the occupancy or the throughput in the different defined airport areas. Fig. 13 and Fig. 14 present the occupancy of the different PMI modules and the incoming number of passengers per minute in the security and passport controls, respectively, for the busiest day of summer 2019 (Saturday 3rd August 2019).

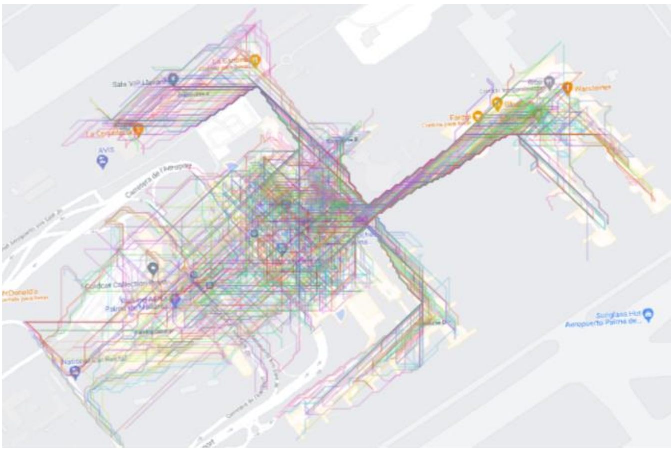
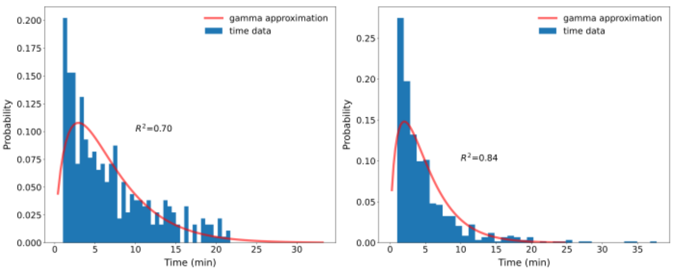


Figure 11. Sample of terminal trajectories extracted from the mobile app data



(a) (b)

Figure 12. Observed and modelled passenger stay times distributions in the passport control area (a) and the commercial area (b) during peak times

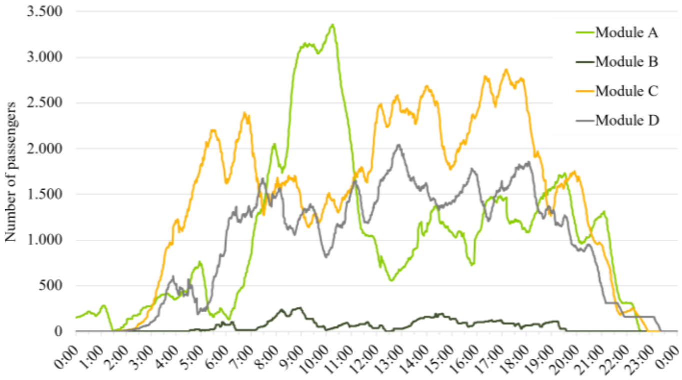


Figure 13. Estimated passenger occupancy in the terminal modules

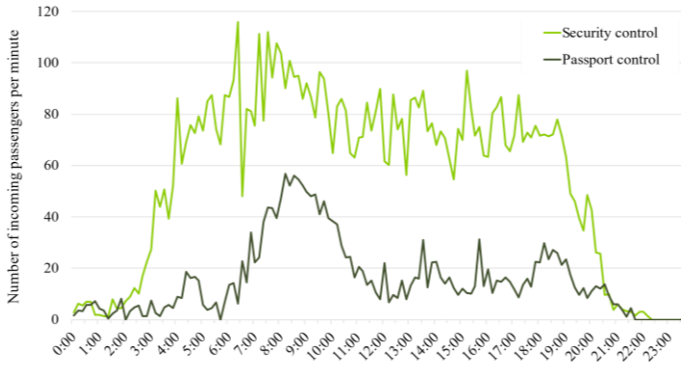


Figure 14. Estimated flow of incoming passengers per minute at the security control and the passport control

IV. Discussion

The methodology for the expansion of the MND sample based on a statistical flight matching algorithm enhances the work done in [11], where flight ticketing data was used for the expansion on a more aggregated level (on a daily basis). This flight-based adjustment approach makes it possible to assign each detected trip to a specific flight. This enables a more accurate expansion of the MND users as well as the assignment of flight-related information such as flight number, gate, airport of origin/destination, etc.

The results obtained for the inference of unknown characteristics through the use of machine learning models demonstrate that the proposed methodology provides satisfactory results and that the models calibrated using the passenger surveys are able to accurately assign the target variables to MND-derived ATDs. This extends the work done in [11], where trip purpose was assigned using rule-based algorithms based on heuristic rules extracted from the passenger surveys, and in [12], which focused on the calibration of the models but did not implement the assignment phase. In this paper, trip purpose and access mode have been inferred; this methodology can however be easily extended to the estimation of a wide range of other passenger features available in the surveys, such as the use of the airport facilities (e.g., check-in desks, shops, restaurants) or the group size in the case of group travelling.

The approach for the estimation of the passenger itinerary within the airport terminal using BCR and mobile apps data also provides satisfactory results, opening the door to a more detailed understanding of the passenger behavior within the airport terminal. Despite the reduced sample size and the limited temporal granularity of the date available for the study, it has been demonstrated that it is possible to reconstruct the passengers' itinerary with a reasonable level of certainty. Future work to enhance the description of the passenger behavior within the terminal should aim at increasing the mobile apps data sample and investigating the integration of other data sources which describe the movement of passengers within the terminal (data from Wi-Fi sensors or beacons, data on the check-in and boarding processes, etc.).

In summary, the methodology proposed in this paper allows the detailed characterization of door-to-gate and gate-to-door passenger flows with a level of detail not available before. This information can be used to adapt the transport system, including both the surface access alternatives and the airport terminal processes, to the needs and preferences of passengers. Future research should explore the application of this methodology to other types of airports, such as airports with a wider range of access alternatives (rail, metro, etc.), in order to see how to coherently integrate the information on the different surface access modes, and large international hubs where connecting flights represent a non-negligible proportion of passengers.

The information extracted using this methodology can be used to build predictive models able to short-term forecast the passenger flows ([13]), in order to simulate the door-to-gate and gate-to-door passenger journey in an integrated manner and assess the operational impact of different passenger flow management measures.

Acknowledgment

The project leading to these results has received funding from the SESAR Joint Undertaking under grant agreement No 891287 under European Union's Horizon 2020 research and innovation program.

References

[1] European Commission. White Paper: Roadmap to a Single European Transport Area – Towards a competitive and resource efficient transport system. 2011.

[2] European Commission. Flightpath 2050. Europe's Vision for Aviation. Report of the High-Level Group on Aviation Research. 2011.

[3] P. García-Albertos, 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", SESAR Innovation Days, 2016.

[4] P. García-Albertos, O. G. Cantú Ros, C. Ciruelos and R. Herranz, "Understanding door-to-door travel times from opportunistically collected mobile phone records. A case study of Spanish airports", SESAR Innovation Days, 2017.

[5] N. Breyer, D. Gundlegård, C. Rydergren, 2021. Travel mode classification of intercity trips using cellular network data. Transportation Research Procedia, 52:211 – 218. doi: 10.1016/j.trpro.2021. 01.024.

[6] J. Reades, F. Calabrese, A. Sevtsuk, and C. Ratti (2007). "Cellular census: Explorations in urban data collection". IEEE Pervasive computing, 6(3), 30-38.

[7] F. Calabrese, G. Di Lorenzo, L. Liu, and C. Ratti (2011). "Estimating Origin-Destination flows using opportunistically collected mobile phone location data from one million users in Boston Metropolitan Area".

[8] 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.

[9] P. García-Albertos, O. G. Cantú Ros and R. Herranz, "Analyzing door-to-door travel times through mobile phone data", CEAS Aeronautical Journal 11, no. 2, pp. 345-354, 2020.

[10] A. Kalatian, Y. Shafahi,. "Travel mode detection exploiting cellular network data" MATEC Web Conf., 81:03008. doi: 10.1051/matecconf/20168103008, 2016

[11] J. Burrieza-Galán, et al. (2022). "A methodology for understanding passenger flows combining mobile phone records and airport surveys: Application to Madrid-Barajas Airport after the COVID-19 outbreak". Journal of Air Transport Management, 100, 102163.

[12] A. Gregg, J. Blasco-Puyuelo, R. Jorda-Munoz, I. Martin-Martinez (2022). "Airport accessibility surveys and mobile phone records data fusion for the analysis of air travel behaviour". 12th International Conference on Transport Survey Methods, 20-25 March, Porto Novo Beach, Portugal. In press.

[13] P. Scala, M. M. Mota, J. Blasco-Puyuelo, O. G. Cantu Ros, C. Blasco (2022). "A novel validation approach for validating the simulation model of a passengers' airport terminal: case study Palma de Mallorca airport". 34th European Modeling & Simulation Symposium. Unpublished.