

A machine learning approach for predicting airport passenger flows

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

Abstract— The European aviation policy envisages a resilient air transport system seamlessly integrated into the European transport network, with the final objective of taking passengers from door to door predictably and efficiently while enhancing air transport experience. Meeting this vision calls for an in-depth understanding of air passengers' travel behavior and the ability to anticipate its impact on the performance of the air transport system. This paper presents a methodology to forecast the upcoming airport passenger flows for a particular day of operations, in order to help airports make more informed passenger flow management decisions and render the air transport system more resilient against adverse circumstances.

Keywords— air travel behaviour, machine learning, airport passenger flows forecasting, modal share, mobile network data

I. INTRODUCTION

A. Background and motivation

The European Commission, in its 2011 White Paper on Transport [1], envisages a multimodal, passenger-centric system able to allow passengers to seamlessly travel from origin to destination by the most efficient and sustainable combination of modes. Aligned with this vision, one of the main objectives of the European aviation policy is a resilient air transport system thoroughly integrated with other transport modes and able to provide passengers with efficient and seamless travel services while enhancing air transport experience [2].

Achieving this vision requires an air transport system able to satisfy passenger preferences, needs and constraints; a necessary condition to build that system is the capacity to characterize air travel demand behavior. Traditionally, the characterization of air passenger behavior has relied on surveys, which provide rich information about the passengers' preferences and profile. However, surveys also entail major drawbacks, as they are expensive and time-consuming, thus limiting the sample size and the frequency of update, which makes it difficult to monitor the increasingly fast changes in mobility patterns and passenger preferences brought about by digitalization.

During the last decade, different studies have addressed how data generated by personal mobile devices can be exploited to analyze air passenger behavior. Mobile network data has been identified as particularly appropriate for this purpose, thanks to the possibility of working with large, well-distributed population samples with high temporal and spatial resolution and obtaining information for all the legs of the

door-to-door journey ([3], [4]). Although mobile network data lacks certain key features about the profile of the users and the characteristics of the identified trips, it can be blended with additional data sources in order to provide a full picture of all the steps of the door-to-door journey [5].

In addition to monitoring passenger behavior, the information extracted from mobile network data can also be used to build predictive models able to forecast the passenger flows and anticipate the transport network behavior. A more up-to-date view of the upcoming passenger flows would allow airports to make a better use of their resources and coordinate with ground transport service providers. This would not only lead to a more efficient allocation of resources, but it would also render the transport system more resilient against adverse conditions, contributing to a better overall passenger experience.

B. Previous work

In recent years, mobile network data have gained recognition in transport planning and are now regularly used to extract mobility indicators. 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]). Different studies have explored the use of mobile network data to analyze air travel behavior. The work developed in [9] 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. The work done in [10] addresses the estimation of unknown trip characteristics, such as trip purpose and airport access mode, using machine learning models previously calibrated with passenger surveys. The work done in [5] extends this work and presents a methodology for the complete door-to-gate and gate-to-door reconstruction of the passenger journey based on the coherent combination of anonymized mobile network data with a wide range of heterogeneous 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 [10], and (iii) the enhancement of the characterization of the

The results presented in this study are part of the project TravelInt, a project funded by the European Union under the Next Generation-EU program. The project has been supported by the Spanish Ministry of Economic Affairs and Digital Transformation, through Red.es.

terminal leg by using boarding card reader and mobile apps geolocation data.

These studies, however, focus on the estimation of past passenger flows, but do not address the problem of forecasting future flows. While the insights provided by descriptive analyses are indeed valuable for the airports, to make more informed decisions airport also need to be able to anticipate future passenger flows in an accurate manner.

C. Objectives of study

This study extends the work developed in [5] by building a set of predictive models able to anticipate passenger behavior. The proposed approach takes as a starting point a historical synthetic population of passengers extracted by applying the methodology described in [5] and uses it to calibrate predictive models able to forecast the passenger arrival curves at the airport and the modal share in the airport surface access. This methodology is tested and validated by applying it to the forecast of passenger flows in the Palma de Mallorca Airport (hereinafter PMI).

II. DATA AND METHODOLOGY

A. Datasets

The data used for this study are the following:

- Mobile network data from July and August 2019 consisting in a set of anonymized mobile phone records obtained through collaboration with one of the main mobile network operators in Spain.
- Flight schedules and passenger demand for PMI during July and August 2019. Airport flight schedules include flight information such as the airport of destination/origin, airline, flight type (regular, charter, training, ambulance, etc.), Scheduled In/Off-Block Time (SIBT/SOBT), etc. Demand information consists in the number of passengers per flight.
- Passenger surveys conducted at PMI between the 18th and the 24th of July 2018 including information on different trip characteristics (destination, stay duration, access mode, etc.) as well as sociodemographic details (age, gender, nationality, etc.)
- Public transport ticketing data for July and August 2019 on the public bus services feeding the airport.

B. Methodology

The methodology presented in this paper takes as a starting point a historical synthetic population of passengers constructed by applying the data processing pipeline described in [5] to the mentioned datasets and uses it to build different predictive models in order to estimate two features which have been considered particularly relevant for airports, the arrival time of passengers at the airport and the access mode.

The steps that have been followed to predict the passenger flows are depicted in Figure 1 and described in the following subsections.

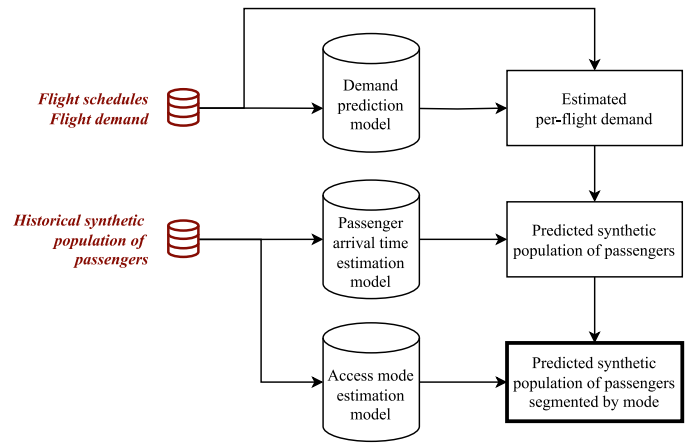


Figure 1. Methodology for forecasting the airport passenger flows

To verify the proposed methodology, the historical synthetic population of passengers from July 2019 were used to calibrate the predictive models. Subsequently, these models were applied to forecast the passenger flows of August 2019, and the results were validated against the actual synthetic population data (ground truth) from August 2019. This evaluation aimed to assess the performance of our methodology in accurately predicting passenger flows.

1) Per-flight demand prediction

Historical flight schedules and per-flight demand were used to calibrate a machine learning model able to predict the future demand of the scheduled flights. The machine learning model takes as input different features derived from the flight schedules and uses them to estimate their associated demand. The algorithm selected for this task is a random forest regressor.

At the time of training the model, the following considerations were taken into account:

- Only regular and charter flights, which represent 99.85% of the total passengers of the airport, were considered (cargo, ambulance, and other residual categories were discarded).
- The study focused on predicting the behavior of departing passengers, as the accurate forecasting of their behavior is considered more valuable for the airport.

The first step to calibrate the model is to extract relevant features able to explain the behavior of the passengers. From the data available, the following features were extracted:

- Flight type (Regular/Charter).
- Airline (RZR, IBE, etc.).
- Airport of destination (MAD, MXP, etc.)
- Country of destination (GB, ES, etc.)
- Scheduled time of the flight (0-23)
- Weekday (Mon-Sun)

Random forest regressor models do not handle categorical variables, therefore the categorical features were processed differently depending on the number of categories available within each variable. The categories ‘flight type’ and ‘weekday’ were encoded using label encoding, which consists in transforming the categorical value into a numerical format, while ‘airline’, ‘airport’ and ‘country’, due to the high number of classes available in each category, were encoded using target encoding, which consists in replacing the categorical value with the average (or any other relevant indicator) of the target variable.

Once the relevant features were extracted, a process for training the model was implemented. In order to maximize the performance of the model, the following processes were applied:

- The dataset was divided into a training set and a test set. The model is trained using only the training subset, while its performance is evaluated using the test subset, with the aim of preventing overfitting. 75% of the samples were used for training and 25% for testing.
- A Recursive Feature Elimination (RFE) algorithm was implemented. This algorithm analyzes different subsets of variables and selects the most relevant for estimating the target variable. As a result, it reduces computational time and minimizes the likelihood of overfitting.
- Hyperparameter tuning of the random forest regressor algorithm was applied. This technique tries different configurations of hyperparameters in order to find which combination provides the best performance.
- A k-fold cross validation resampling method was applied. The k-fold cross-validation method divides the training data into k separated portions, and trains the model k times using as training sample k-1 of those groups and validating the result on the remaining portion. The number of folds or groups used in the cross-validation process was set to 5; this leads to an 80/20 partition of the training data, which seems appropriate given the size of the dataset.

During the training process, the average validation score was computed for each data stratification and for every combination of hyperparameters, in order to select the combination that provides the highest performance.

Finally, a random forest regressor model was fitted on the whole training set using the hyperparameters selected in the previous step. The performance of this final model was evaluated on the test set that was set aside in the first step to obtain a final indicator of how well the model behaves with unseen data.

Once the model was calibrated, it was used to forecast passenger behavior for August 2019, comparing its results with the observed demand behavior in order to assess the

performance of the model for predicting demand for different months to those used for calibrating the model.

The results obtained are presented in Section III.A.

2) Passenger arrival time prediction

To estimate the arrival time at the airport of each of the passengers forecast in the previous step, information derived from the historical synthetic population of passengers is used. To model this behavior, the passengers’ arrival time before departure (hereinafter earliness) is computed. The distribution of this feature can be accurately approximated by gamma probability distributions.

The passengers’ earliness behavior is strongly influenced by factors such as the flight’s destination, time and day of the week. For instance, passengers travelling to international destinations typically arrive earlier to the airport than those travelling to domestic destinations. Similarly, passengers departing during the early hours of the day tend to arrive later to the airport compared to other times, as no congestion is expected during those hours. This behavior can be observed in Figure 2, which displays passenger presentation curves (earliness cumulative distribution function) extracted from the historical synthetic population, categorized by passenger destination and the time of the day.

Other factors influencing passengers' arrival time at the terminal include trip purpose. Typically, business passengers tend to arrive later at the airport, especially when traveling with minimal luggage. Unfortunately, the dataset obtained in the previous step, which primarily comprises flight information and passenger demand, does not include such details, making it impossible to incorporate these features into the analysis.

Considering this, the methodology to model and assign this behavior to the forecast demand is outlined as follows:

- Passenger presentation curves calculation. As the behavior is highly dependent on the destination, time and day of the week, the passenger presentation curves are computed for each crossed-segmentation of variables.

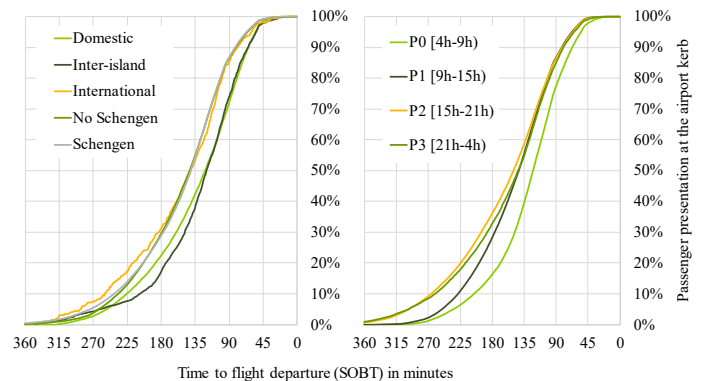


Figure 2. Passenger presentation curves at the airport terminal depending on the destination (left) and the time of the day (right).

- **Clustering.** The R^2 metric is employed to calculate the correlation matrix between each of the curves. Subsequently, hierarchical clustering is utilized to group curves with similar behaviors. The choice of hierarchical clustering is driven by the advantage of not needing to pre-specify the number of clusters; instead, the clustering depends on the inherent characteristics of the data. Furthermore, it is validated that each cluster maintains a minimum number of observations, ensuring the representativeness of each group. If any cluster fails to meet this requirement, it is merged with the most similar group based on the R^2 metric.
- **Fitting.** The passenger presentation curves for each of those groups are calculated and fitted to gamma probability distributions.
- **Assignment.** To reproduce the observed behavior in the predicted demand, the arrival time is assigned to each passenger by assigning a random value following the distributions extracted, depending on the flight characteristics (destination, hour and weekday).

This process allows the estimation of passenger arrivals at the airport for each passenger detected in the previous step. It is worth noting that this methodology enables the transformation of the aggregated demand generated in the previous step into a synthetic population of individual passengers

This approach has been used to forecast the arrival time to the airport of the predicted demand of August 2019. The results are compared with the actual passenger arrival curves in order to evaluate the performance of the proposed methodology.

The results of this process are presented in Section III.B.

3) *Passenger access mode estimation*

The objective of this process is to predict the transport mode used by each passenger to access the airport. To this end, information regarding the transport modes used by the users, derived from the historical synthetic population of passengers, is used to calibrate a machine learning model able to predict the most likely transport mode for each passenger. The transport mode classes available on this dataset are (i) private bus from tour operators, (ii) public bus, (iii) rental car, (iv) private car and (v) taxi/ride-sharing services. Given the nature of the task as a classification challenge, the chosen algorithm for this process is a random forest classifier.

The historical synthetic population of passengers provides a comprehensive array of features useful to explain passenger behavior regarding the chosen transport mode for accessing the airport. These features encompass factors such as trip purpose (where business passengers often prefer taxis, while leisure passengers lean towards public transport modes), or place of residence (as residents of Palma de Mallorca have the possibility to use private cars for airport access compared to non-residents). However, it is important to note that when calibrating the model for application to the synthetic population

of predicted passengers, only features available in this specific dataset can be used. These features consist on flight-related details and estimated passenger arrival times. Therefore, from the data available, the following features have been derived:

- Destination (Domestic, Schengen, etc.)
- Scheduled time of the flight (0-23)
- Weekday (Mon-Sun)
- Arrival time to the airport prior to the flight departure time in minutes

As the categorical features ('destination' and 'weekday') do not have a large number of categories, both have been encoded using label encoding.

During the training process, the following processes were considered:

- The dataset was divided into a training set and a test set; the split used was 75% of the samples for training and 25% for test.
- Recursive Feature Elimination (RFE) was not implemented this time due to the reduced number of input variables.
- To ensure that the model is able to learn the features of all the classes, it is important that all classes are well represented, as otherwise it may be difficult for the model to accurately understand their behavior and may be neglected in favor of other classes. This is the case of the 'Public bus' class, which represents less than 4% of the total sample. To address this issue, a Random Undersampling method was selected aimed to not only balance all the classes, but also to reduce the number of registers and simplify the training process
- A hyperparameter tuning of the random forest classifier algorithm was applied.
- A k-fold cross validation resampling method was applied, setting 5 as the number of folds.

During the training process, the average validation score for each partition of the data and every combination of hyperparameters was obtained, in order to select the combination that provided the best performance. Finally, a random forest classifier model was fitted on the whole training set using the hyperparameters selected in the previous step. The performance of this final model was then evaluated on the test set, which was set aside in the first step, in order to obtain a final indicator of how the model performs with unseen data.

After calibrating the model, it was used to predict the transport modes used by the forecast passengers for August 2019. The results were compared to the actual data in order to assess both the model's performance and the methodology.

The results obtained are presented in Section III.C.

III. RESULTS

A. Per-flight demand prediction

The methodology described in II.B.1) was used to build a model able to anticipate the passenger demand for scheduled flights. Considering that the final goal is to use data from August 2019 to validate the results, data from flight schedules prior to August 2019 was used to calibrate the models.

After applying the RFE algorithm, the features ‘type’ and ‘country’ were removed, as their inclusion did not increase the model’s performance. This may occur because these features do not provide significant information to predict the target variable, causing the model to disregard them. This is particularly evident in the case of the ‘country’ feature, which the model is likely to be ignoring because there is a more informative feature, ‘airport’, which provides a more detailed level of information. Ultimately, the ‘country’ feature can be understood as clusters of the ‘airport’ feature. Regarding the ‘type’ feature, given that approximately 90% of the flights were regular flights and the difference in occupancy between regular and charter flights is not substantial, this feature did not introduce significant variability to the model.

In Figure 3, the final set of selected variables is presented along with their respective importance to the model. The most relevant features for the model are ‘airline’ and ‘airport’. This is reasonable since different airlines employ aircraft with significantly different passenger capacities; which also depends on the final destination (‘airport’). For example, based on the data used in this study, the average occupancy of domestic destinations was barely 120 passengers, while for international destinations it was almost 165 passengers per flight.

Table I displays presents the performance of the machine learning model during the calibration phase (test results) and when applied to the August 2019 dataset. The performance was evaluated using (i) the Coefficient of determination (R^2), a widely used indicator to measure how well the data fits the regression model, with values closer to 1 indicating higher accuracy; (ii) the Mean Absolute Error (MAE), which measures the average value of the errors in the predictions, irrespective of their direction; and (iii) the Mean Absolute Percentage Error (MAPE) which computes the average of the absolute percentage of the errors to calculate how accurate the predictions were in comparison with the actual values. It can be observed that the model’s performance slightly decreases when applied to the August 2019 dataset, however, the model still demonstrates a satisfactory performance.

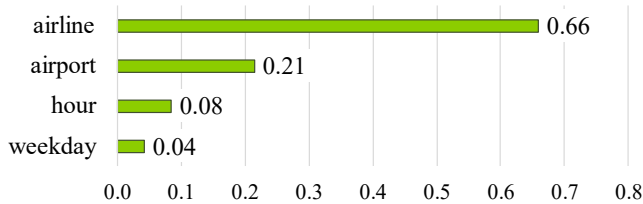


Figure 3. Feature importance in the demand prediction model.

	R^2	MAE	MAPE
Test	0.79	14.88	13.07%
August 2019 data	0.75	15.90	16.35%

In Figure 4 and Figure 5, the correlation between the predicted and the actual demand is presented for both the calibration phase and its application to the August 2019 dataset, respectively. When comparing both figures, it becomes evident that there is a greater dispersion in the prediction results of August 2019, consistent with the performance results presented in Table I.

This process allows an accurate forecasting of the flight demand for August 2019. Figure 6 illustrates a daily comparison between the predicted and actual demand. The average daily error of the model is around 3,3867 passengers per day, which accounts for just 5.6% of the actual PMI daily departing demand. As it can be observed, the model tends to slightly underestimate the demand. This tendency can be attributed to the model being trained on data from months preceding August, which typically experiences higher passenger occupancy levels. This phenomenon may also account for the slight decline in performance noted in Table I. This could be corrected by using data from complete previous years to train the model, allowing it to better capture this behavior and consider potential seasonality of the data. Unfortunately, this data was not available for the study.

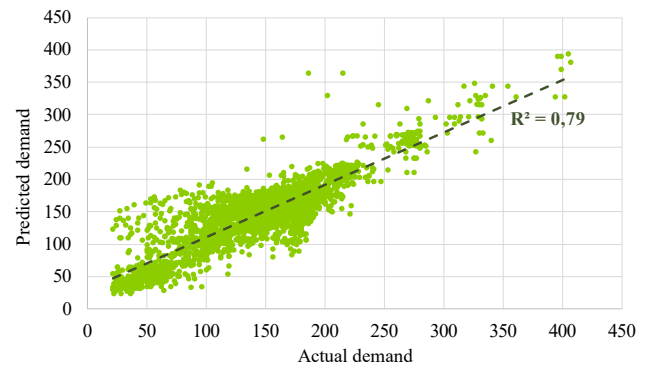


Figure 4. Correlation between actual and predicted flight demand in test

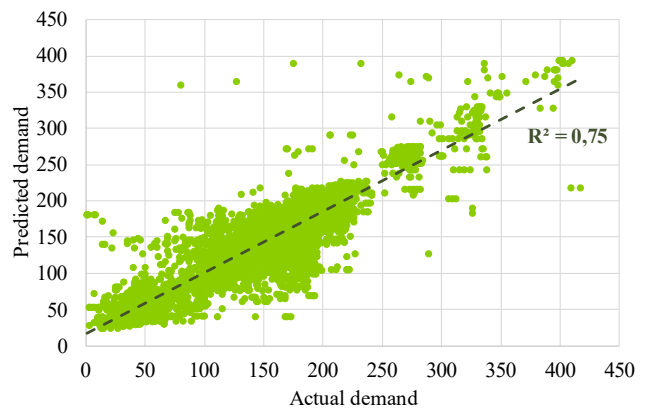


Figure 5. Correlation between actual and predicted flight demand of August 2019

TABLE I. PERFORMANCE OF THE DEMAND PREDICTION MODEL

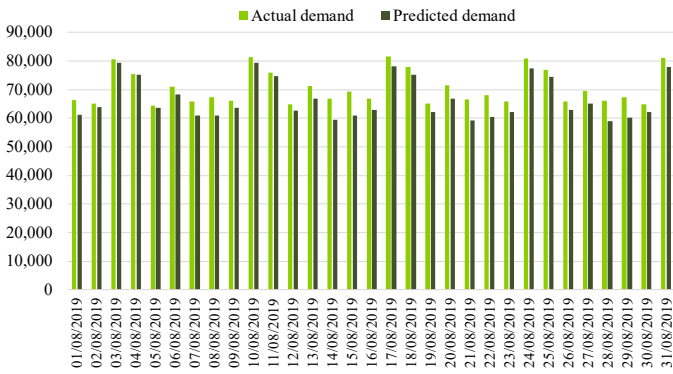


Figure 6. Daily comparison between the actual and predicted demand of August 2019

B. Passenger arrival time prediction

The methodology described in II.B.2) has been applied to the historical synthetic population of passengers from July 2019 aiming at reconstruct the passenger behavior regarding their arrival time to the terminal.

The calculation of the passenger presentation curves for each crossed-segmentation of flight features, considering destination, time of the flight and day of the week, led to a total of 505 different curves. Subsequently, the correlation matrix between each pair of curves was computed and a hierarchical clustering was performed in order to arrange curves with similar behaviors. In Figure 7, the hierarchically-clustered correlation matrix between all the passenger presentation curves is presented. From this information, the clustering algorithm generated 19 clusters. Subsequently, clusters with a low number of observations (the minimum number of observations to consider a representative behavior was set to 500) were removed. After merging these clusters with the most similar ones, a final set of 11 final clusters was selected.

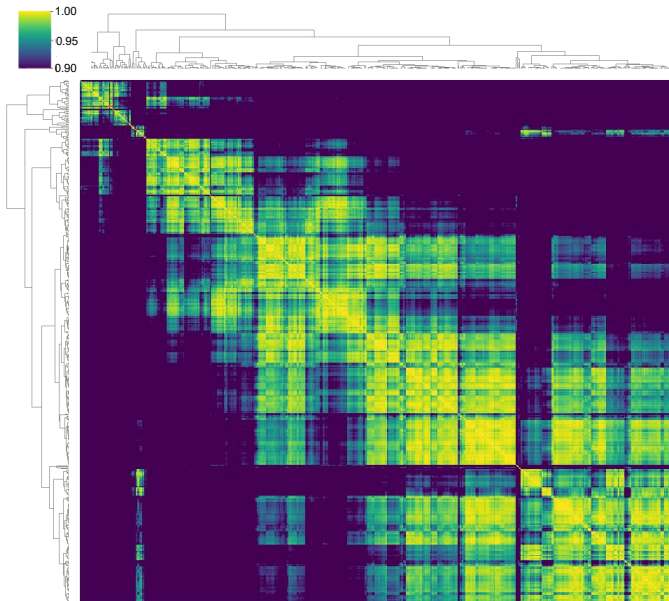


Figure 7. Hierarchically-clustered correlation matrix between the passenger presentation curves.

Figure 8 (a) illustrates an example of the clustering results for Schengen flights, displaying the passenger presentation curves for each cluster included within the ‘Schengen’ class. These curves reveal divergent behaviors among the different clusters. In Figure 8 (b) shows which time of the day and day of the week match to each cluster allowing to better analyze the results. As it can be observed, during the early morning (before 8h), the patterns mainly correspond to clusters 2, 7 and 5, characterized by later passenger arrivals at the airport. During the rest of the day, the observed behavior corresponds to clusters 1, 9 and 10, indicating earlier arrival of passengers to the airport. The time window from 15h to 17h and after 20h corresponds to clusters with even earlier arrivals.

Once the different clusters were selected, the passenger presentation curves of each cluster were approximated by gamma probability functions. Figure 9, presents a comparison between the actual data and its approximation using the gamma function for two of the clusters: one with a strong correlation (cluster 1) and one with a lower correlation (cluster 8). Although noticeable differences can be appreciated in cluster number 8, the approximated curves provide an accurate representation of the real data. For all the clusters obtained, the correlation between the actual passenger presentation curves and their estimation by gamma functions exceeds a R^2 score of 0.98.

After modelling the passenger arrival time behavior, this behavior is applied to the demand predicted in the previous step by assigning each passenger an estimated arrival time based on the gamma functions. The Figure 10 shows the hourly comparison between the actual and the predicted demand for the first complete week of August 2019. The methodology yields satisfactory results by accurately replicating passengers’ behavior, achieving a R^2 metric of 0.96 for the entire month of August.

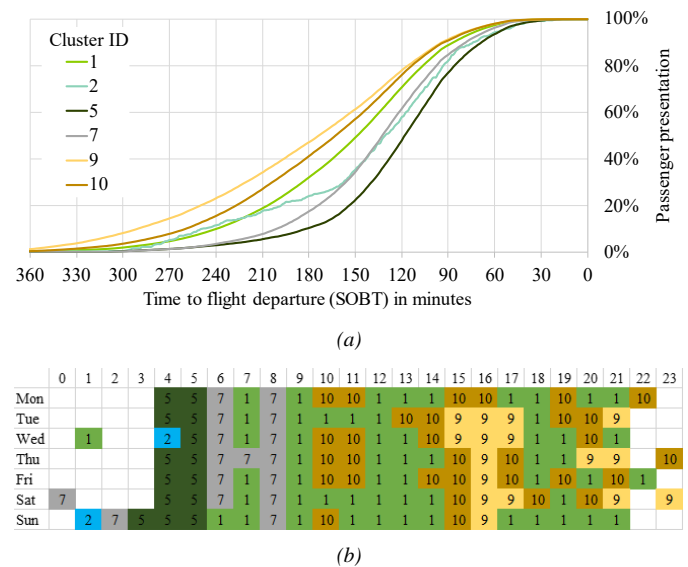


Figure 8. Example of clustering results for Schengen flights. Passenger presentation curves for the clusters present in Schengen flights (a) and its correlation to the time of the day and day of the week (b).

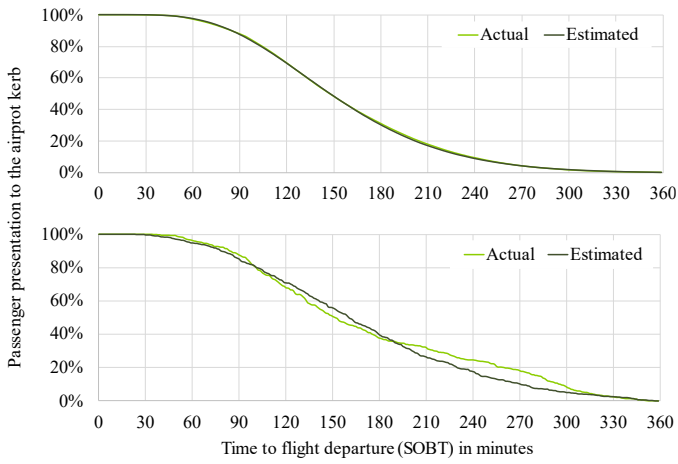


Figure 9. Comparison between the actual passenger presentation curves and their estimation by gamma functions for cluster number 1 (top) and cluster number 8 (bottom)

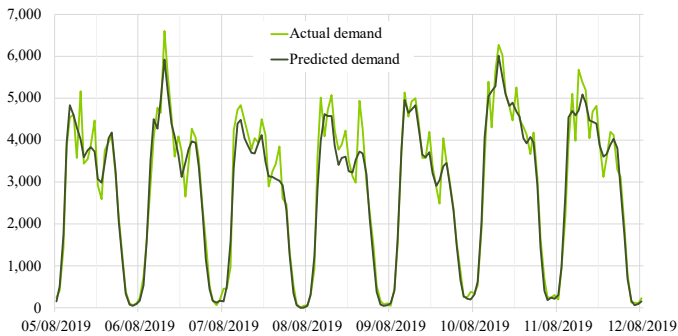


Figure 10. Hourly comparison between the actual and the predicted demand for the first complete week of August 2019.

C. Passenger access mode estimation

The methodology described in II.B.3) has been applied to the historical synthetic population of passengers from July 2019 with the objective of calibrate a machine learning model able to estimate the passengers' transport mode used for accessing the airport.

In this case, the most relevant feature for the model is the 'earliness', as the arrival time at the airport before departure tends to vary depending on the transport mode used to access the airport. Passengers using public transport modes are expected to arrive earlier at the airport as the access leg usually involves additional steps (travelling to the public transport stop, waiting for the airport service, etc.). This leads them to allocate additional buffer time in case any issue appears, such as missing the public service and waiting for the next one. On the other hand, passengers accessing the airport by taxi or private car usually tend to arrive later at the airport since they do not encounter these additional factors.

Table II presents the performance of the machine learning model. Since predicting the transport mode from a set of classes is a classification task, the performance has been measured using the following metrics: (i) precision, which indicates the number of elements correctly classified from the total elements predicted in the class; (ii) recall, which measures

the number of elements correctly classified compared to the actual elements of that class; and (iii) F1-score, which is the harmonic average of precision and recall. Additionally, in Figure 11 it is presented the confusion matrix obtained during test, where it can be assessed the performance of the classification model across the multiple classes by providing a comprehensive breakdown of correct and incorrect predictions for each class.

The random forest classifier demonstrates an excellent performance with a final F1-Score of 0.87, which is fairly satisfactory for a classifier with 5 classes. When analyzing each class separately, it can be appreciated that the performance is similar for all of them, and no class significantly underperforms. In this regard, the most challenging class was the 'public bus' due to the previously mentioned class imbalance, making it difficult for the model to accurately understand its behavior. However, the undersampling method applied seems to work satisfactorily, as the classification results for this class are not significantly different from the rest of classes.

This model was then applied to estimate the access mode to the airport of the passengers forecast in the previous step. The modal share for the complete month of August 2019 is presented in Table III, where it is also compared to the actual August 2019 data. The correlation between both is noticeably high, which indicates that the model is able to reproduce the actual passenger behavior.

TABLE II. PERFORMANCE OF THE TRANSPORT MODE ESTIMATION MODEL

	Precision	Recall	F1-Score
Private bus	0.89	0.85	0.87
Private car	0.86	0.88	0.87
Public bus	0.77	0.92	0.84
Rental car	0.91	0.87	0.89
Taxi/ride-sharing services	0.84	0.90	0.87
Average	0.85	0.88	0.87

True label	Predicted label				
	Private bus	Private car	Public bus	Rental car	Taxi/ride-sharing services
Private bus	155,940	6,965	2,902	7,754	9,522
Private car	5,324	111,417	1,082	4,985	4,478
Public bus	674	376	23,050	429	449
Rental car	7,513	7,680	1,948	162,308	6,930
Taxi/ride-sharing services	5,320	3,125	921	3,460	115,256

Figure 11. Confusion matrix obtained during test. Each cell contains the count of elements based on the model's predictions and the true class labels.

TABLE III. ACTUAL VS PREDICTED MODAL SHARE FOR AUGUST 2019

	Actual	Predicted
Private bus	28.1%	25.5%
Private car	19.6%	24.7%
Public bus	3.9%	4.9%
Rental car	28.7%	25.3%
Taxi/Ride-sharing services	19.7%	19.6%

In Figure 12, the hourly comparison between the actual and forecast passenger flows segmented by transport mode is presented for the first complete week of August 2019. It is noticeable that the forecast passenger flows fairly reproduce the actual flows, exhibiting a satisfactory level of correlation. When examining the different modes individually, it becomes apparent that the ‘Public bus’ curve is the one presenting major difficulty to fit the actual data. The R^2 correlation obtained for each transport mode for the entire month of August 2019 is also included on the figure. It can be noticed that the ‘Public bus’ class exhibits a slightly lower correlation compared to the others, which is likely to be due to the difficulties of the model to distinguish that class due to imbalances during the training. Nevertheless, the achieved correlation is still satisfactory.

IV. DISCUSSION

The proposed methodology extends previous research efforts aimed at characterizing passengers’ air travel patterns. It

uses historical descriptive data to build predictive models able to forecast the passengers’ behavior. This approach not only considers historical behaviors, but also allows us to consider additional exogenous variables, such as flight schedules, as used in this study, or other potential features that could be explored in future research, such as the weather forecast, which can affect the passenger behavior when accessing the airport.

The methodology used to estimate the passenger demand, involving the calibration of a machine learning model, has shown promising results. The inclusion of additional data from other years could potentially enhance the model’s performance by enabling it to better capture patterns across different months. However, the obtained performance was more than sufficient to the purpose of this study. Future research efforts could focus on enhancing model performance by assessing different machine learning algorithms such as XGBoost, LightGBM or neural networks. Future research also should focus on extending this methodology to various types of airports. It must be remarked that this study was conducted for Palma de Mallorca Airport, a highly tourist-focused airport, during summer season, when high flight occupancy is expected. Replicating this methodology in different airport types could help determine whether the models can adapt to more volatile behaviors. This includes regional airports with significantly lower flight volumes, where occupancy is expected to vary widely from one airport to another; or large international hubs where connecting flights represent a non-negligible part of passenger traffic.

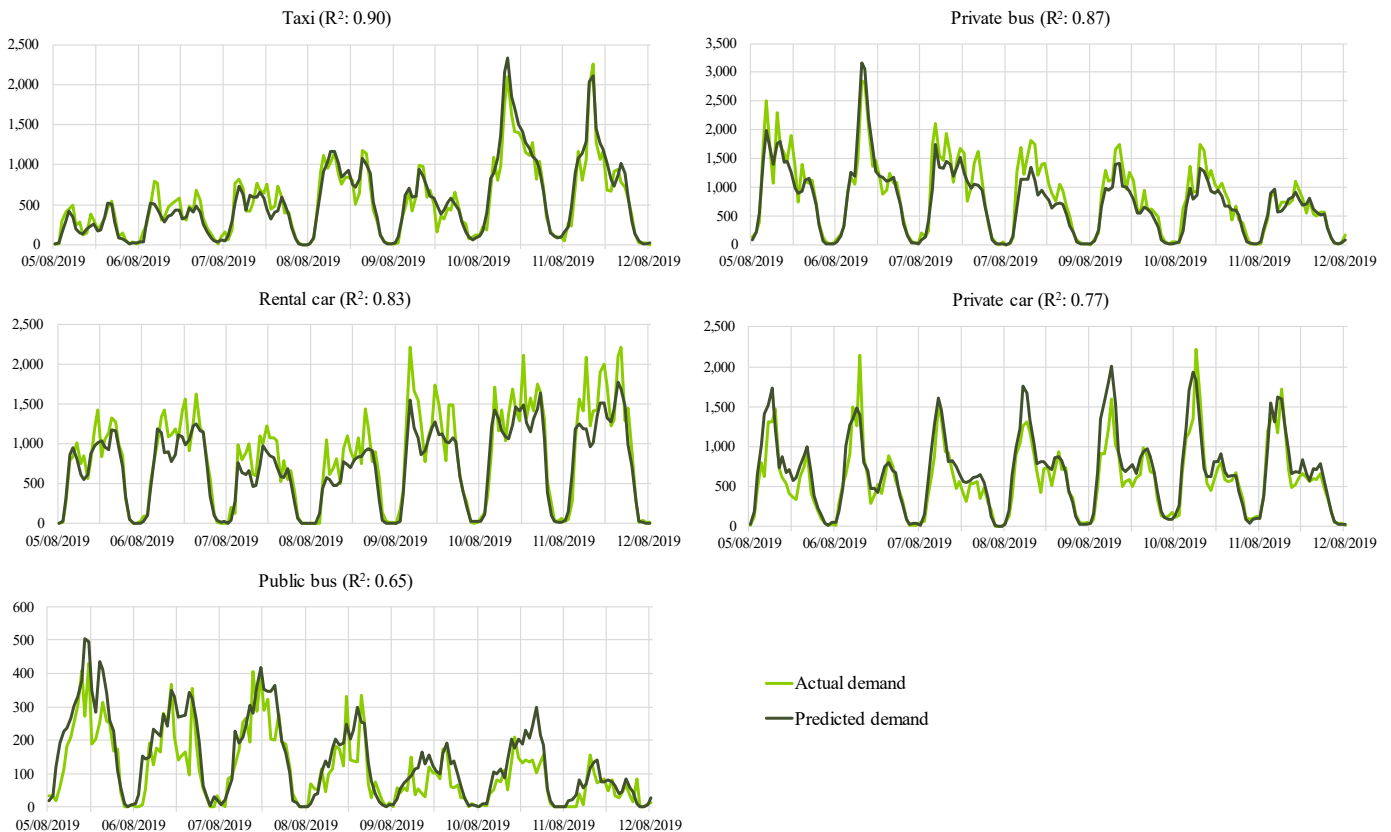


Figure 12. Hourly comparison between the actual and the predicted demand by mode for the first complete week of August 2019.

The proposed approach for estimating passenger arrival times at the airport has yielded two valuable insights. Firstly, the passenger arrival at the airport before departure can be accurately estimated using probability distributions. Secondly, this behavior is highly influenced by the passenger's final destination, time of the flight and day of the week. The methodology developed in this study aims to cluster similar behaviors considering these factors. However, the final number of clusters generated may depend on the level of detail the user desires or even the specific airport of application. This is because patterns observed are likely to vary significantly between airports. Therefore, an historical data exploration is always required in order to adapt the methodology to the patterns and behaviors of passengers in the target airport.

The estimation of the transport mode used by passengers to access the airport was a challenging task. Typically, the calibration of machine learning classification algorithms requires a higher number of features in order to enable the models to effectively learn patterns. Additionally, the class distribution was remarkably imbalanced, further complicating the problem. These main issues were addressed by introducing more complexity to the model to compensate for the lack of features, while also being careful to avoid overfitting. Data resampling algorithms were implemented to assist the model in identifying the key patterns. This resulted in favorable performance, enabling accurate forecasting of passenger access mode. Future research could, again, aim to improve the performance of the model by evaluating various machine learning classification algorithms. Additionally, future studies should focus on applying this methodology to different airports to assess its robustness. Palma de Mallorca presents a demanding scenario, with a significant volume of rental cars and private buses from tour operators, even though access to the airport is primarily by road. Implementing this methodology at other types of airports with a wider range of modes such as, underground, commuter, or train could also be a challenge for model training.

The methodology presented in this paper not only facilitates the detailed forecasting of departing passenger flows under normal conditions but also allows for different what-if analyses. These analyses can assess how passenger flows would change if specific initial conditions were modified, given that each forecast passenger is matched to an existing flight. For example, it could analyze how flight delays would affect passenger flows if passengers were informed on time before starting their journey to the airport. It could also help to

identify which flights are more likely to be affected by a ground surface disruption.

In summary, the methodology proposed in this paper enables the detailed forecasting of departing passenger flows and the identification of the mode used to access the airport, two indicators which are undeniably useful for the airports. This information can be used by airport operators to optimize resource allocation based on an up-to-date view of the upcoming passenger flows. It can also be used to coordinate with the ground transport providers and authorities in order to optimize the ground transport system.

ACKNOWLEDGMENT

The results presented in this study are part of the TravellInt project, a project funded by the European Union under the Next Generation-EU program. The project has been supported by the Spanish Ministry of Economic Affairs and Digital Transformation, through Red.es.

These results build on the outcomes achieved during the IMHOTEP project. The project received funding from the SESAR Joint Undertaking under grant agreement No 891287 under European Union's Horizon 2020 research and innovation programme.

REFERENCES

- [1] European Commission. 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 – Maintaining global leadership and serving society's needs, 2011.
- [3] P. García-Albertos, et al. "Big data analytics for a passenger-centric ATM system: A case study of door-to-door intermodal passenger journey inferred from mobile phone data", SIDS 2016.
- [4] P. García-Albertos, et al. "Understanding door-to-door travel times from opportunistically collected mobile phone records. A case study of Spanish airports", SESAR Innovation Days, 2017.
- [5] J. Blasco, et al. (2022). Data fusion for the analysis of air travel behavior: Application to Palma de Mallorca Airport. Proceedings of the 12th SESAR Innovation Days. <https://shorturl.at/uAHRX>
- [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, et al. "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] J. Burrieza-Galán, et al. (2022). "A methodology for understanding passenger flows combining mobile phone records and airport surveys". Journal of Air Transport Management..
- [10] A. Gregg, et al. (2022). "Airport accessibility surveys and mobile phone records data fusion for the analysis of air travel behaviour". 12th International Conference on Transport Survey Methods. In press