Understanding airline behaviours on trajectory choices through opendata

P. Bassi, T. Bolić, L. Castelli, A. De Lorenzo
Dipartimento di Ingegneria e Architettura, Università degli Studi di Trieste, Italy
paola.bassi@dia.units.it, tbolic@units.it, castelli@units.it, andrea.delorenzo@units.it

In the air transport industry there is a widespread interest in understanding the factors influencing the trajectory choice between any given origin and destination (O/D) airport. Some studies already exist (see, e.g., Delgado, 2015) but are mostly based on proprietary data. Through a historical (data-driven) analysis of 3D trajectories of the European airspace, this paper investigates the potential of using opendata to identify airline/flight choices and behaviours. Relying on a dataset downloaded from OpenSky and composed of 3.5 million flights spanning from May to September 2017, trajectories are clustered according to the Hausdorff distance through a Density-Based Spatial Clustering of Applications (DBSCAN) algorithm. In addition, and differently from most studies available in the literature, here we also consider the flight altitude (3D trajectories) as it plays an important role in proper route clustering. For example, an aircraft type A and aircraft type B can fly exactly the same 2D trajectory (latitude, longitude), but the type A cannot go over an altitude X, while type B can. Thus, when taking 3D environment into account the trajectories of these two aircraft types should belong to different clusters. This requires a slight modification of the algorithm as the horizontal distance between the elements is measured in kilometres, while the difference in altitude is much lower, measured in meters (e.g. the distance between two flight levels is 1000 ft which is about 300 m). Another feature of our clustering procedure is the empirical identification of the departure and arrival airport of each trajectory (this piece of information is not available in the ADS-B messages), leading to about 2700 Origin-Destination (O/D) pairs.

Our preliminary results show that trajectory choices may indeed depend on airline characteristics. Relying on Pearson's χ^2 test (the lower the p-value, the higher the correlation), we see that in the 27.8% of the cases aircraft types and trajectories are statistically and significantly correlated (p-value < 0.1). Such percentage remains higher than 20% also for stricter p-values (24.7% with p-value < 0.05 and 22.5% for p-value < 0.02). Aircraft types are clustered according to their maximum take-off weight, as suggested by Cook and Tanner (2015). Similarly, considering the relationship between trajectories and individual airlines, we observe a significant correlation in 15.6% of the O/D pairs (13.7% with p-value < 0.05 and 11.9 with p-value < 0.02). Since in many cases only one cluster exists (this also happens because of the short duration of most European flights: the longer the flight, the larger the number of different trajectories), our findings may suggest that, when there are different trajectory options, these are likely to be chosen according to some specific explanatory variables.

Future research will try to incorporate other determinants of the trajectory choices. The major breakthrough in the use of opendata would be the capability to move from 3D to 4D trajectories, namely including the temporal dimension to analyse delays and turnaround times.

References

Cook, A. J., & Tanner, G. (2015). European airline delay cost reference values. EUROCONTROL Performance Review Unit.

Delgado L. (2015) European route choice determinants. 11th USA/Europe Air Traffic Management Research and Development Seminar, Lisbon, 23 -26 June 2015