TOPOLOGICAL DATA ANALYSIS IN ATM:
THE SHAPE OF BIG FLIGHT DATA SETS

Manuel Cuerno*, Luis Guijarro*, Rosa María Arnaldo** & Fernando Gómez**

Universidad Autónoma of Madrid* (UAM) & Universidad Politécnica of Madrid** (UPM)

INTRODUCTION

Airports and ATM Systems are complex sociotechnical structures that are highly interdependent, making them difficult to analyze and understand. The interconnectedness, interdependences and complexity of the system is reflected in the amount of data generated by its operation. Flight trajectory data offer a big potential to grasp the features and behaviour of such intricate system that often produces complex high-dimensional sparse data sets; moreover, they are affected by inconsistencies, errors, and high levels of variability. These sets are difficult to analyse due to several reasons: *multiple variables, continuous data, multi-level interactions, dynamic changes* and *high dimensionality*.

TDA offers a powerful tool for analysing complex and high-dimensional aviation data sets. By identifying topological features and patterns, TDA can reveal hidden relationships and help airlines and airports make better decisions about flight scheduling, maintenance, and safety. In this work we propose to use TDA to **analyse flight trajectory data**, **identify patterns in the movement of aircrafts**, and **determine the relationships between different variables** involved in the spatial and temporal flight trajectories and delays. These strategies will highlight common patterns and anomalies in airport operation and congestion, and help to recognise underlying causes of delays and develop more effective strategies for reducing them.

DEVIATION DISTANCE

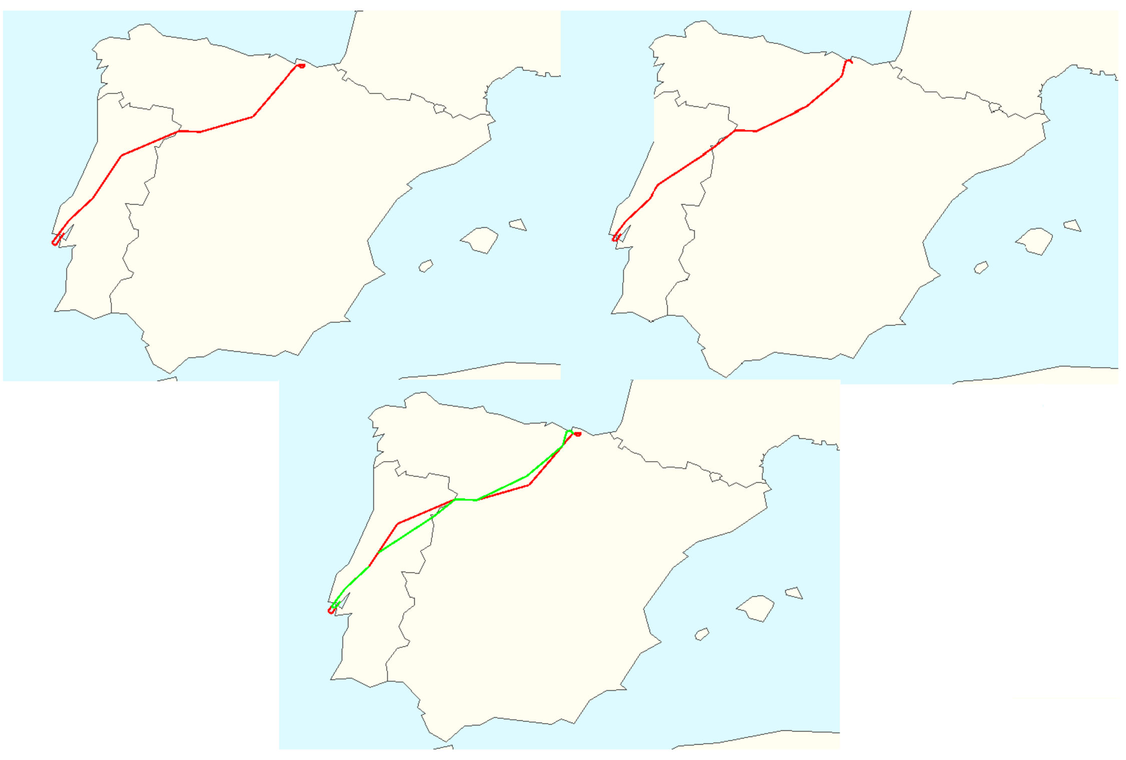

Fig. 1: Planned and real trajectories from Santander to Lisbon

Each trajectory is given as a set of 4–dimensional ordered points which have the following configuration:

$(t =$ time, $L =$ latitude, $l =$ longitude, $a =$ altitude (km)$)$.

We define a **distance** between (t_0, L_0, l_0, a_o) and (t_1, L_1, l_1, a_1) where $t_0 = t_1 = T$ as

$$\mathrm{d}_T((L_0, l_0, a_o),(L_1, l_1, a_1)) := \sqrt{\mathrm{d}_H((L_0, l_0),(L_1, l_1))^2 + |a_0 - a_1|^2},$$

where d_H is well-known *haversine* (or *great circle*) distance used for many geolocalization purposes such as mobile apps.

After we normalize the time series of both planned and real trajectories $pT = \{(\widehat{t}_l, L_l, l_l, a_l)\}_{l\in K}$ and $rT = \{(\widehat{t}_l, \widetilde{L}_l, \widetilde{l}_l, \widetilde{a}_l)\}_{l\in K}$ with $K = [0, k_f]$, we define the **deviation distance** as

$$\mathrm{d}_{\mathsf{Dev}}(pT, rT) := \sum_{l=0}^{k_f} \mathrm{d}_{\widehat{t}_l}((L_l, l_l, a_l),(\widetilde{L}_l, \widetilde{l}_l, \widetilde{a}_l)).$$

FOOTPRINT OF AN AIRPORT

In order to apply TDA techniques to a given problem, we require a point cloud to work with. We are going to focus on the *flights that arrive and departure from a certain airport*. We want to investigate how deviations and delays affect its performance.

Suppose that this airport has N flights per day. As discussed in the previous subsection, we compute the distances between the planned and actual trajectories of each flight, resulting in a list of ten distances denoted as $\{d_1, \ldots, d_N\}$. Then, we can define our point cloud as follows:

$$\mathcal{V} = \{(\widetilde{d}_i, r_i)\}_{i\in[1,N]} \subset \mathbb{R}^2,$$

where $\widetilde{d}_i = (-1)^p d_i$ with $p = 0$ if the flight arrived late or $p = 1$ if the flight arrived on time or sooner, and r_i corresponds to the quantity $\widetilde{t}_{i_f} - t_{j_f}$, i.e., the difference between the time the flight landed and the expected one.

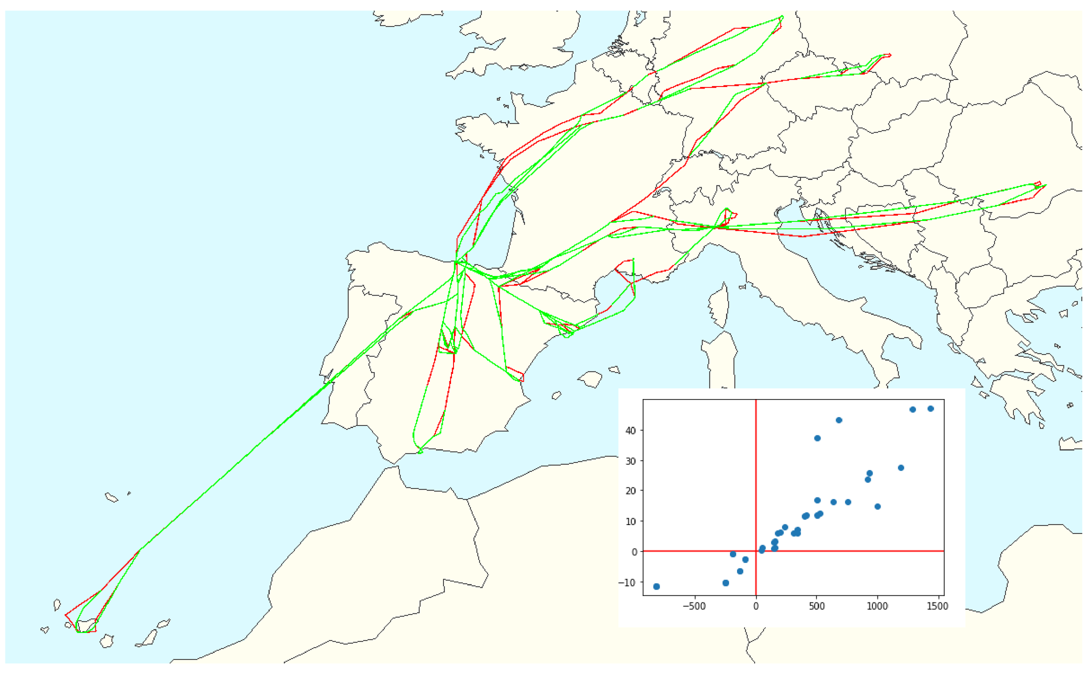

Fig. 2: Point cloud of Santander's airport on the 29th June 2019

Once we obtain the point cloud, we compute the 0–persistence diagram.

We can do this for a sequence of days, obtaining a family $\mathcal{D}$ of persistence diagrams. Following the theory of persistence landscapes [1, 2], we transform $\mathcal{D}$ into a family of persistence landscapes $\mathcal{P}$ and we compute its **average persistence landscape**.

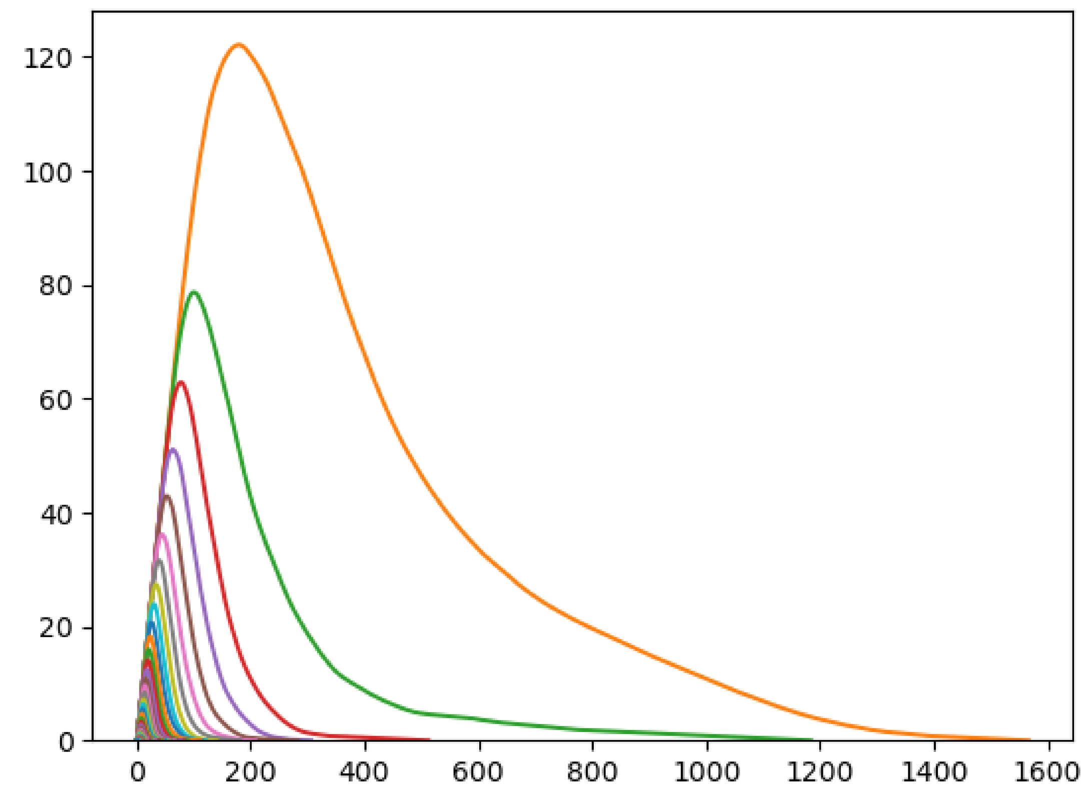

Fig. 3: Average Persistence Landscape of Santander's airport in the Summer Season of 2018

ATM: Air Traffic Management

EXPERIMENT AND CONCLUSIONS

The purpose of this work is to introduce the concepts of Topological Data Analysis (TDA) and explore their application in the context of ATM. We aim to demonstrate the efficacy of TDA through an analysis of real-world data, specifically the Spanish network of airports during the Summer Season of 2018, as clasified by AENA (a Spanish public company incorporated as a public limited company that manages general interest airports in Spain).

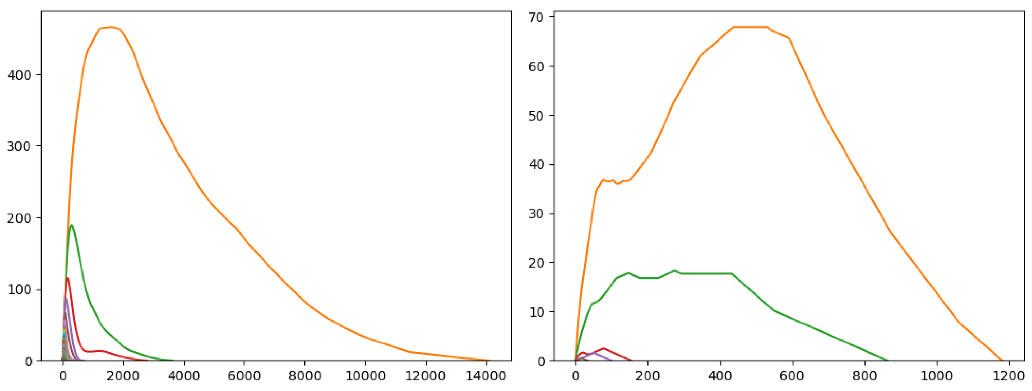

Fig. 4: On the left: Madrid-Barajas. On the right: Huesca. Both on the summer season of 2018

The results of the analysis show how different airport groups follow a sort of cluster, even though there are few in number in each group. It also allows the identification of airports that clearly differ from their preassigned group (such as Zaragoza's airport, which differs from the rest of the airports of Group 2). It also helps detect when an airport is isolated and far from any other airport in the network, such as Adolfo Suarez Madrid-Barajas and Josep Tarradellas Barcelona-El Prat's airports are from any other airport in Spain.

To the best of our knowledge, although there have been some generic works outlining the possible application of TDA in aviation [4], no rigorous one has been performed by applying this method to a large amount of aircraft trajectory data. Our attempt tries to anticipate and identify deviation and anomalies in aircraft space/time trajectories, infer patterns of behavior at different airports, and classify and characterize airports depending on the distribution of their daily flights via trajectory deviation and delay.

My personal webpage:

References

[1] Peter Bubenik. "Statistical topological data analysis using persistence landscapes". In: J. Mach. Learn. Res. 16 (2015).

[2] Peter Bubenik. "The persistence landscape and some of its properties". In: Topological data analysis—the Abel Symposium 2018. Vol. 15. Abel Symp. Springer, 2020.

[3] Manuel Cuerno et al. "Topological Data Analysis in ATM: the shape of big flight data sets". In: (2023). arXiv: 2304.08906.

[4] Max Z. Li et al. "Topological data analysis for aviation applications". In: Transportation Research Part E 128 (2019), pp. 149–174.

Contact: manuel.mellado@uam.es