Statistical modelling of aircraft trajectories

Rémi PERRICHON

Advisors: Thierry KLEIN (ENAC) & Xavier GENDRE (ISAE)

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

ENAC

Context

One of the main goals of the DSNA is to modernize air navigation. To achieve so, the French Civil Aviation Authority (DGAC) has been designated project owner of SALTO. This tool is currently used at the strategic, pre-tactical and tactical levels of airspace management. Yet, the tool does not include weather data. Matching trajectories with the experienced weather could be a great starting point to perform some spatio-temporal modelling. Despite an unquestionable efficiency in predicting or classifying trajectory data, deep learning methods are not adapted to deal with classical problems in statistics such as testing. The thesis aims at exploring methods in the fields of spatial and spatio-temporal statistics to better understand how complex data (weather, airspace configurations) are associated with some features of trajectories (delays for instance).

Data

For the moment, two main data sources have been taken into account:

- **(Trajectory data) R&D data from Eurocontrol**
  The R&D data archive contains more than four years of data, that is to say to more than 14 million flights as of April 2021. The data are collected from all commercial flights operating in and over Europe.

  Data are available for 4 months each year: March, June, September and December. Data include the last-filed flight plans and the actual route, the airspace and the route network that was in place at that time.
  For the moment, the following flights are taken into account:
  - Toulouse Blagnac to Paris Orly: 12,400 flights (2015 to 2018)
  - Paris Orly to Toulouse Blagnac: 12,406 flights (2015 to 2018)
  - Brussels to Frankfurt: 3,879 flights (2015 to 2018)
  - Frankfurt to Brussels: 3,819 flights (2015 to 2018)
  - Madrid to Brussels: 4,895 flights (2015 to 2018)
  - Brussels to Madrid : 4,892 flights (2015 to 2018)
  - London Heathrow to Madrid : 6,057 flights (2015 to 2018)
  - Madrid to London Heathrow : 6,053 flights (2015 to 2018)

- **(Weather data) ERA 5 data**
  ERA5 is the fifth generation European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis for the global climate and weather. I focus on ERA5 hourly data on pressure levels from 1979 to present.
  Out of 37 available pressure levels (ranging from 1,000 hPa to 1 hPa), only 23 (from 1,000 hPa to 200 hPa) are kept.

References

[1] Cressie, Noel A. C., and Christopher K. Wikle. Statistics for Spatio-Temporal Data. Hoboken, N.J.: Wiley, 2011.

[2] Chris J. Needham and Roger D. Boyle. Performance evaluation metrics and statistics for positional tracker evaluation. In James L. Crowley, Justus H. Piater, Markus Vincze, and Lucas Paletta, editors, Computer Vision Systems, pages 278–289, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.

Objectives

Trajectories are spatio-temporal data. Nearby (in space and time) observations tend to be more alike than those far apart: the independence assumption is violated. The fact that trajectories are 4D (three space dimensions, one time dimension) and have an irregular sampling (in both space and time) is the reason why specific statistical tools should be used **[1]**, even to answer simple questions such as:

Is there something as a typical trajectory for a given air link? How to detect abnormal trajectories? How to group similar trajectories? How to model the experienced weather along a specific trajectory?

Answering these questions relies on creating a suited distance for trajectory data **[2]**. To compare trajectories, there are many things to take into account: the departure point, the arrival point, but also the duration, the global shape and the associated time window. Many distances already exist. Yet, they badly take into account irregular sampling, and trajectories of different durations.

The main objective of the thesis is to develop a rigorous approach to better understand how complex data (weather, airspace configurations) are associated with some features of trajectories (delays for instance).

0. Data cleaning

Data cleaning is the mandatory step before doing any visualization

1. Matching trajectories and weather

First, I want to associate a weather value (for example the wind speed) to each point of a flight (one-to-one matching). To associate each point of the trajectory to a weather value, I need to match along three dimensions:

- Closest pressure level (altitude)
- Nearest time
- Closest position (longitude, latitude)

This first step has already been done.

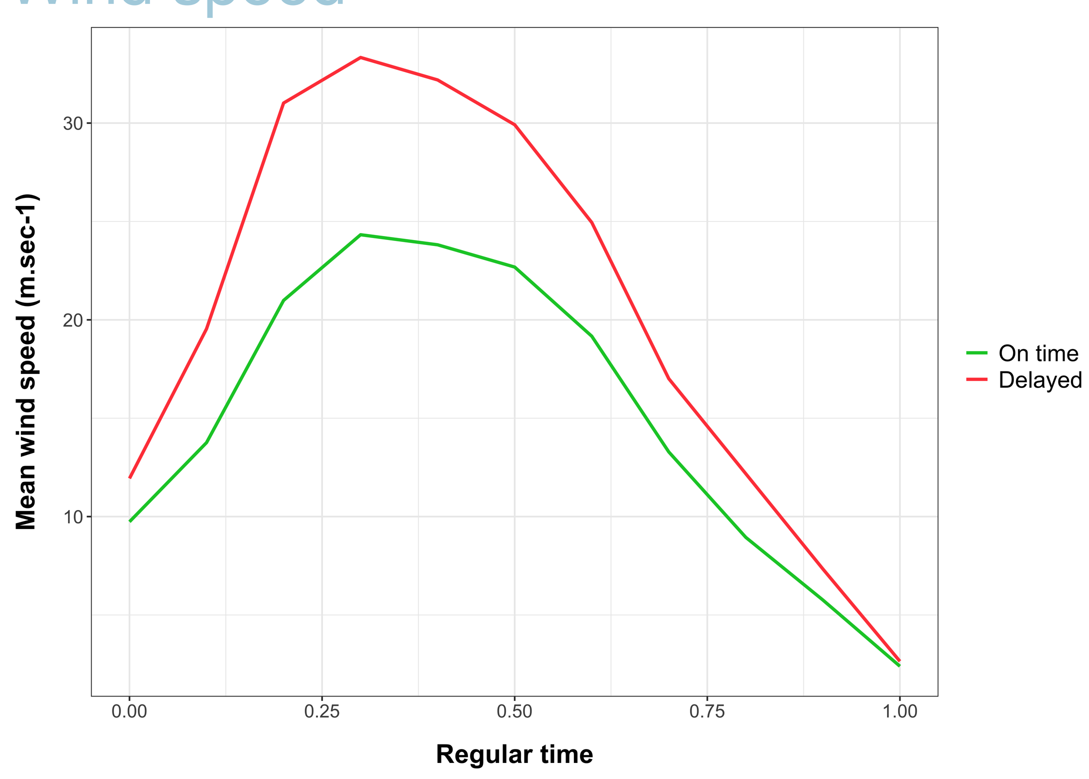
Second, I want to have a summary of the weather experienced by the plane. In this case, I am interested in building an average experienced weather based on a good distance (for instance, using Wasserstein barycenters).

2. Studying the impact of the wind on delays

First, we are interested in delays accumulated after the take-off, that is to say during the flight. We hypothesize that the delayed flights experience a different wind compared to flights that are on schedule. For a given air link and a given year (for instance flights from Toulouse Blagnac to Paris Orly in 2015), the 20% latest flights relative to the mean duration are said to be delayed. The duration of a flight is scaled between 0 and 1 for the flights to be comparable. The experienced wind during a flight is viewed as a stochastic process $Z = \{Z(t), t \in [0, 1]\}$ for which we observe n values of the wind speed $z(t_1), ..., z(t_n)$. A linear interpolation is made to define the empirical mean wind speed with values: $\bar{z}(t) = \frac{1}{N}\sum_{i=1}^{N} z_i(t)$, N being the number of flights with $t \in \{t_1, ..., t_n\}$.

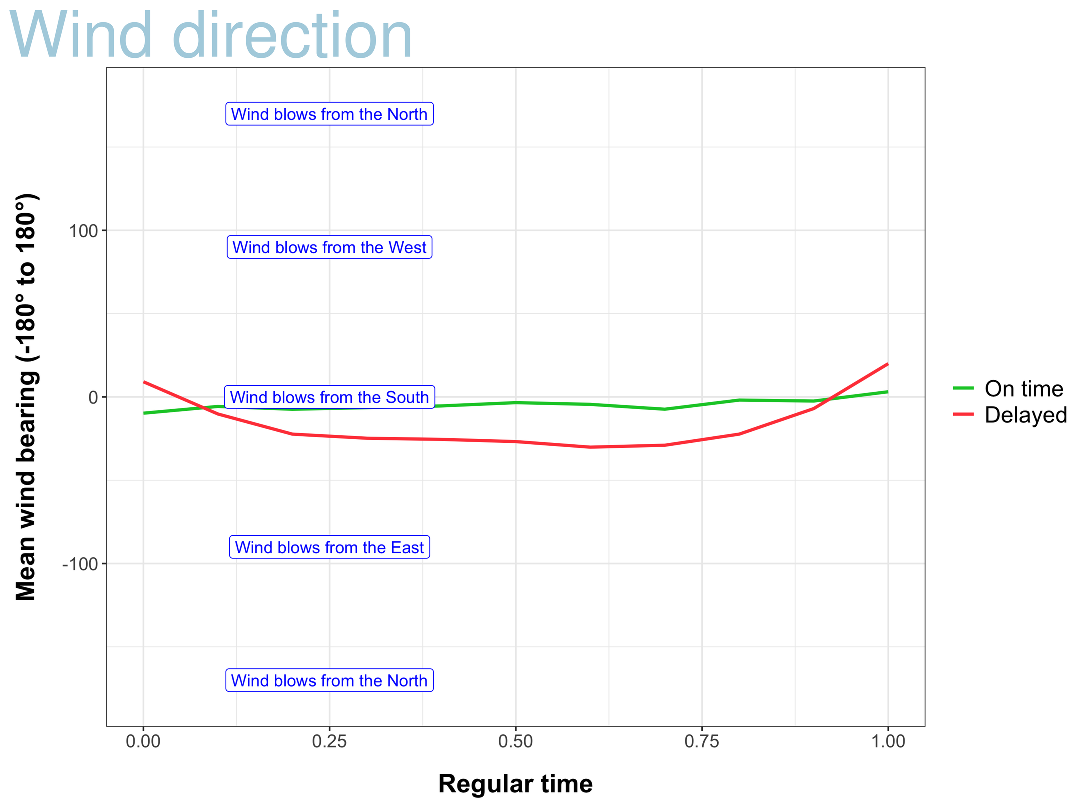
First results - wind and delays

Wind speed



In 2015 for the flights departing from Toulouse Blagnac and landing at Paris Orly, the average experienced wind speed is higher for each part of the flight profile for the delayed aircrafts. This result remains the same for other years and can be generalized to other airports. A formal statistical test is to be done.

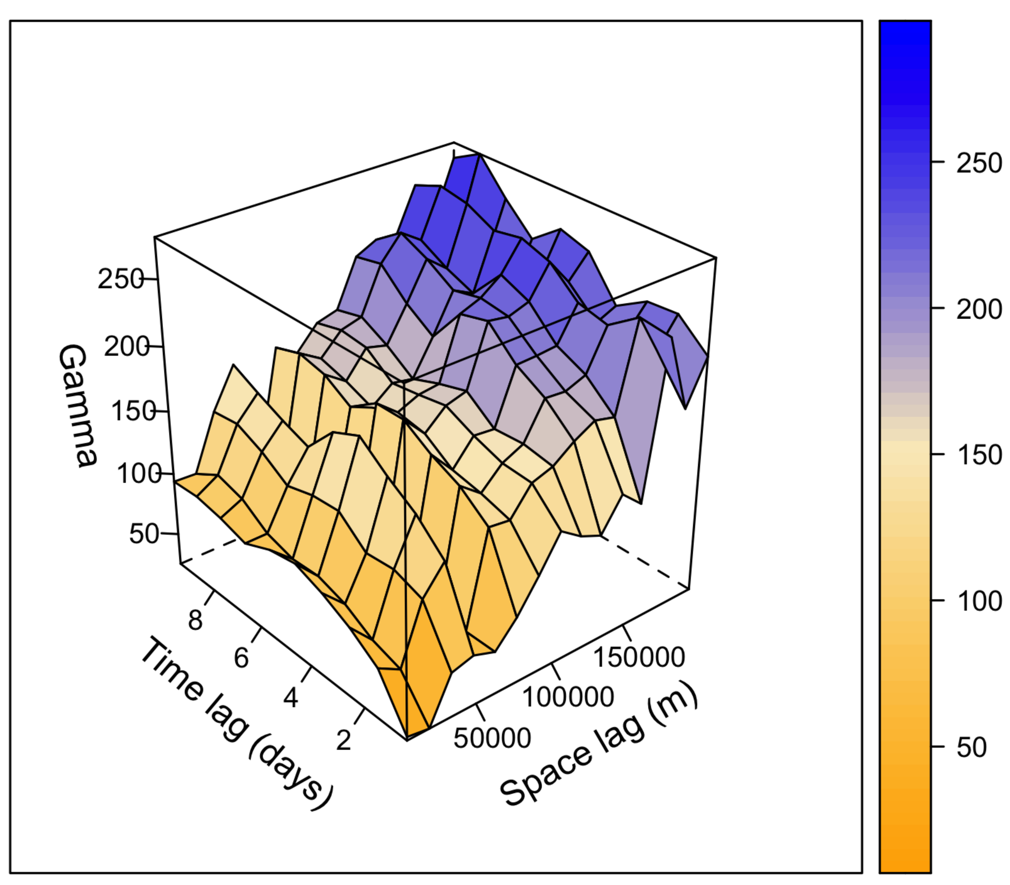
Wind direction



In 2015 for the flights departing from Toulouse Blagnac and landing at Paris Orly, the average direction of the wind is from South to North for the two groups. Yet, the direction seems less favorable for delayed flights knowing that the perfect angle between Toulouse and Paris is 7.5° (perfect tail wind for a straight flight path). A formal test is to be done.

Towards spatio-temporal statistics

In the spatio-temporal framework, the points of the flights are viewed as the realization of a spatio-temporal process $Z = \{Z(s,t) : s \subset \mathbb{R}^3, t \in T\}$. In spatio-temporal statistics, we are often interested in estimating the semivariogram γ. It gives some information on the dependence of the wind speed values in space and time. The following plot shows that the dependence in time is relevant at the daily time scale. The distance profile seems to be the same as time goes. The color scale gives the value of γ.



Contact: remi.perrichon@enac.fr