A Framework to Evaluate Aircraft Trajectory Generation Methods

Xavier Olive   Junzi Sun   Mayara Condé Rocha Murça   Timothé Krauth

ATM Seminar 2021

ONERA (France)   TU Delft (The Netherlands)   ITA (Brazil)   ZHAW (Switzerland)

The trajectory generation problem

Construct a trajectory reflecting:

- the **physical behaviour** of the aircraft;
- the **statistical properties** of the spatio-temporal bounds.

Subject to performance optimisation (fuel consumption, runway throughput, deconfliction, etc.)

Many applications, including

- simulations,
- workload predictions,
- safety analysis,
- emergency trajectories,
- etc.

Outline

Main question:

- How to evaluate/compare different trajectory generation methods?

Outline

- Expectations around trajectory generation methods
- Description of the evaluation framework
- Case studies
- Discussion

Framework

Literature Review

Current state of the art heavily depends on the application.

Methods can be split into two categories:

- **Model-driven** trajectory generation methods
  - based on kinematics, aerodynamics (optimise fuel consumption)
  - based on aircraft performance models (BADA, OpenAP)
- **Data-driven** trajectory generation methods
  - statistical approaches, unsupervised ML
  - importance of dimensionality reduction
  - importance of distance metrics

Expectations about generated trajectories

	operational realism	physical realism
human factors	high	average
ATC training	high	average
safety analyses	focus on outliers	average
emergency situations	low	high
flow management	high	low

It often boils down to finding the proper **balance between operational and physical realism**.

Proposed metrics (1/2)

- **Expert evaluation**

  Assess the operational relevance of trajectories. Show real and generated trajectories to operational people and ask if situations happened in real-life.

  Metric: f1-score

- **Similarity evaluation**

  Has this situation already been encountered in a dataset. Compute the minimal distance between the generated trajectory t and all samples t_r in the reference dataset.

  Metric:

$$s(t) = \min_{t_r \in \text{history}} d(t, t_r)$$

Proposed metrics (2/2)

- **Statistical evaluation**

  Focus on the statistical properties of a set of generated trajectories.

  Metric: Kolmogorov Smirnov test, KL divergence, Kolmogorov, Wasserstein distance, etc.

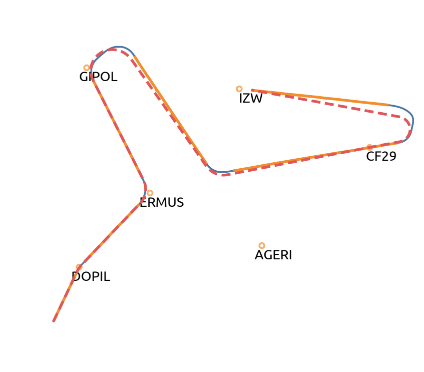
- **Physical evaluation**

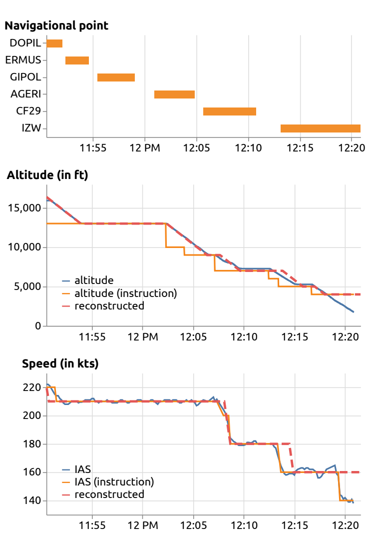
  Ensure that generated trajectories are reasonable wrt physics and aircraft performance?

  Metric (proposition):

  1. Extract features from the generated trajectory
  2. Replay the trajectory in a simulator (BlueSky)
  3. Compare the replay to the generated trajectory

Feature extraction and trajectory replay





Guidelines

- **State** the context of the application
- **Rate** priorities in terms of expectations.
- **Favour** metrics which **do not serve as base of the implementation** of the assessed generation metrics
- **Apply** metrics to real trajectories as well

Case studies

Description of the problem

- Research interest around collision risk modelling in TMA
- Dataset of trajectories landing at Zurich airport `LSZH` (October and November 2019)
- Data freely available from:
    - The OpenSky Network
      https://www.opensky-network.org
    - https://doi.org/10.6084/m9.figshare.11406735.v1
    - Python `traffic` library
      https://www.github.com/xoolive/traffic

```
from traffic.data.datasets import landing_zurich_2019
```

Gaussian Mixture Model-based trajectory generation

Murça and Oliveira, 2020

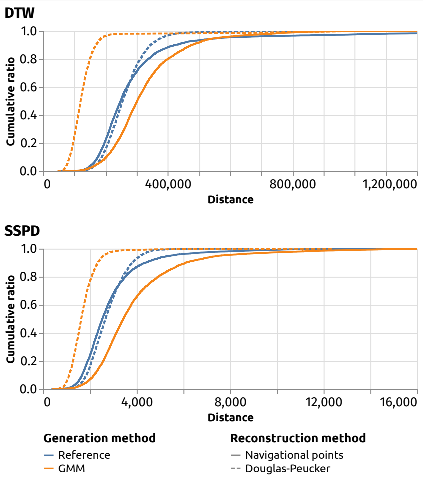
General principle

- Identify major arrival trajectory patterns
- Model each pattern with a Gaussian Mixture Model
- Sample patterns, then sample trajectories within patterns

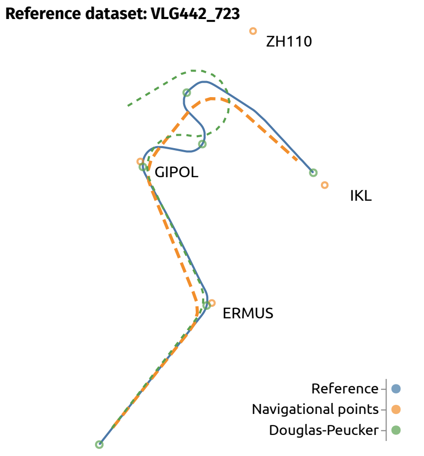
Physical evaluation

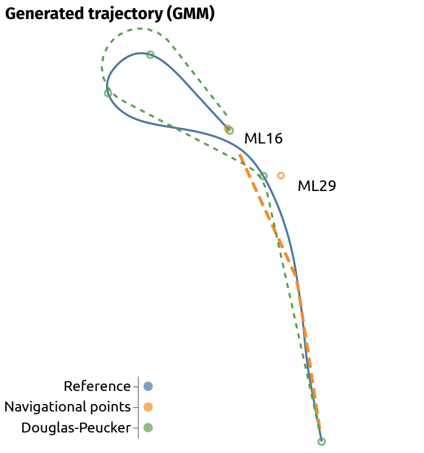
- Extract trajectory features based on
  1. navigational points i.e., `DIRECT` instructions
  2. Douglas–Peucker simplications
- Measure the distance between real/generated trajectories and their replay in BlueSky
  https://github.com/bguillouet/traj-dist

Results

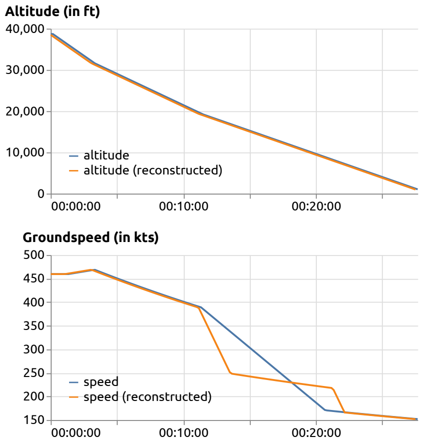

DTW
Cumulative ratio
1.0
0.8
0.6
0.4
0.2
0.0
0
400,000
800,000
1,200,000
Distance
SSPD
Cumulative ratio
1.0
0.8
0.6
0.4
0.2
0.0
0
4,000
8,000
12,000
16,000
Distance
Generation method
Reference
GMM
Reconstruction method
Navigational points
Douglas-Peucker

Limitations due to the modelling and capabilities of the simulator.





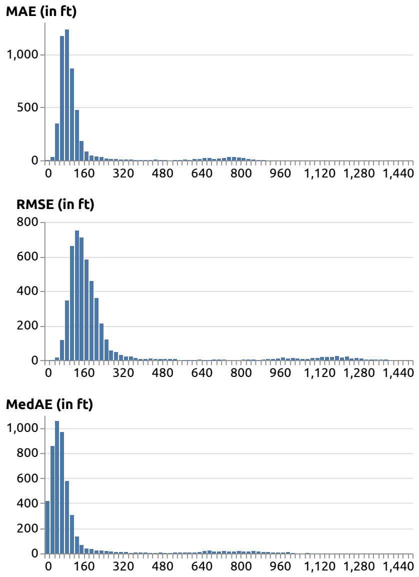
Generation of vertical profiles



Using aircraft kinematic models from OpenAP

- constant Mach/CAS, crossover altitudes, vertical rates segments
- Altitude successfully regenerated.
- Difference in speed default deceleration value after the constant CAS descent (OpenAP vs. BlueSky)

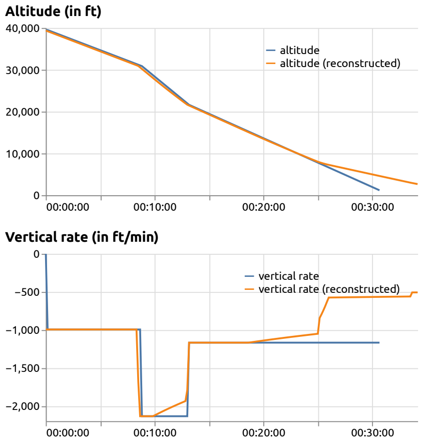
Statistics of generated vertical profiles



Average error below 200 ft

(5000 generated trajectories)

Vertical profile generation goes wrong (outlier)



Divergence between the two trajectories at lower altitudes (below 10,000 ft).

Higher than desired vertical rate (limited by the performance model in the simulator)

Conclusion

Key take-aways

1. Clearly state the expectations for the use case
   Consider relevant metrics in accordance with the original requirement
2. When comparing methods, avoid metrics on which one of the implementations is based
   Caution **if you reinject the metric as an objective to improve the generation method**
3. Challenge metrics with real trajectories
4. Take metrics for what they are, use them to improve all steps in your overall approach