BOEING

ICRAT 2018

Castelldefels,
June 25, 2018

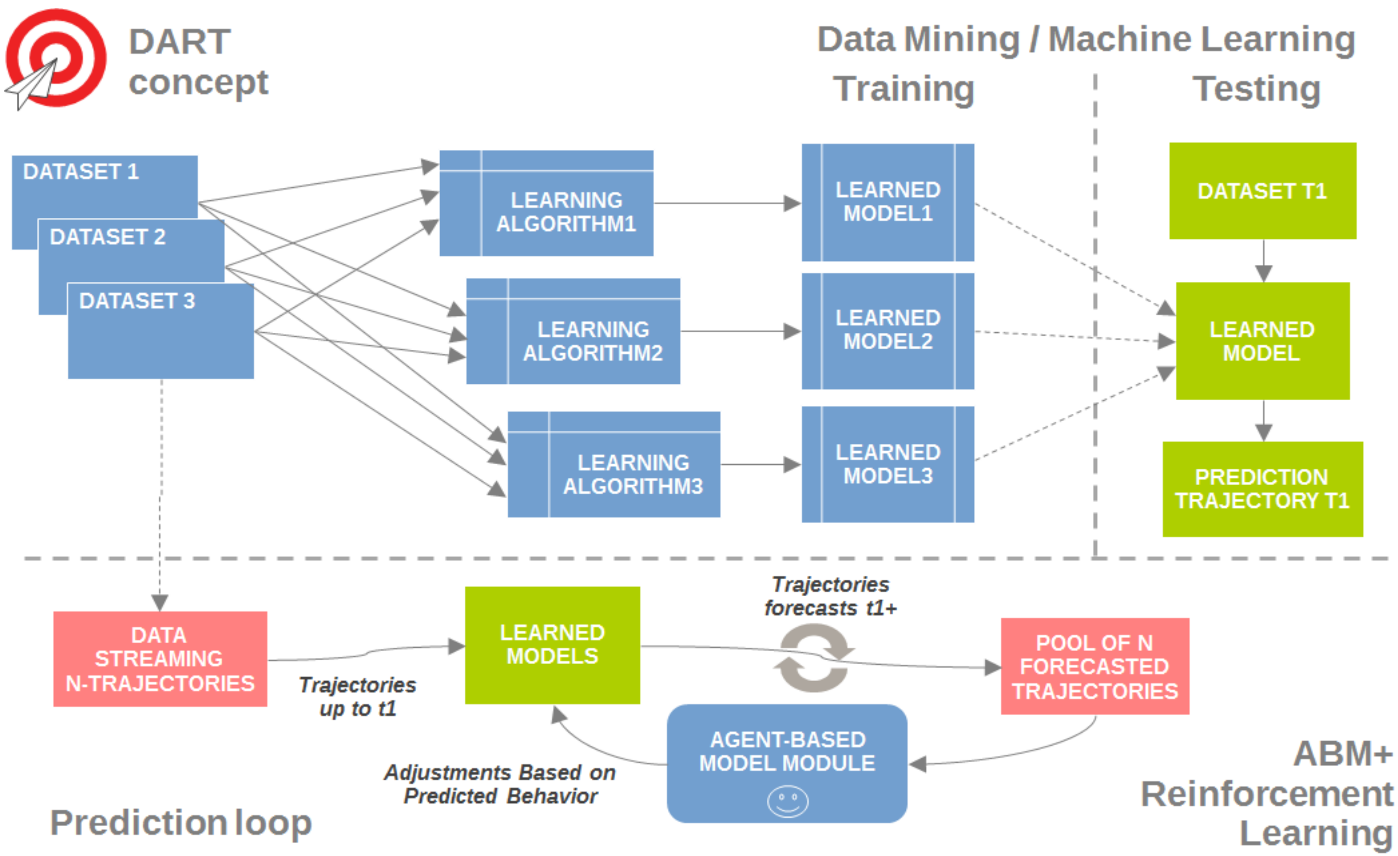
Hybrid data-driven trajectory prediction analytics

Data-Enhanced Trajectory Based Operations Workshop

DART Concept

Objectives
DART will deliver **understanding on the suitability of applying data-driven and agent-based models** for enhancing our abilities to **increase predictability** of aircraft trajectories.

Increasing predictability <-> Reducing uncertainty



Motivation

- **Flight Planning** refers to the process of planning and optimizing a flight strategically in accordance to the airline business model
  - **Optimization** of fuel- and time-related costs

- Nowadays, this is a **deterministic process** based on the description of different models (model-based approach):
  - Aircraft Performance
  - Weather Forecasts
  - Airline Preferences (e.g., Cost Index)
  - Operational Context
  - Intended vertical and horizontal profiles

- Current approaches **do not leverage** historical data
  - Planed optimized flight conditions are rarely reached

Is it possible to learn from the past to improve the planning of future flights?

Single Trajectory Prediction

LEARNING FROM RAW DATA

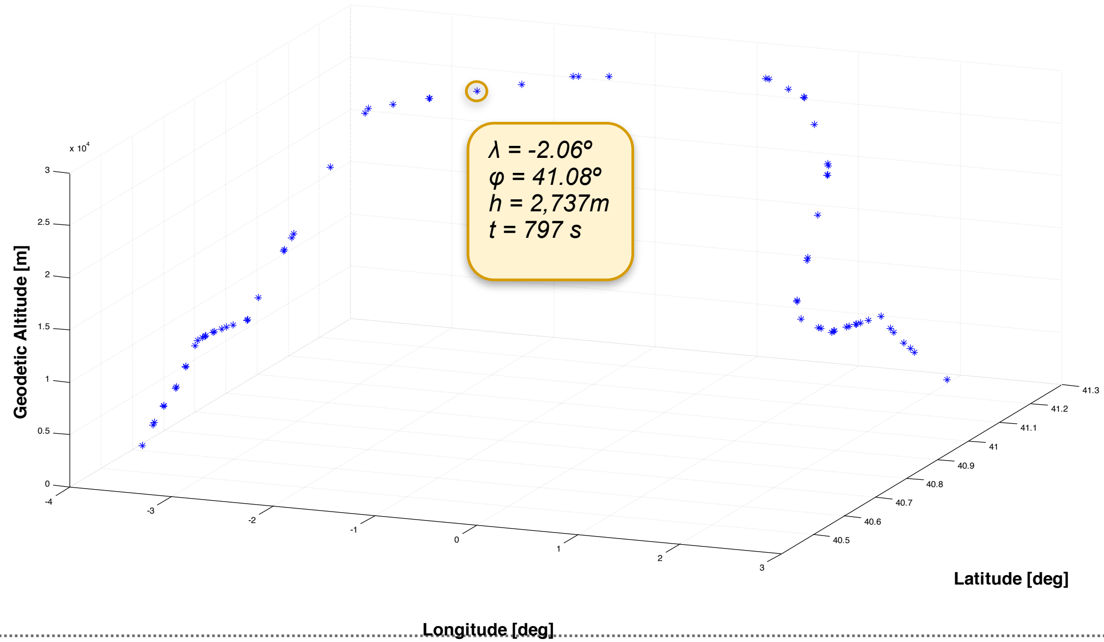
Flight Plan

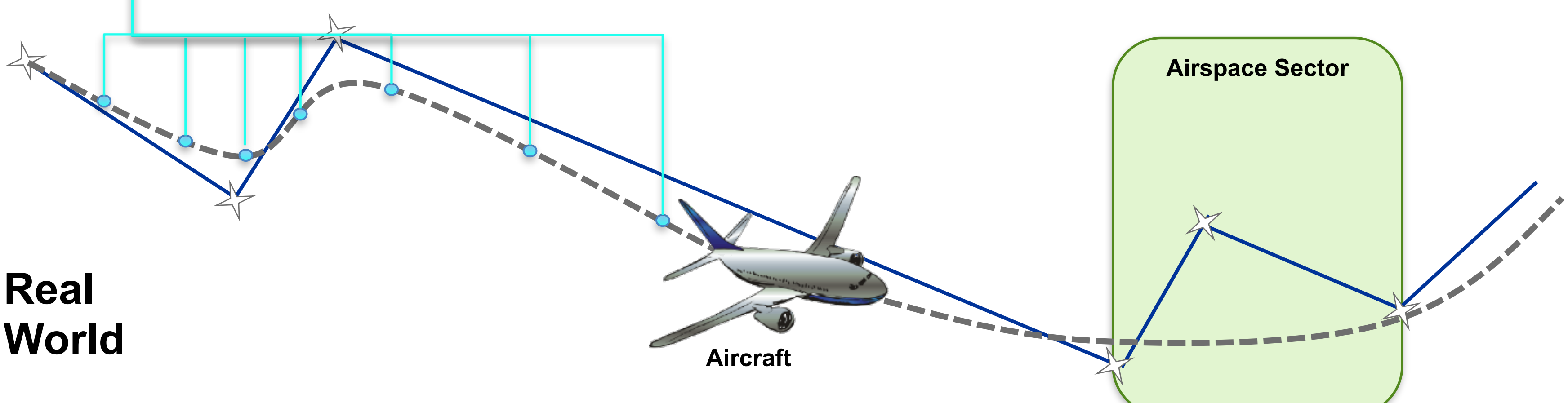
Speed
Flight Level (FL)
Route
Estimated Times
Acft ID

Surveillance Data

ADS-B
Radar Tracks

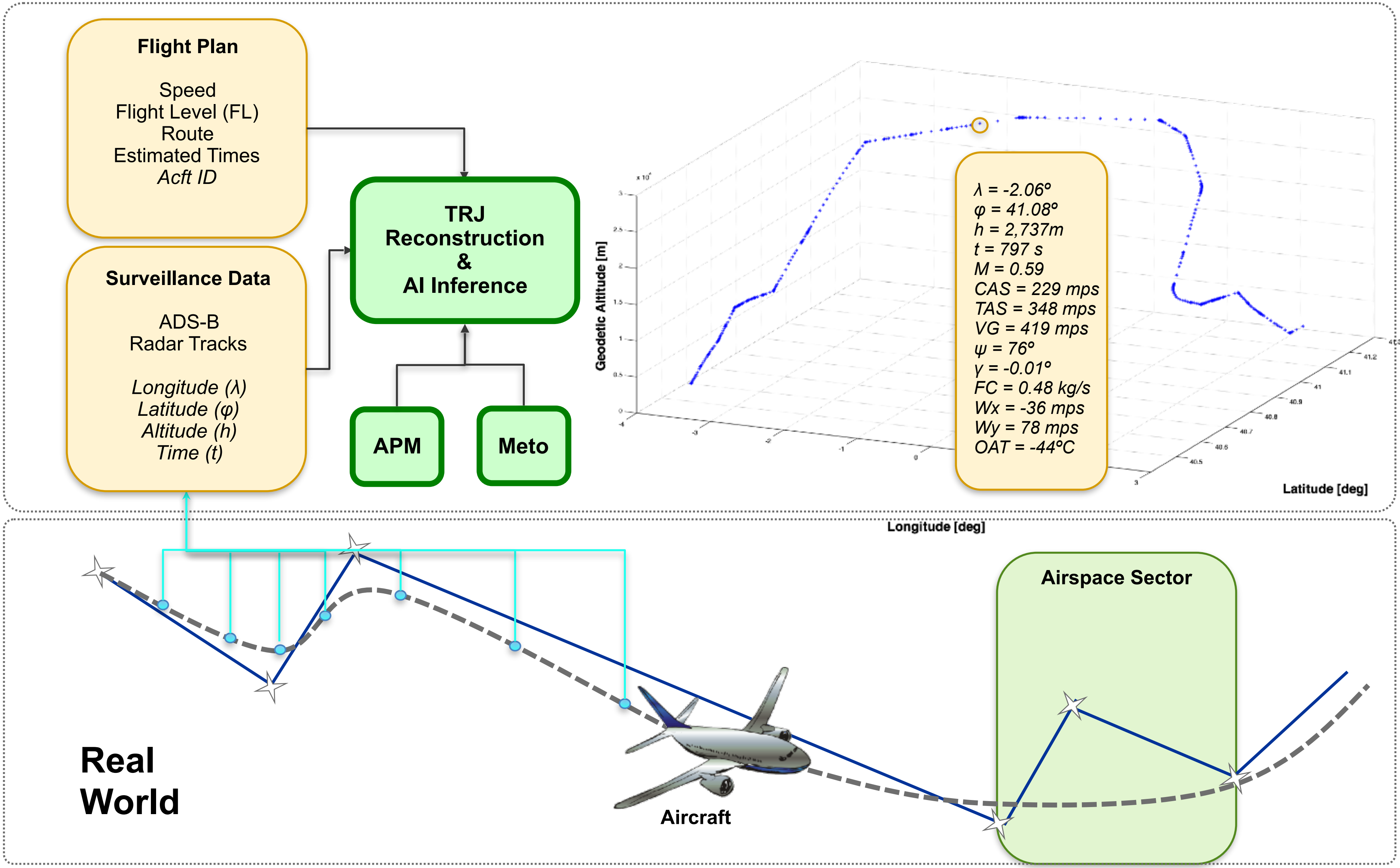
Longitude (λ)
Latitude (φ)
Altitude (h)
Time (t)





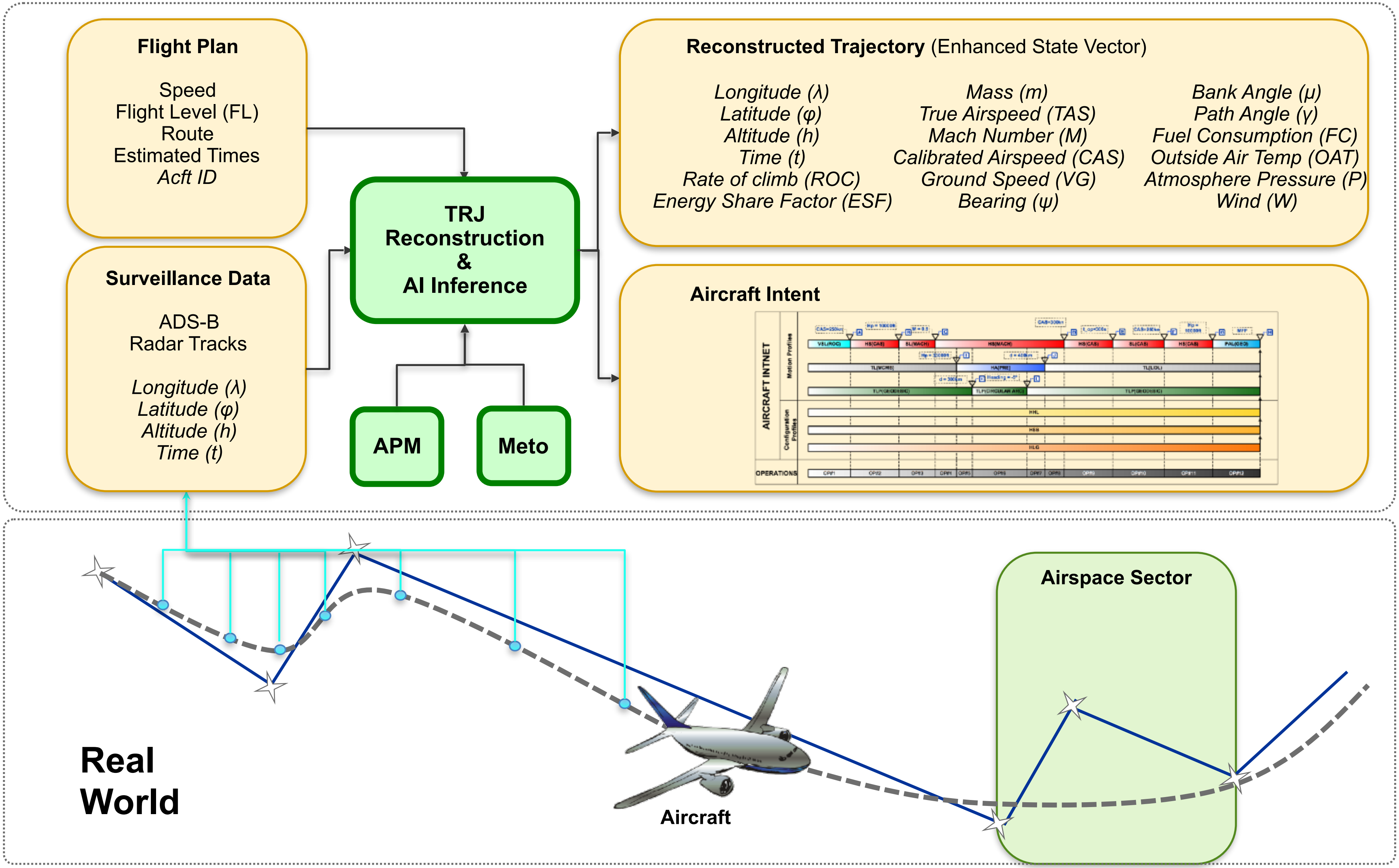
Single Trajectory Prediction

LEARNING FROM RECONSTRUCTED TRAJECTORIES



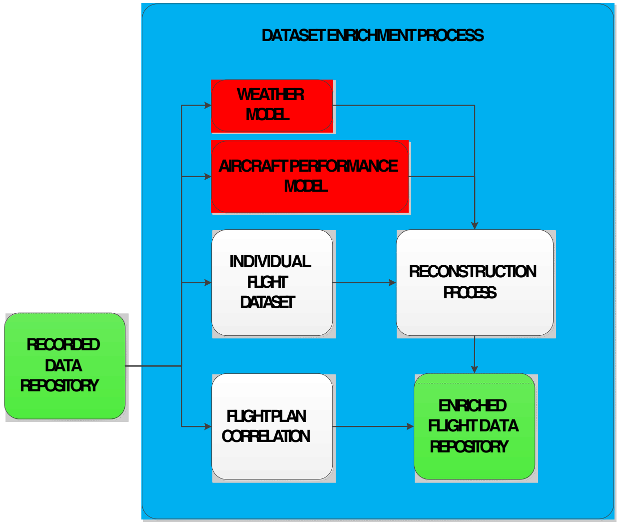
Single Trajectory Prediction

LEARNING FROM AIRCRAFT INTENT



Learning from data: Enrichment Process

DATA ENRICHMENT PROCESS

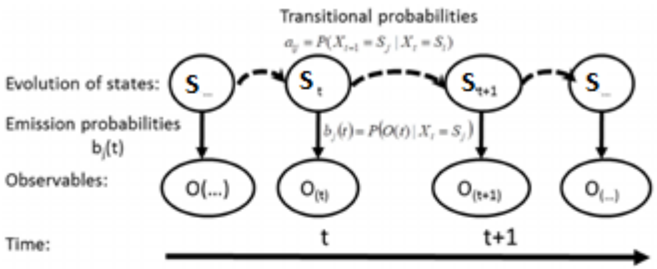


- **Weather Model**: information regarding wind and atmosphere conditions
- **Aircraft Performance Model**: basic aircraft information about drag, thrust and fuel consumption
- **Individual Flight Dataset**: recorded data onboard or surveillance data
- **Flight Plan Correlation**: correlation of main features that represent the flight and are linked to the flight data
- **Reconstruction Process**:
  - La Civita, Marco. "Using aircraft trajectory data to infer aircraft intent." U.S. Patent No. 8,977,484. 10 Mar. 2015.
  - Luis, P. D., La Civita, M., Lopez, J., & Miguel, A. Vilaplana. "Computer-implemented method and system for estimating impact of new operational conditions in a baseline air traffic scenario." U.S. Patent Application No. 15/155,754. 22 Dec. 2016.
- **Enriched Flight Data Repository**: collection of all enriched flight dataset including additional state variables with a higher sample rate

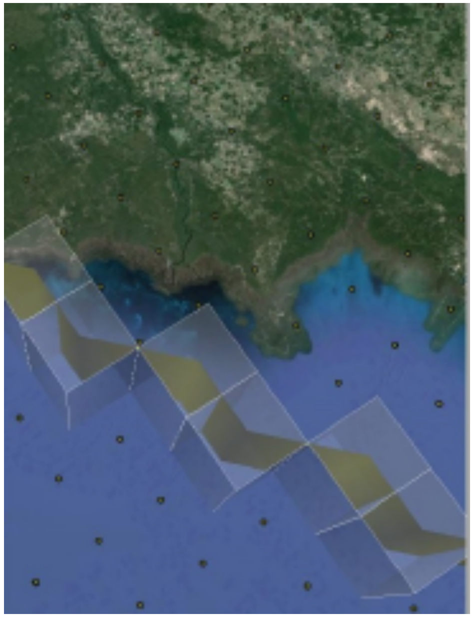
Dataset Enrichment Process

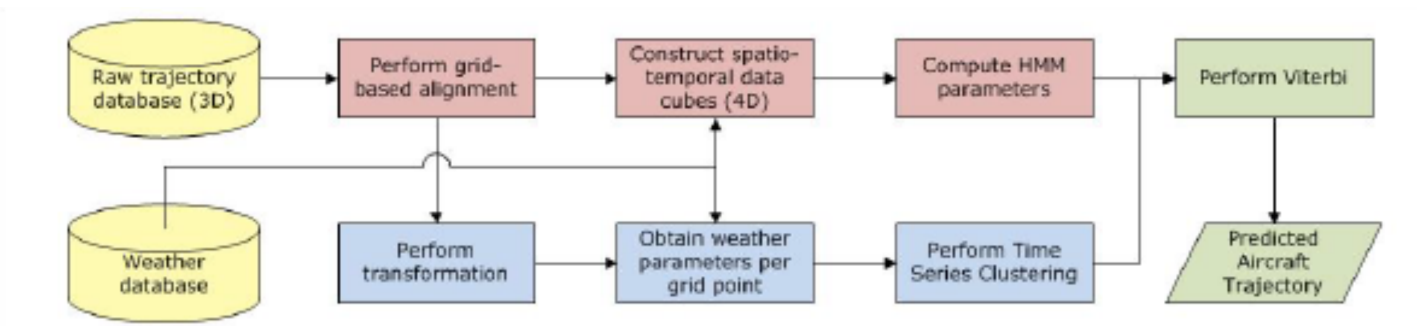
Trajectory Prediction from RAW Data
Hidden Markov Models

Given a set of historical raw or **reconstructed trajectories** for specific aircraft types along with pertinent historical weather observations, we aim at learning a model that reveals the correlation between weather conditions and aircraft positions and predicts trajectories in the form of a time series.



- States $S = \{S_1, S_2, ..., S_K\}$ are represented by reference points' coordinates *(latitude, longitude, altitude)* that form aligned trajectories.
- Transition probabilities $A = \{a_{ij}\}$, $1 \leq i, j \leq K$, i.e. a_{ij} is the probability of an aircraft discretely transitioning from one state S_i to another S_j along its aligned trajectory, $\bar{T}$.
- Emission probabilities $B = \{b_i(o)\}$, $1 \leq i \leq K$ is the probability of discrete weather parameters having been observed at a specific state, S_i.
- Initial probabilities $\pi = \{\pi_i\}$, $1 \leq i \leq K$ is the probability of an aligned trajectory beginning at a specific state, S_i.



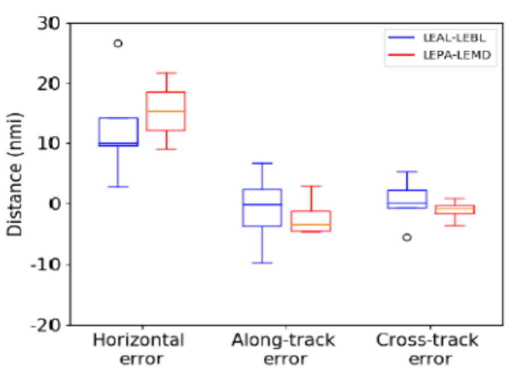


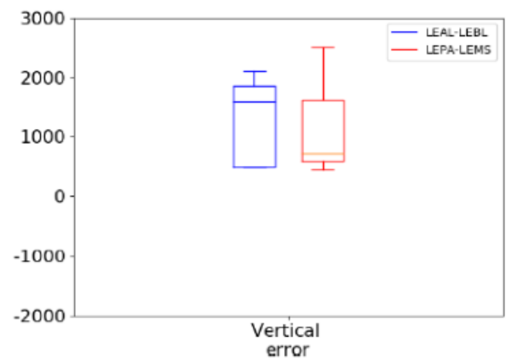
Trajectory Prediction from RAW Data: Hidden Markov Models.

Results from Raw trajectory data based algorithms

Hidden Markov Models & Gradient Boost Machine Regression

TestCase#	Route	TrainingSetSize		TestSetSize		Route	TrainingSetSize		TestSetSize	
		#trjs	#pts	#trjs	#pts		#trjs	#pts	#trjs	#pts
1	LEAL-LEBL	1118	55116	200	9860	LEMD-LEIB	2572	125623	200	9769
2	LEAL-LEBL	1118	55116	19	937	LEMD-LEMH	1056	68141	19	1226
3	LEAL-LEBL	1118	55116	152	7493	LEPA-LEMD	5116	306128	152	9095
4	LEBL-LEMG	1704	127451	43	3216	LEMD-LEMH	1056	68141	43	2775
5	LEBL-LEMG	1704	127451	180	13463	LEPA-LEMD	5116	306128	180	10771
6	LEBL-LEZL	2404	183343	41	3127	LEMD-LEAM	1434	70128	41	2005
7	LEBL-LEZL	2404	183343	46	3508	LEMD-LEMH	1056	68141	46	2968
8	LEBL-LEZL	2404	183343	164	12508	LEMG-LEMD	1403	75408	164	8815
9	LEBL-LEZL	2404	183343	210	16016	LEPA-LEMD	5116	306128	210	12566
10	LEIB-LEBL	1360	53443	259	10178	LEPA-LEMD	5116	306128	259	15498
11	LEIB-LEBL	1360	53443	158	6209	LEPA-LEVC	1426	50467	158	5592
12	LEMG-LEBL	1563	114767	38	2790	LEMD-LEAM	1434	70128	38	1858
13	LEMG-LEBL	1563	114767	46	3378	LEMD-LEIB	2572	125623	46	2247
14	LEZL-LEBL	2380	186299	40	3131	LEMD-LEIB	2572	125623	40	1954



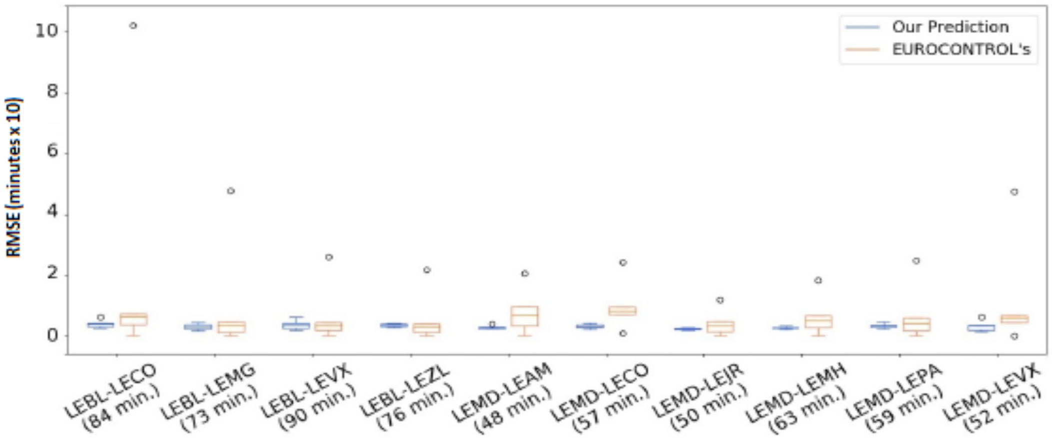


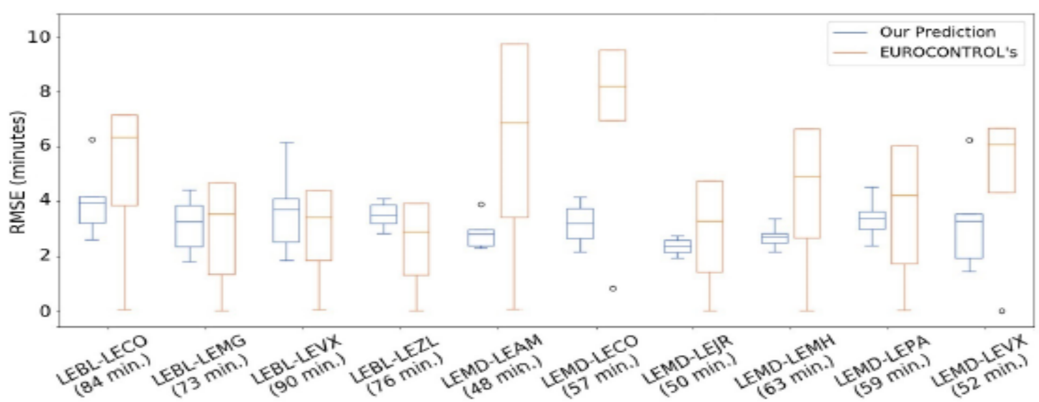
The mean value for the cross-track error and vertical error along the entire test trajectories in all 14 route pairs is **7.692nmi** and **1589.452ft**

Trajectory Prediction from RAW Data: Hidden Markov Models.

Results from Raw trajectory data based algorithms

ETA comparison with EUROCONTROL on flight departure

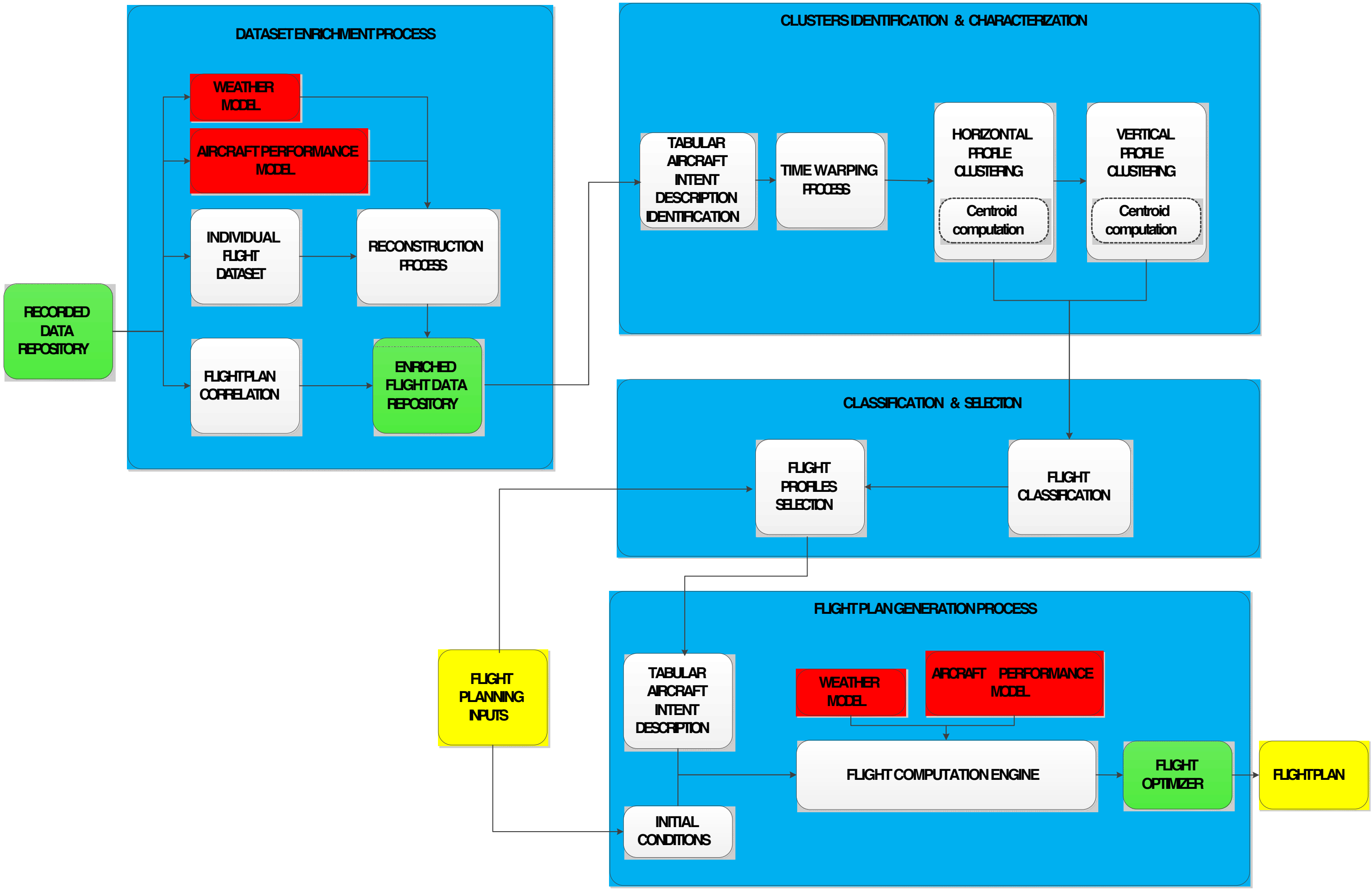




Compare data driven and model based generated trajectories
Actions:
- Get from Eurocontrol DDR2 database 2016 and 2017 predicted trajectory data (CTFM, FTFM, RTFM)
- Extract model based CTFM trajectories (predictions) from the dataset
- Perform ETA and trajectory comparison between CFTM trajectories and HMM

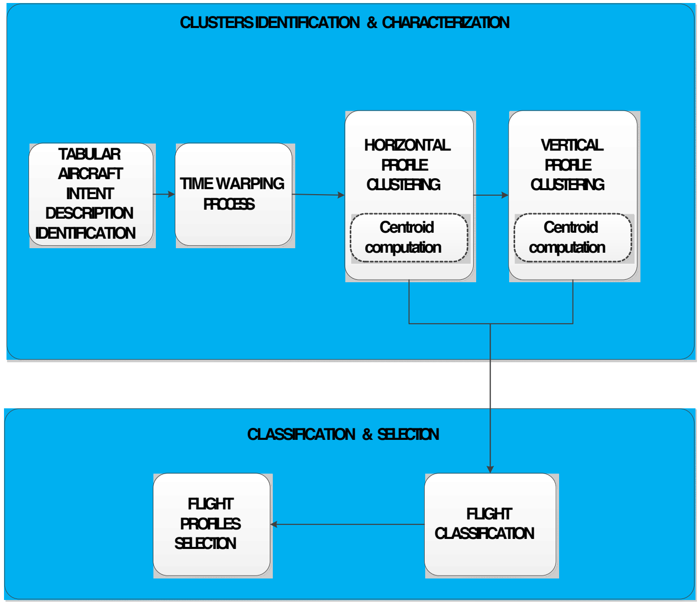
1. Our prediction yields 9% better median scores on eight routes, while the Eurocontrol's ETA shows better median scores on two routes (LEBL-LEVX and LEBL-LEZL).
2. The standard deviation values in Eurocontrol's ETAs are much larger, resulting in larger windows of predictability at arrival times.
3. Boxplots representing Eurocontrols's ETAs show extreme outliers.

Hybrid Data-Driven Flight Planning Architecture: using Aircraft Intent



Flight Planning Tool Overall Approach

Hybrid Data-Driven Flight Planning Architecture



DATA-DRIVEN PROCESS

- **Clusters Identification & Characterization**
  - **Tabular Aircraft Intent Description Identification**: variations of pressure altitude (Hp), Mach Number (M) and aerodynamic bearing (X_{TAS})
  - **Time Warping**: normalization of time duration of all flights to the interval [0,1]
  - **Horizontal Profile Clustering**: flights are grouped according to X_{TAS}
    - Computation of the centroid that characterizes each cluster
  - **Vertical Profile Clustering**: flights are grouped according to Hp and M
    - Computation of the centroid that characterizes each cluster
- **Classification & Selection**
  - **Flight Classification:** based on Flight Planning Features
  - **Flight Profiles Selection**: selection of centroids that define the flight according to the input FP.

WIND at Origin and Destination, Cruise WIND & TEMP, Day of the Week, Cruise FL, Wake Vortex Category

Hybrid Data-Driven Flight Planning Architecture

Classification & Selection

- Selection of most likely centroid according to the Flight Planning features
  - Random Forest (RF), algorithm that grows many classification trees. Each generated tree gives a classification for a new input, while the forest (i.e., the ensemble of all trees) provides an overall classification weighing the outcomes of all individual trees.
    - It runs efficiently on large data bases.
    - It can handle thousands of input variables without variable deletion.
    - It gives estimates of what variables are important in the classification.
    - It generates an internal unbiased estimate of the generalization error as the forest building progresses.
- The centroids determined the lateral and vertical flight profiles by described the time evolution of M, Hp & X_{TAS} :
  - It ensures that a model-based approach can be used afterwards to compute a nominal flight once a weather forecast is available and there is a description of the aircraft performance.

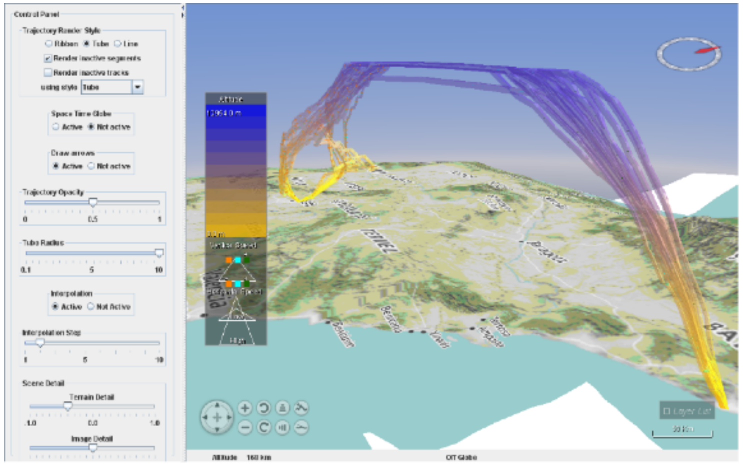
RF is fast and is not prone to overfit

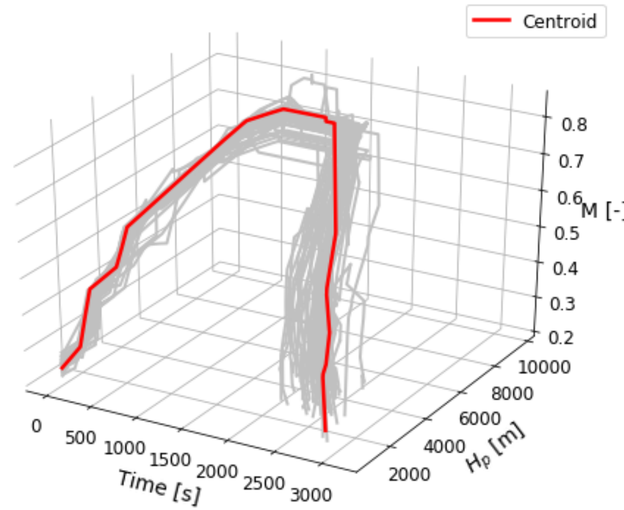
Case Study

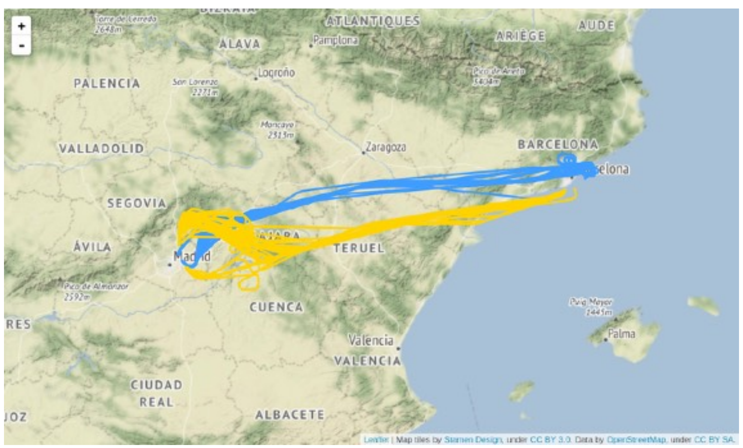
- **Objective**:
  - Planning a flight departing form LEBL and arriving to LEMD making use of the proposed hybrid planning approach
- **Historical Recorded data**:
  - 1/2-year of surveillance data (fused radar tracks) **3,356 flights**
  - 1/2-year of Flight Plans (correlated to surveillance data)
  - 1/2-year of weather models (downloaded from NOAA)
- **Results**:
  - Comparison with the Planned Flight
  - Comparison with the actually flown Flight

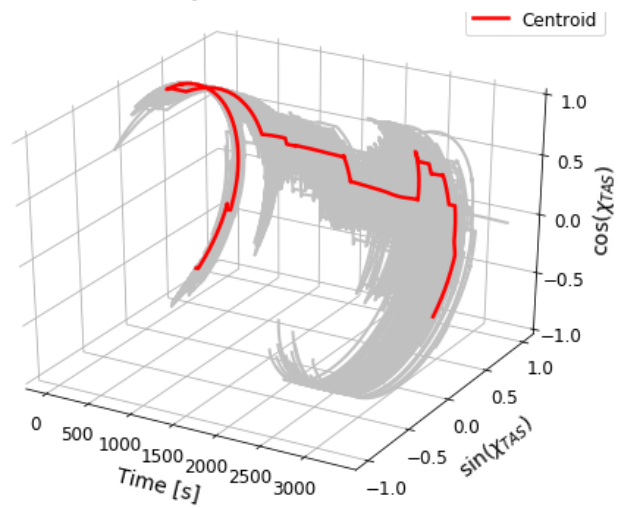
Case Study Description

Case Study - Clustering



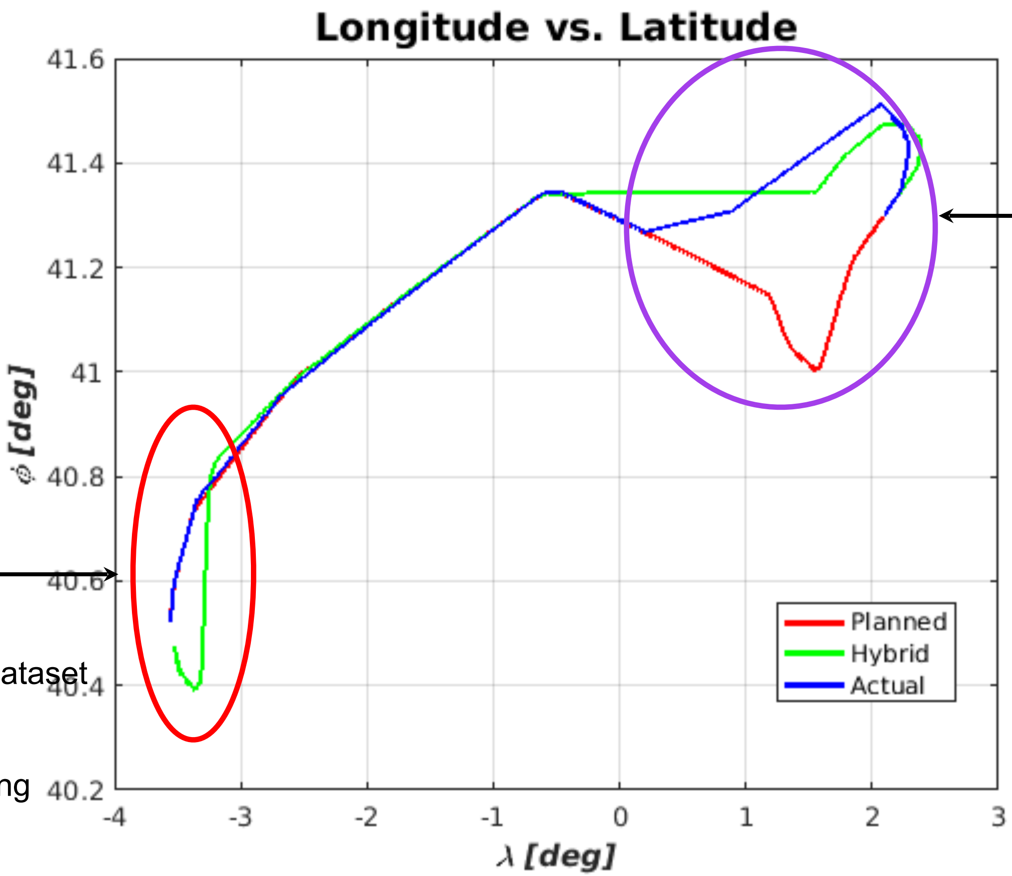






Cluster Visualization

Case Study - Results: Lateral Profile



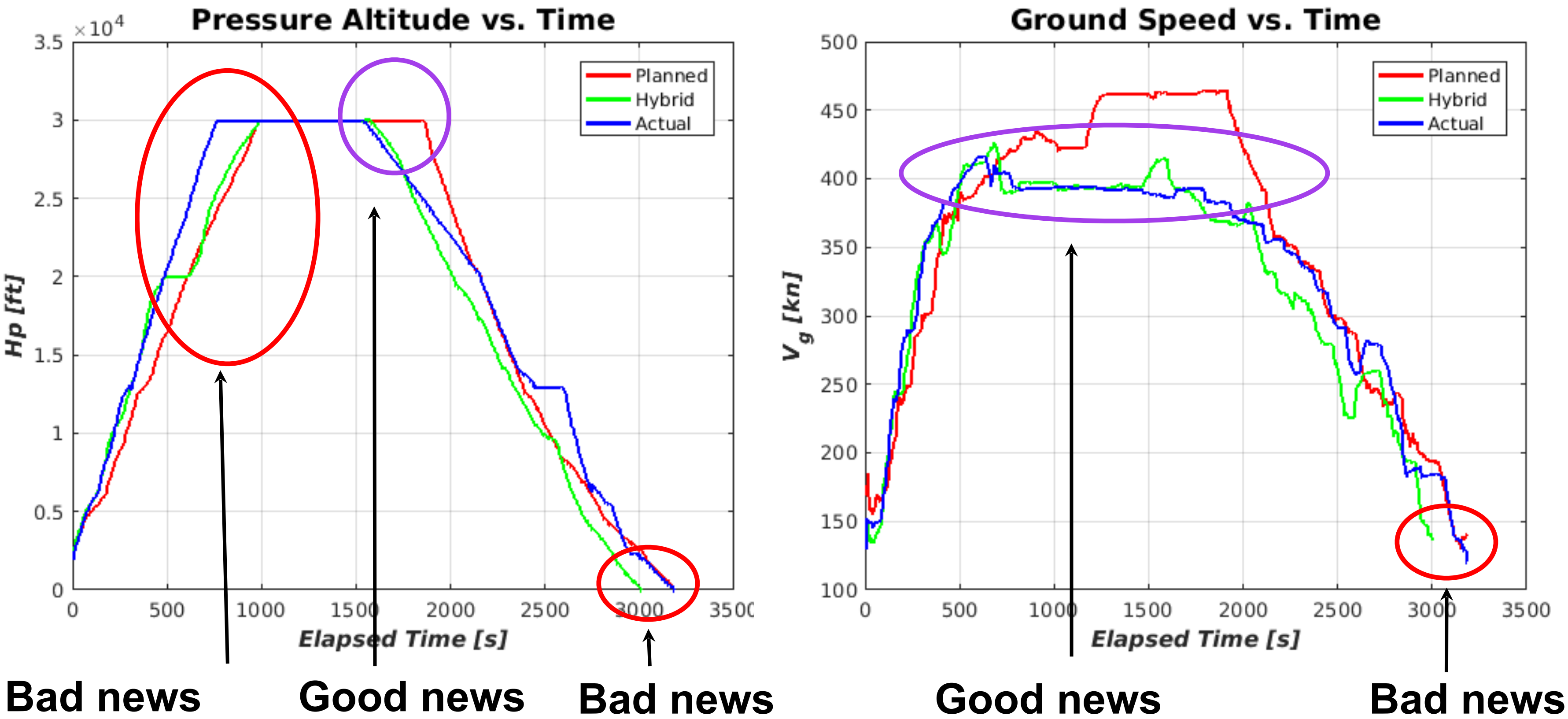
Good news

Bad news

1. Increase learning dataset
2. Review clustering process
3. Features engineering

Potential improvements in the lateral profile

Case Study – Results: Vertical Profile



1. Use of centroids 2. Features engineering 3. Explore other ML algorithms

Improved ~10% RMS_{Hp} & ~50% RMS_{Vg}

Remarks & Future Work

- An hybrid flight planning architecture has been implemented
  - It leverages historical data
  - It make use of models to improve pure data-driven Flight Planning
  - It ensures that hidden patterns are considered during Flight Planning

- Required an extended analysis with higher number of trajectories (i.e., extended learning dataset)

- Some potential improvements have been identified:
  - **Clustering Process**
    - Principal Component Analysis (PCA)
    - Robust K-Means Clustering
  - **Classification methods**
    - Boosted Trees (e.g., XGBoost)
  - **Features engineering**
  - **Flights segmentation**

What improvements can be achieved in the future?