# Data-driven Aircraft Trajectory Prediction

H. Georgiou, N. Pelekis, Y. Theodoridis, G. A. Vouros
Univ. of Piraeus (UPRC), Greece

Data Enhanced TBO Workshop @ ICRAT 2018

The EU Framework Programme for Research and Innovation
HORIZON 2020
SESAR JOINT UNDERTAKING
Founding Members
EUROPEAN UNION EUROCONTROL
DART
datAcron

# DART Operational Context

**Overall Aim :**

Demonstrate how DART predictive analytics capability can **improve trajectory prediction in support of DCB processes at planning phase**, further reducing uncertainty and improving ATM operations and services provided.

**Scenario 1** (AUs): aims to **compute the predicted trajectory** that an aircraft will fly during an operation day **without considering traffic**.

**Scenario 2** (ANSPs): aims to **study and determine the complexity to be considered in trajectories due to the influence of the surrounding traffic**, at the planning phase, taking into account flight plans and/or individual trajectory predictions.

# Single Trajectory Prediction

**Objective:** Given a flight plan, predict the actual trajectory, i.e., the 3-D route of the corresponding flight,  w.r.t. to information that really matters …. local weather, aircraft type, ... Etc.

**…. In a Data-Driven Way**

**…. Exploiting historical data on actual <u>enriched</u> trajectories**

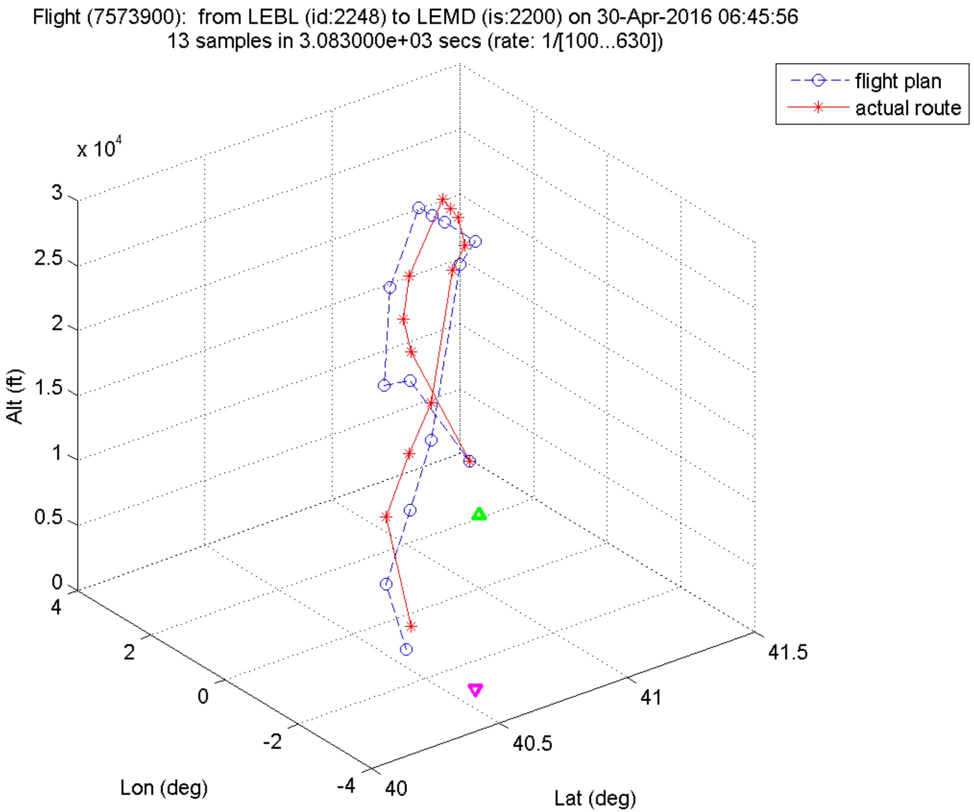
# Single Trajectory Prediction

**Objective:** Given a Flight Plan, predict the actual trajectory, i.e., the 3-D route of the corresponding flights,  w.r.t. to information that **really matters** …. local weather, aircraft type, ... Etc.

Predicted trajectories, in the context of DCB problems, can be used for predicting evolution of demand per sector, and thus hotspots.

**Our case:** Trajectories are predicted one by one, without considering traffic.

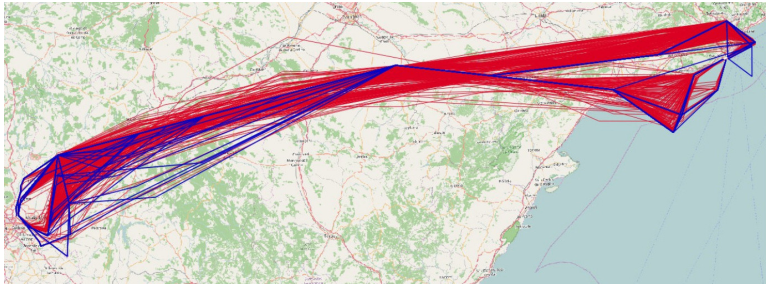
# Hybrid Clustering

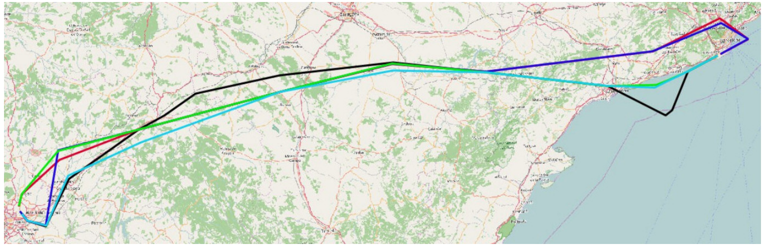



**Basic idea – Method outline:**

1. Input: Flight plans, actual routes, local weather, aircraft type, (...)
2. **Stage-1**: Cluster the actual enriched routes, producing medoids of clusters as “representatives”
3. **Stage-2**: Build a Pred.Model for each cluster, associate it with the cluster flight plans (emissions)
4. **Stage-3**: For a new flight plan, find the *k* closest matches (Pred.Model)
5. Output: $k \geq 1$ best-matches of the query FP, for prob.estim. or further refinement

# Hybrid Clustering

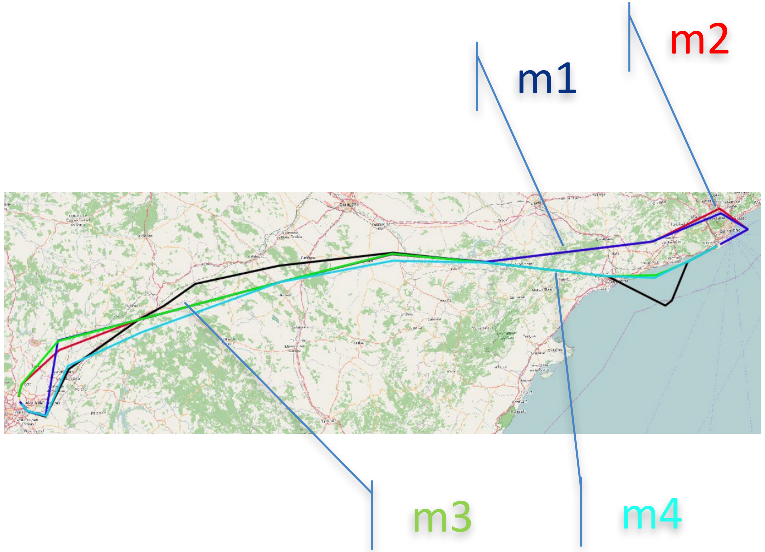




## Basic idea – Method outline:

1. Input: Flight plans, actual routes, local weather, aircraft type, (...)
2. **Stage-1**: Cluster the actual enriched routes, producing medoids of clusters as "representatives"
3. **Stage-2**: Build a Pred.Model for each medoid, associate it with the cluster flight plans (emissions)
4. **Stage-3**: For a new flight plan, find the *k* closest matches (Pred.Model)
5. Output: $k \geq 1$ best-matches of the query FP, for prob.estim. or further refinement

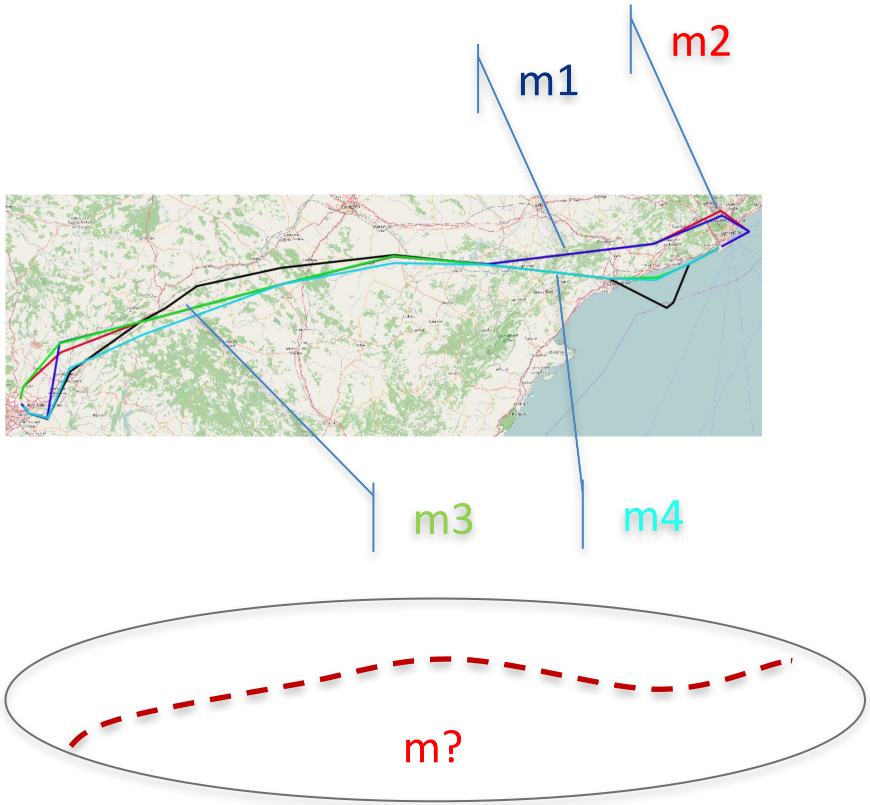
# Hybrid Clustering




**Basic idea – Method outline:**

1. Input: Flight plans, actual routes, local weather, aircraft type, (...)
2. **Stage-1**: Cluster the actual enriched routes, producing medoids of clusters as "representatives"
3. **Stage-2**: Build a Pred.Model for each medoid, associate it with the cluster (emissions: flight plans vs. medoid)
4. **Stage-3**: (optional) For a new flight plan, find the *k* closest matches (Pred.Model)
5. Output: $k \geq 1$ best-matches of the query FP, for prob.estim. or further refinement

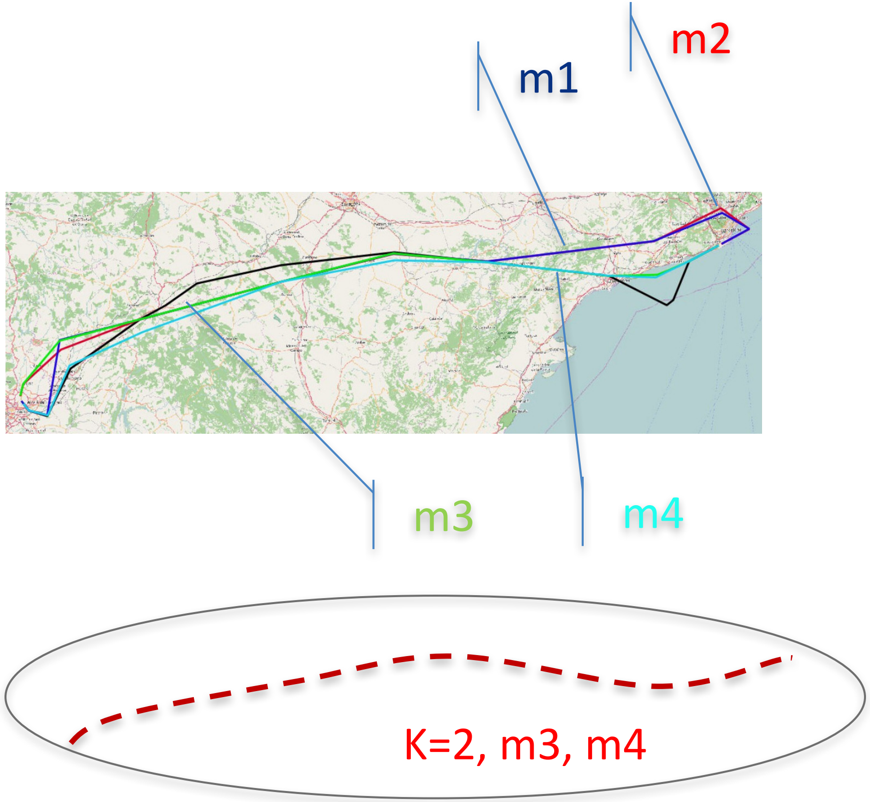
# Hybrid Clustering




**Basic idea – Method outline:**

1. Input: Flight plans, actual routes, local weather, aircraft type, (...)
2. **Stage-1**: Cluster the actual enriched routes, producing medoids of clusters as "representatives"
3. **Stage-2**: Build a Pred.Model for each medoid, associate it with the cluster flight plans (emissions)
4. **Stage-3**: For a new flight plan, find the *k* closest matches (Pred.Model)
5. Output: $k \geq 1$ best-matches of the query FP, for estimation or further refinement

# Hybrid Clustering



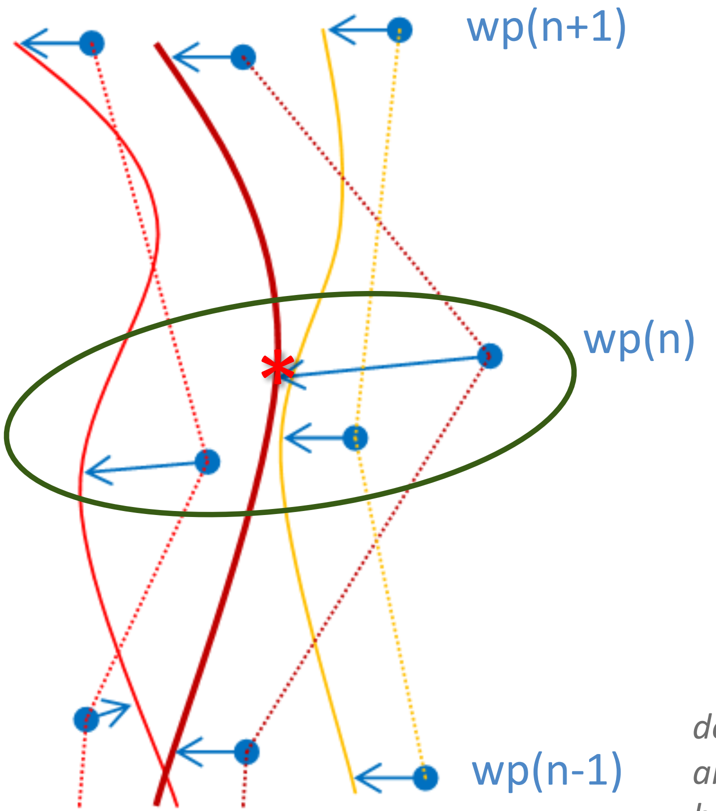

**Basic idea – Method outline:**

1. Input: Flight plans, actual routes, local weather, aircraft type, (...)
2. **Stage-1**: Cluster the actual enriched routes, producing medoids of clusters as "representatives"
3. **Stage-2**: Build a Pred.Model for each medoid, associate it with the cluster flight plans (emissions)
4. **Stage-3**: For a new flight plan, find the *k* closest matches (Pred.Model)
5. **Output:** *k* ≥ 1 best-matches of the query FP, for estimation or further refinement

# How to build predictive models (stage-2)

- **Hidden Markov Model (HMM)**
- **Linear Regressor (LR)**
- **Decision Tree (CART)**
- Support Vector Machine (RBF kernel)
- Gaussian Process (Sqr.Exp. kernel)
- Bagged Trees (ensemble)
- **Neural Network (NN-MLP)**
- Extreme Learning Machines (ELM)

# Hybrid Clustering-Pred.Model (HMM)

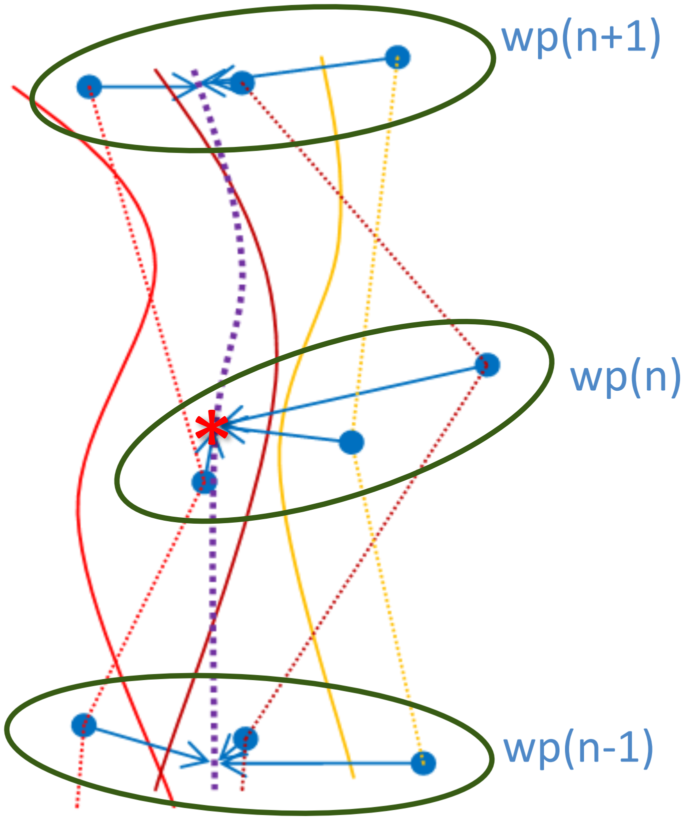



**Stage-2, HMM approach:**

RT(n)-FP(n) statistics are used to build a probabilistic model (HMM emissions) for ref. point wp(n).

*dotted line: flight plan (FP), solid line: actual route (RT)*
*arrows: FP/RT deviations, star: current pred. point (wp(n))*
*bold solid line: cluster medoid, pred. route for query FP*

# Hybrid Clustering-Pred.Model (LR, CART, NN-MLP)



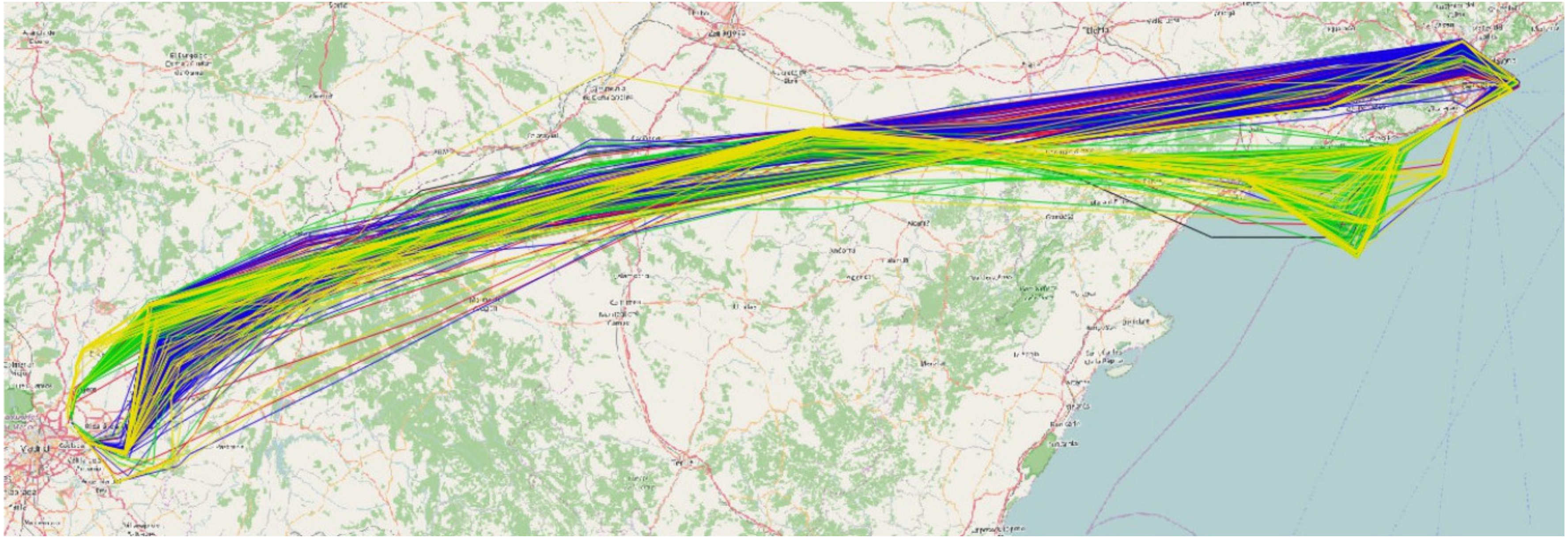

**Stage-2, LR or CART or NN-MLP approach:**

RT(n) is estimated as synthetic from multiple/all FP(*) ref. points, used to build a LSE-minimum linear prediction model.

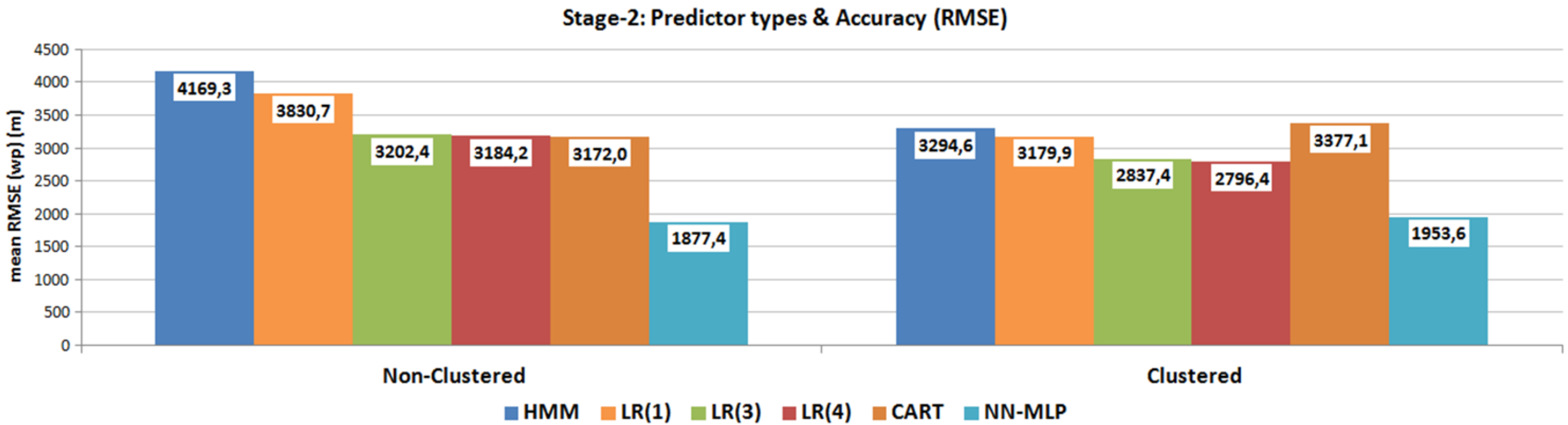
*dotted line: flight plan (FP), solid line: actual route (RT)*
*arrows: FP/RT deviations, star: current pred. point (wp(n))*
*bold solid line: cluster medoid, pred. route for query FP*

# Experimental dataset:

## Spain (Madrid-Barcelona), April 2016



# TP Performance Summary: LEMD/LEBL, April 20




**Example NN-MLP distribution of prediction errors (signed MAPE)(m) for Lat for one waypoint.**

- Flight plans provide optimization constraints, i.e., ‘guidelines’ for the TP training
- They are also a realistic assumption about the intended (a priori) flight path
- **Per-waypoint TP accuracy is in the order of 2-3 km (3-D RMSE), length-invariant**

# Research Outcomes

- **Multi-stage (hybrid) approach provides modularity & flexibility**

- Stage-1: **Clustering** provides
  a) improved **compactness** to the training subsets
  **b) Encapsulation of N-dim** input ('enrichments'), before the actual TP
  **c) improved accuracy in the actual TP**

- Stage-2: HMM provides max. expected errors with high confidence
- Stage-2: LR alternatives provide variable complexity vs. accuracy w.r.t. input dim.
- Stage-2**: NN regressors (non-linear) provide maximum accuracy, at training cost**
- Stage-2: CART (single) falls between LR and NN, but sensitive to noise & outliers

# Gains & Insights

- ➢ Using flight plans as ‘guidelines’ provides **four main advantages**:
  - **Scaling-down** to smaller TP sequences, e.g. 11-18 instead of 680-730 points
  - **Waypoint-based inherently parallelizable** TP design
  - **Length-invariant TP accuracy**, along the flight plan ‘intended’ path

- ➢ Prospects for real-world deployment:
  - ✓ **Predictive model is modular**, can be selected according to scale/resources
  - ✓ LR: Low-complexity, very fast re-training with
    - ✓ **2.8-3.5 km** accuracy (3-D rmse)
  - ✓ NN-MLP: High-complexity, slower training
    - ✓ **1.8-2.0 km** accuracy (3-D rmse)
  - ✓ *...**But** more complex pred. models can work single-stage (no clustering).*

# Data-driven Future Aircraft Trajectory Prediction

H. Georgiou, N. Pelekis, Y. Theodoridis, G. Vouros

Univ. of Piraeus (UPRC)

## Thank you for your attention!

This project has received funding from the SESAR Joint Undertaking under the European Union's Horizon 2020 research and innovation programme under grant agreement No [number]

This work was partially supported by the projects **datAcron** (No:687591) and **DART** (No:699299), funded by the European Union's **Horizon 2020** (H2020) programme.

http://datacron-project.eu
http://dart-research.eu