ΜΕΘΟΔΟΙ PREDICTIVE ANALYTICS ΣΕ BIG MOBILITY DATA ΚΑΙ ΙΑΤΡΙΚΕΣ ΕΦΑΡΜΟΓΕΣ

Χάρης Γεωργίου (MSc,PhD)
Data Science Lab @ Πανεπ. Πειραιά
Email: harris@xgeorgio.info

12 Δεκ. 2018 @ Ινστιτούτο Πληροφοριακών Συστημάτων (ΙΠΣΥ),
Ερευνητικό Κέντρο "Αθηνά"

Overview

- **DART** – Data-driven AiRcraft Trajectory prediction research
- **datAcron** – Big Data Analytics for Time-Critical Mobility Forecasting
- **Track & Know** – R&D in Big Data Mobility in transport, insurance & health

- **Assurance** – Sparse model learning in health Informatics (fMRI) → *ΕΚΠΑ*

** See: H. Georgiou et al. (2018), "Predicting the next steps of moving objects: A survey". (under preparation)*

** See: Y. Kopsinis, H. Georgiou, S. Theodoridis (2014). "fMRI Unmixing Via Properly Adjusted Dictionary Learning", 22nd European Signal Processing Conference (EUSIPCO 2014), 1-5 Sept 2014 @ Lisbon, Portugal.*

DATA STORIES
Data Science Lab. @ University of Piraeus

UNIVERSITY OF PIRAEUS RESEARCH CENTER

DATA STORIES

Data Science Lab. @ University of Piraeus

UNIVERSITY OF PIRAEUS RESEARCH CENTER

- The Data Science Lab at the Univ. Piraeus, established in 2015
- Aims advance research on a wide range of Data Science subjects, including:
  - **big data management**
  - **statistics and data (incl. text, audio) analytics**
  - **machine learning**
  - **semantic integration**
  - **mobility data exploration**
  - **data privacy**

Data Science Lab. @ University of Piraeus

UNIVERSITY OF PIRAEUS RESEARCH CENTER

DART

Data-driven AiRcraft Trajectory prediction research

Founding Members

EUROPEAN UNION

EUROCONTROL

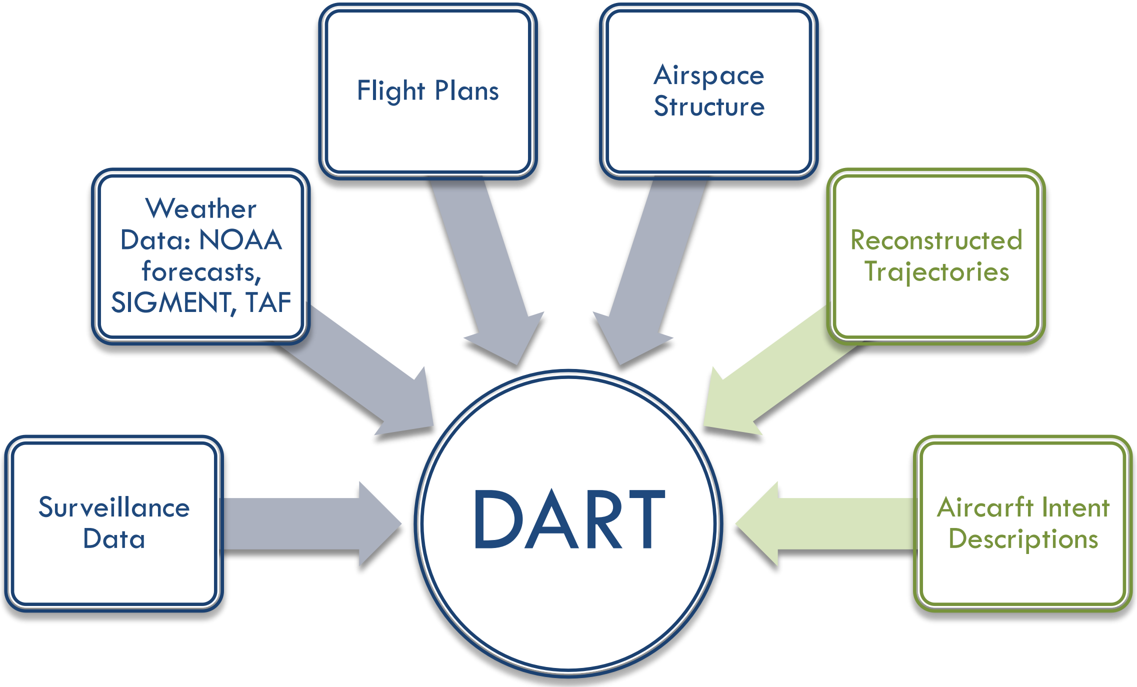
SESAR JOINT UNDERTAKING

- **Trajectory Prediction** in the Aviation domain is devoted to the analysis and evaluation of a wide range of *data-driven* techniques that could potentially be applied to the aircraft trajectory prediction problem.
- Main **specific research objectives** addressed:
  - Study of the application of *Big-data* techniques to trajectory-related data gathering, filtering, storing, prioritization, indexing and segmentation to support the generation of reliable and homogenous/fused input datasets.
  - Study of different data-driven learning techniques to describe how a reliable trajectory prediction model will leverage them.
  - Exploration of advanced visualization processes for data-driven model algorithms design, tuning and validation, in the context of 4-D trajectories.

** See: H. Georgiou et al. (2018), "Predicting the next steps of moving objects: A survey". (under preparation)*

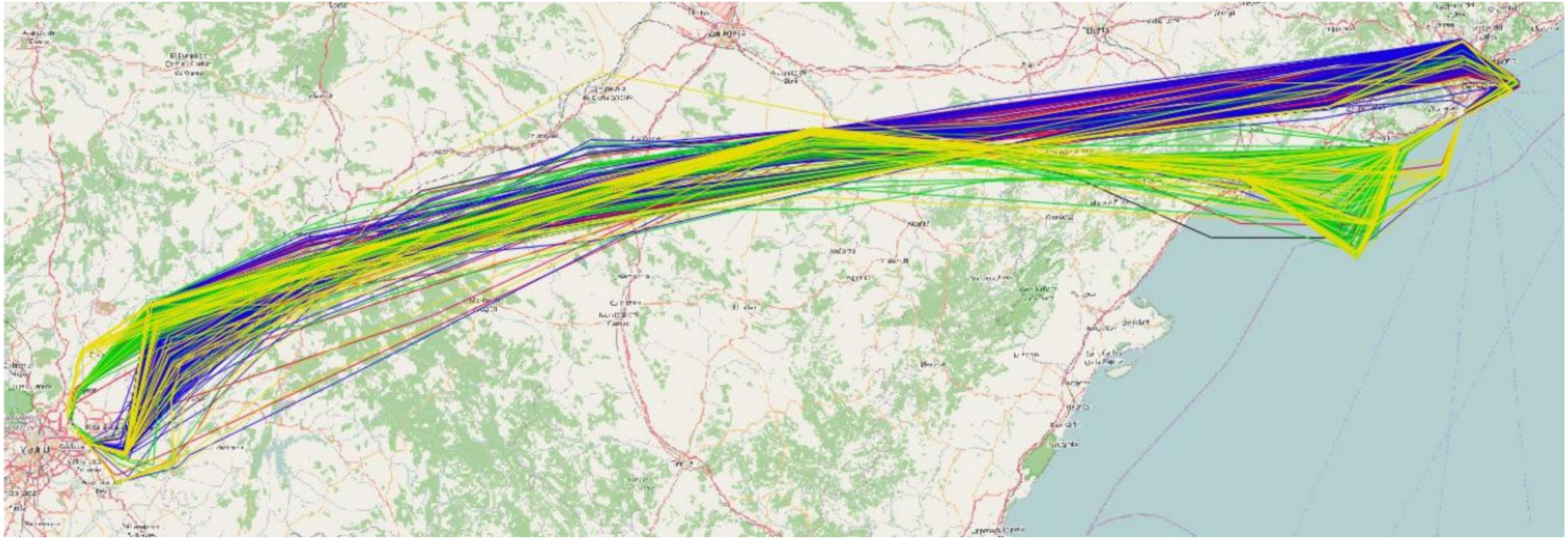
Data Science Lab. @ University of Piraeus

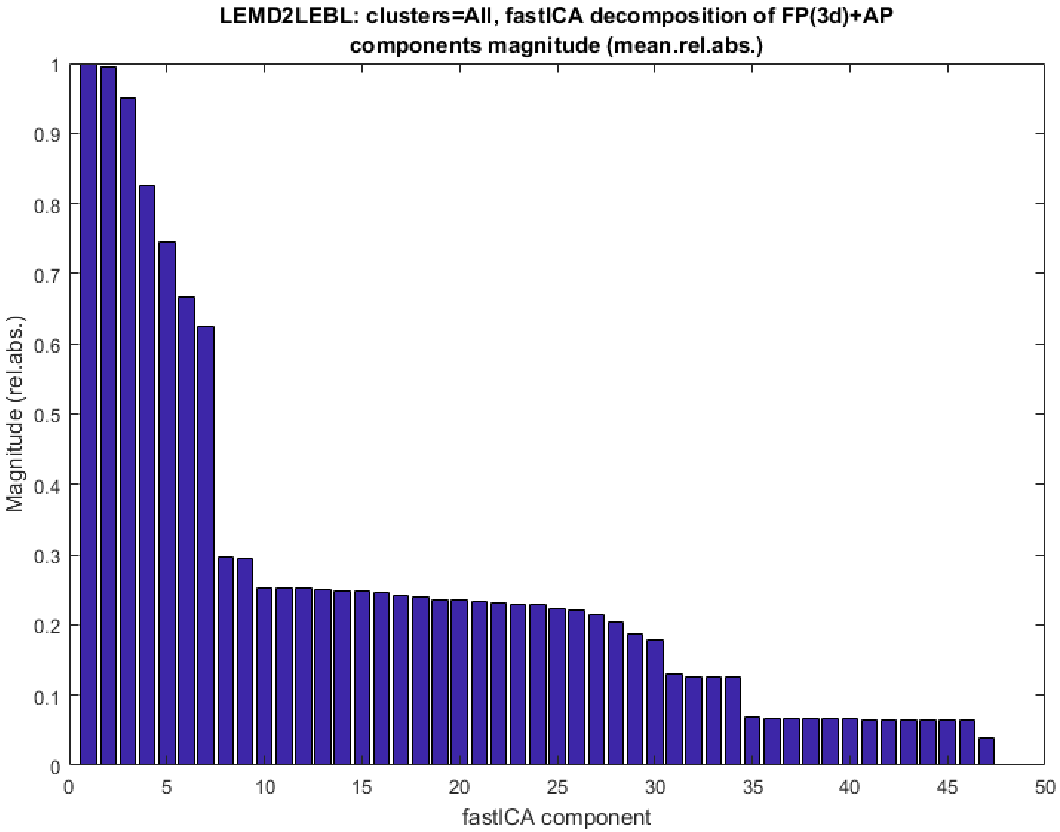
UNIVERSITY OF PIRAEUS RESEARCH CENTER



Example: Madrid/Barcelona route

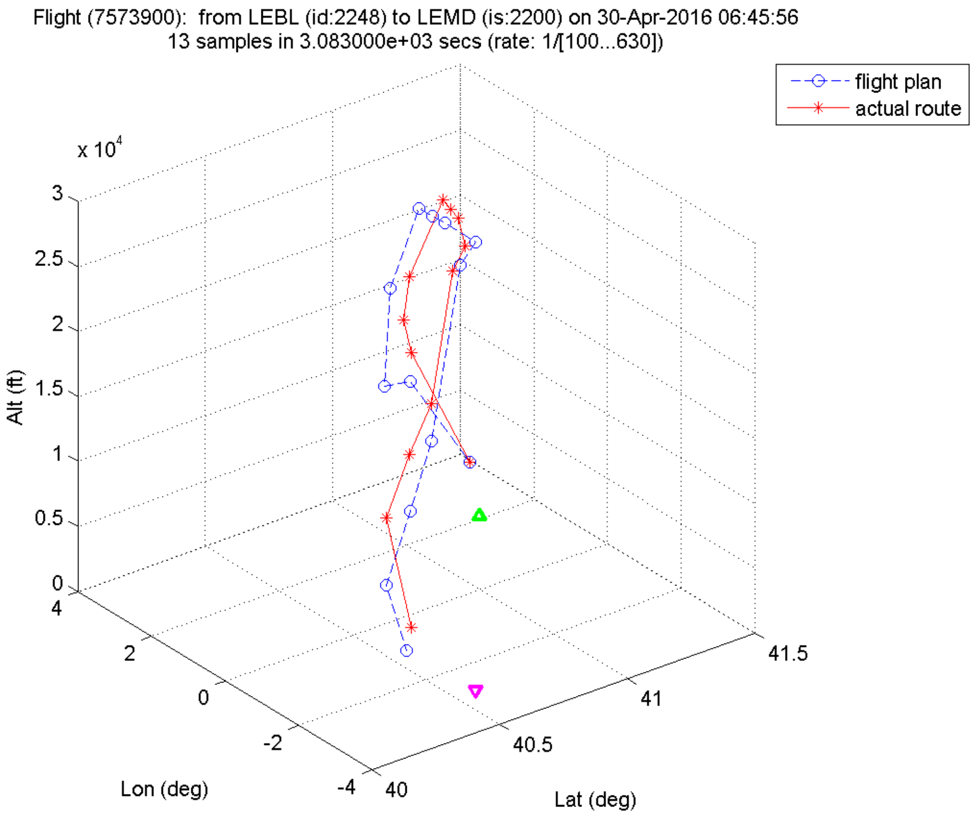
(about 700 flights per month per direction)





TP-C: fastICA spectrum of FP/RT dataset, Lat-only

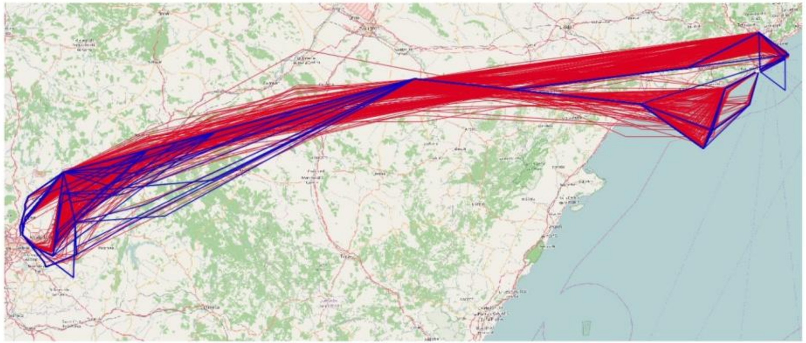
Hybrid Clustering-Pred.Model (HMM, LR, DT, NN)



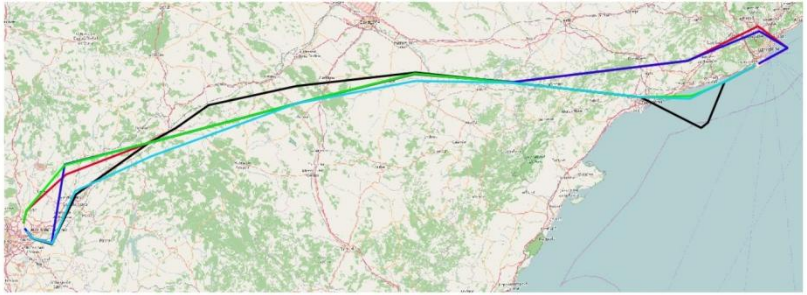
Basic idea – Method outline:

1. Input: Flight plans, actual routes, local weather, aircraft type, (...)
2. **Stage-1**: cluster the actual routes, produce semantic-aware medoids 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-estim. (HMM: true, LR: synthetic) of the query FP, for prob.estim. or further refinement

Hybrid Clustering-Pred.Model (HMM, LR, DT, NN)



Input: flight plans (blue) and actual routes (red)

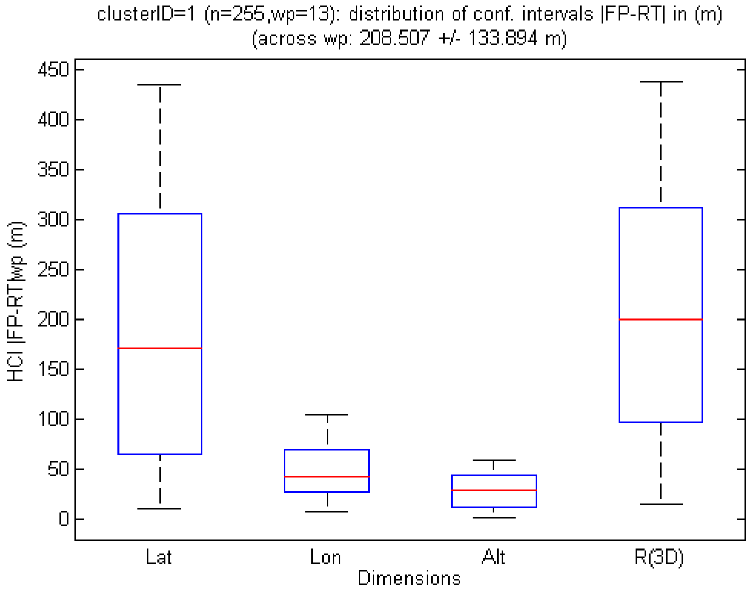


Output: semantic-aware cluster medoids (colored)

Step 1: clustering semantically annotated trajectories

- Input: Reference waypoints, matched with flight plans
- Each waypoint is enriched with semantics, i.e., local weather, aircraft type, etc.
- Use semantic-aware similarity metric, not just spatio-temporal track data
- Output: Semantic-aware cluster medoids

Hybrid Clustering-Pred.Model (HMM, LR, DT, NN)



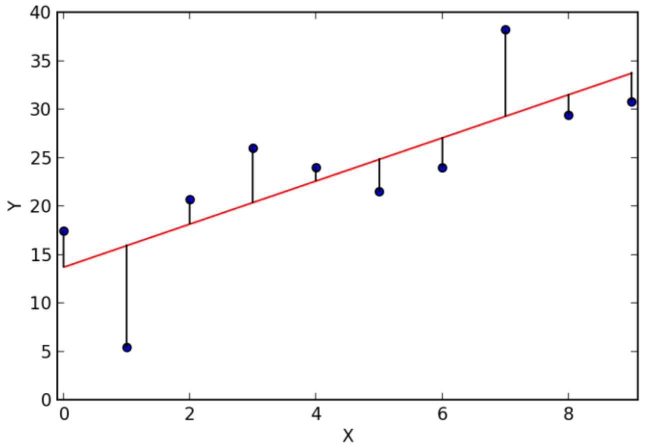
Step 2: building **HMM** for prediction

- <u>Input</u>: Enriched flight plans + medoid for each cluster, query FP
- <u>Output</u>: Max. likelihood estim. (emissions) of per-waypoint FP/RT deviations for query FP
- HMM approach enabled us to quickly estimate the confidence intervals (HWCI) of the accuracy, s.t. further experiments.

$$\mathrm{Route}(q)_j \approx \mathrm{FlightPlan}(q)_j - \mathrm{MeanDeviation}(k)_j \pm \mathrm{HWCI}(k)_j$$

** See: H. Georgiou et al. (2018), Semantic-aware aircraft trajectory prediction using flight plans. (submitted)*

Hybrid Clustering-Pred.Model (HMM, LR, DT, NN)



Step 2: building Linear Regr. for prediction

- Input: Enriched flight plans for each cluster, query FP
- Output: LSE-optimized estim. of per-waypoint value of predicted route
- The previous HMM approach is functionally a special case of LR, which is more generic.

$$\boxed{\mathrm{Route}(q)_j \approx \mathrm{FlightPlan}(q)_j * \mathrm{B}_j + \mathrm{B0}_j \pm \mathrm{HWCI}(k)_j'}$$

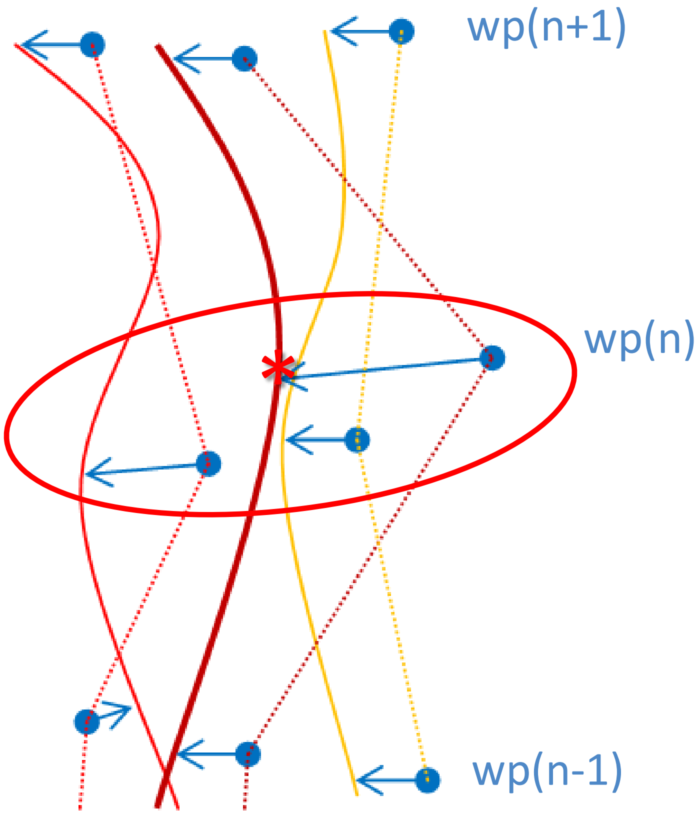
(DT: per-node/leaf model)

Note(1): Decision Tree (DT) regressor can be used instead or LR as replacement (e.g. CART).
Note(2): HWCI for LR error here is expected to be at most equal to the corresponding HWCI for HMM.

Hybrid Clustering-Pred.Model (HMM, LR, DT, NN)

DART

SESAR JOINT UNDERTAKING



Stage-2, HMM approach:

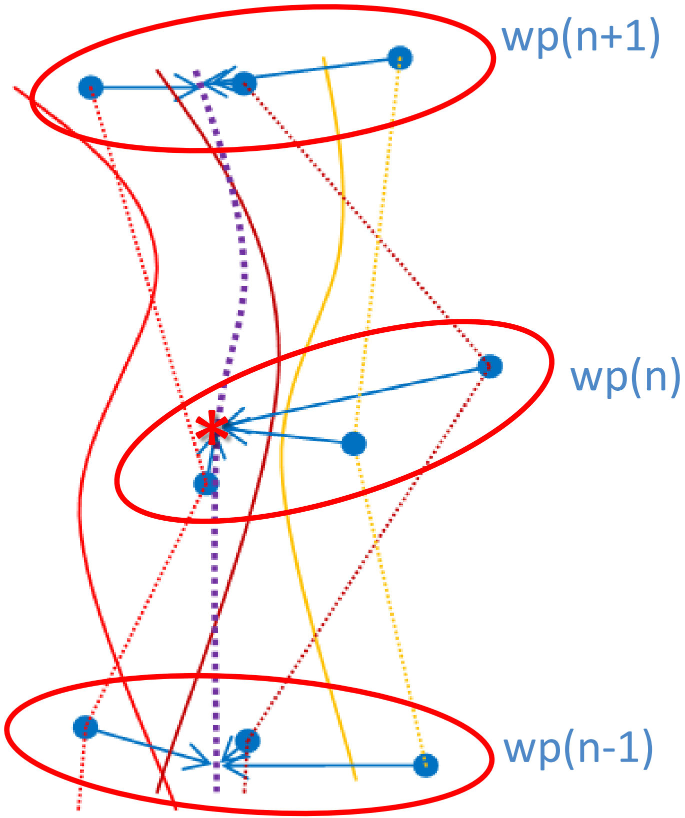
RT(n)-FP(n) statistics are used to build a probabilistic model (HMM emissions) for ref. point wp(n).

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 (HMM, LR, DT, NN)

DART

SESAR JOINT UNDERTAKING

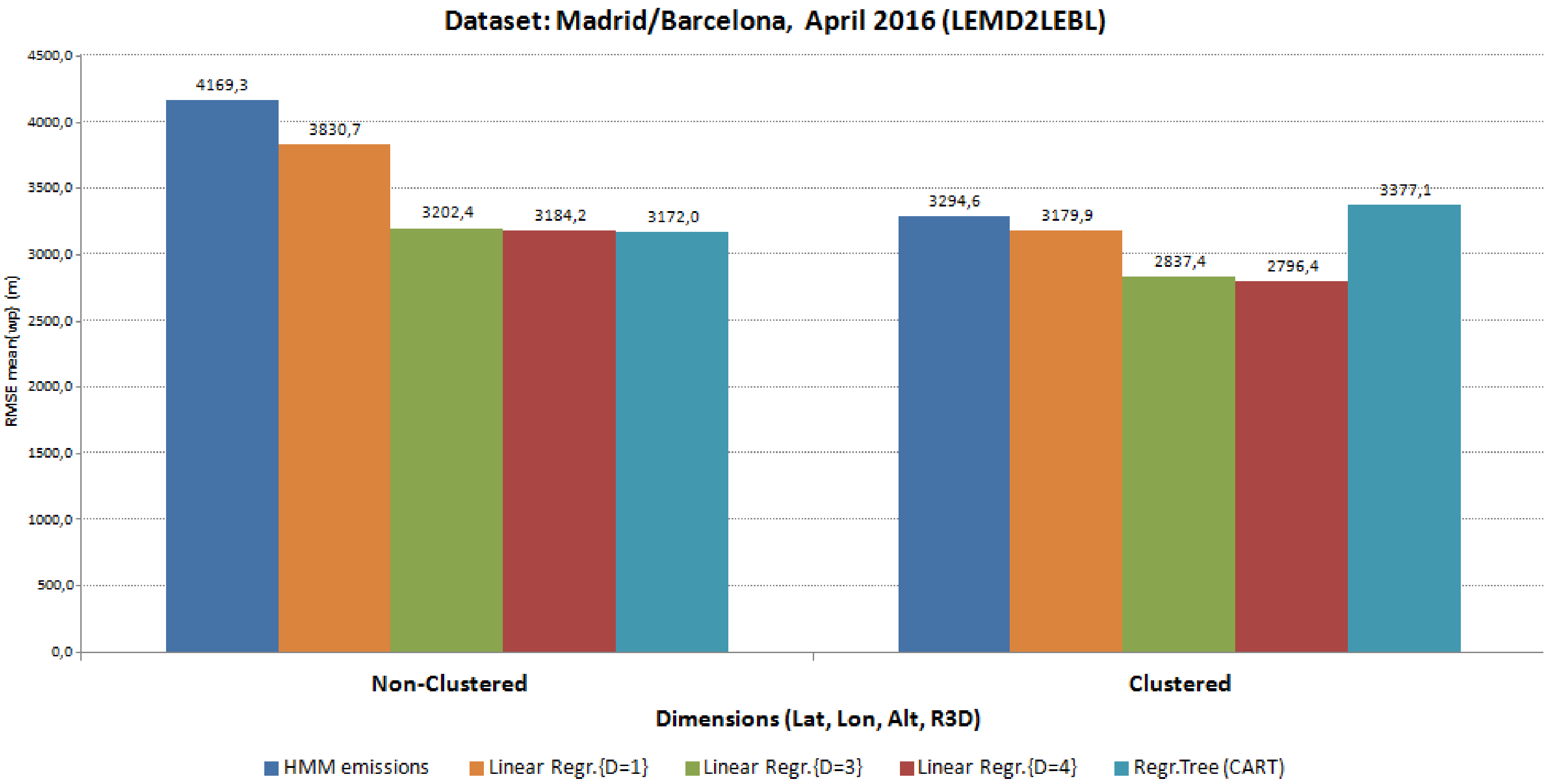


Stage-2, LR or DT approach:

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

Overview: TP performance tests



Non-linear Regressors (NN)

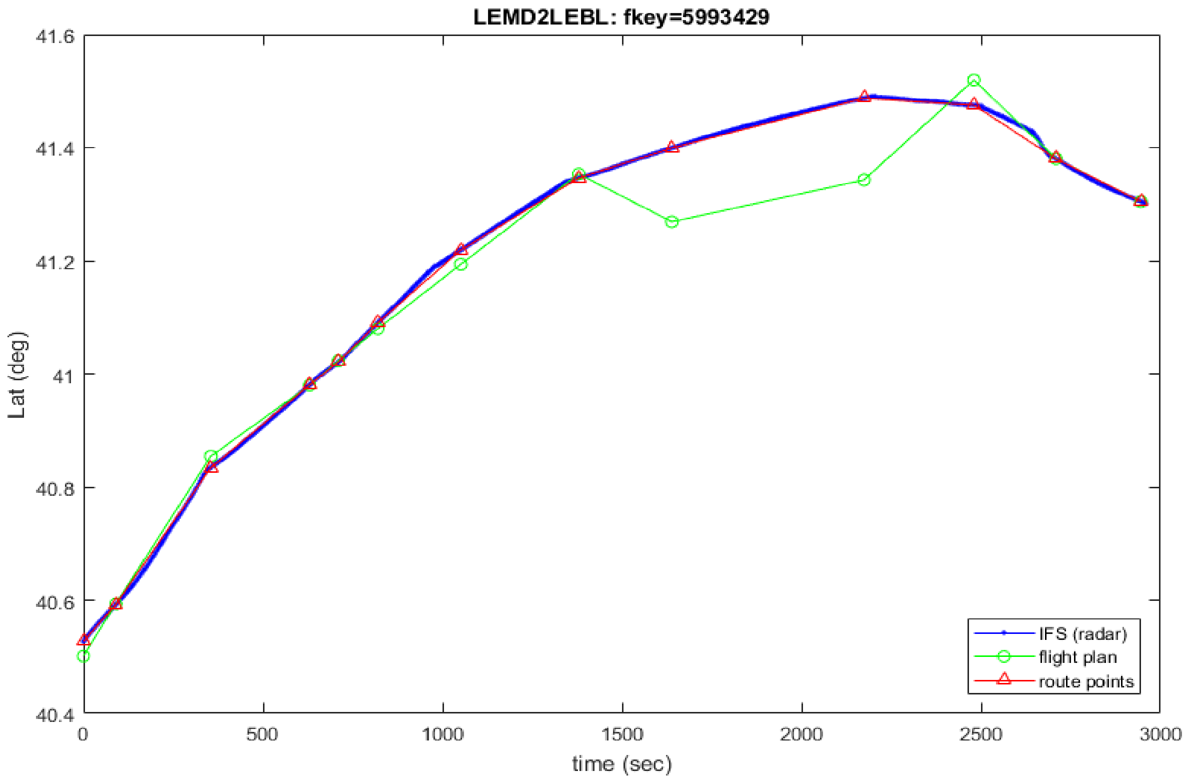
			NN/MLP/h10(BR): RMSE (m)			
CLUSTER	**Nk**	Output:	**RT(j)Lat**	**RT(j)Lon**	**RT(j)Alt**	**RT(j)R3D**
(*)	696	Input:				
CVk	**Nwp**	wp1	240,7	139,3	31,5	**279,9**
(90/10)	11	wp2	105,0	71,1	67,1	**143,5**
		wp3	1500,4	301,2	221,6	**1546,3**
		wp4	2813,7	1424,6	244,0	**3163,2**
		wp5	1709,7	1298,6	342,8	**2174,2**
		wp6	1519,2	670,9	235,9	**1677,4**
		wp7	1314,6	1744,8	252,3	**2199,1**
		wp8	2255,7	1808,6	172,7	**2896,4**
		wp9	1928,0	832,7	189,5	**2108,7**
		wp10	2033,6	764,4	215,5	**2183,2**
		wp11	1691,6	1504,9	269,9	**2280,1**
		mean:	1555,7	960,1	203,9	**1877,4**
		stdev:	798,6	632,4	88,7	**943,4**

LEMD2LEBL (CV.k=10): input=FP(3D)+AP → output=Lat/Lon/Alt (NN)

Advanced pre-processing: Wiener filtering

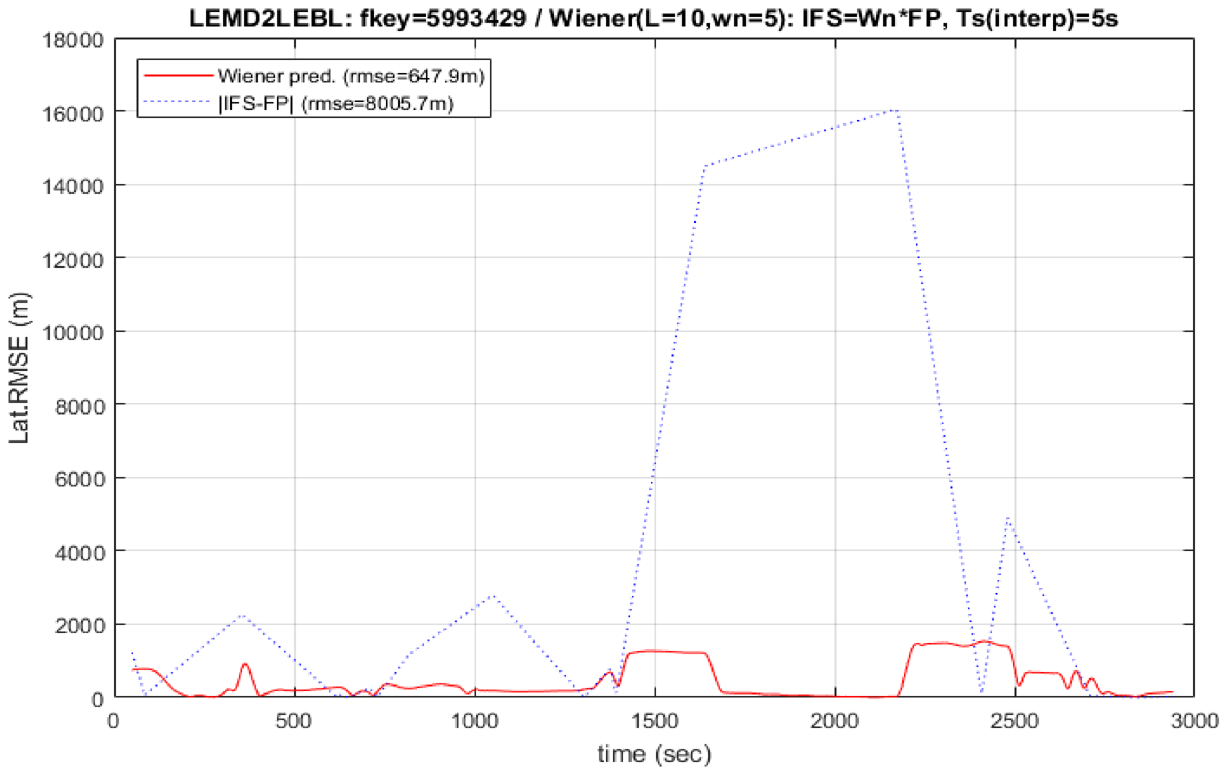
DART

SESAR JOINT UNDERTAKING



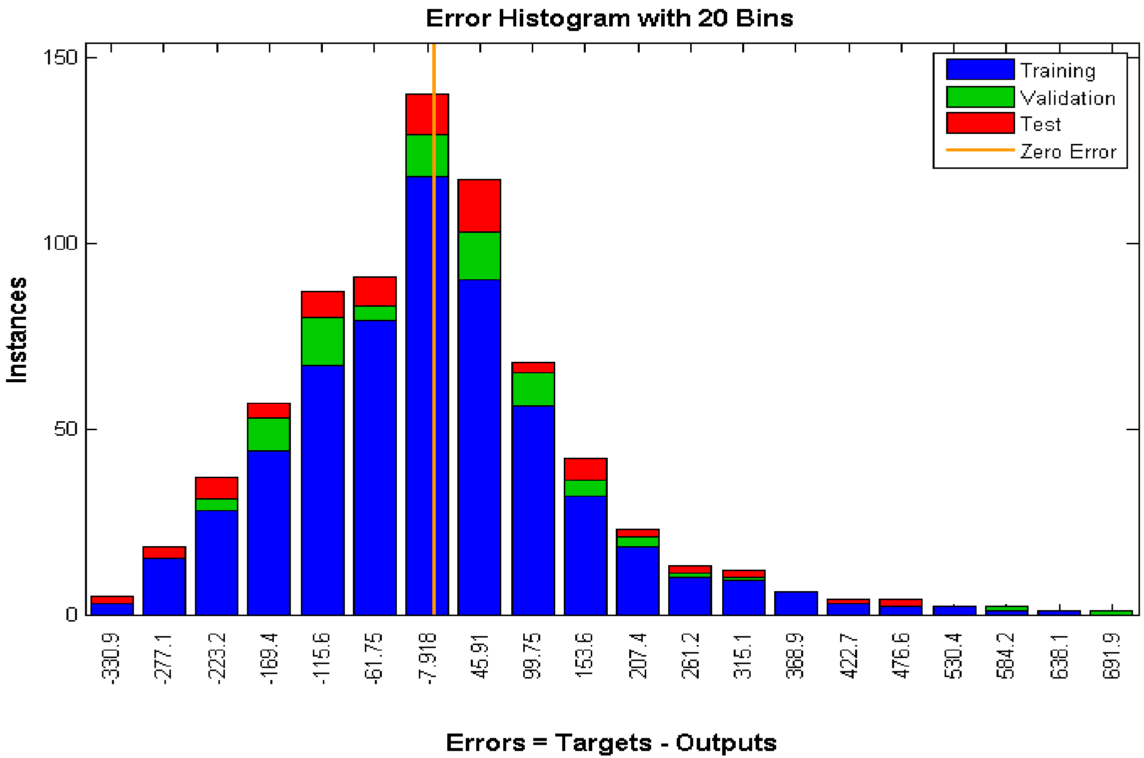
TP-C: Close-up view of a FP/RT data sample (single flight), Lat-only

Advanced pre-processing: Wiener filtering



TP-C: Wiener forward predictor for FP/RT (single flight), Lat-only

Special predictive modeling: ETA (aviation)



NN/MLP: ETA pred. errors on FP/RT dataset (top-20 of 217 stat.feat.)

DATA STORIES

Data Science Lab. @ University of Piraeus

UNIVERSITY OF PIRAEUS RESEARCH CENTER

datAcron

Big Data Analytics for Time-Critical Mobility Forecasting

The EU Framework Programme for Research and Innovation

HORIZON 2020

Grant Agreement No: 687591

Objective	Current Performance	Target Performance
Computation of data synopses	<95% compression	>95% data compression without harming the quality of analytics results
Real-time trajectory reconstruction	Offline trajectory reconstruction	Real-time computations
Efficient large-scale mobility data analytics	Gb size of datasets	Tb size of datasets
Real-time trajectories forecasting for ATM and maritime, resp.	Short forecasting horizon, depending on current speed of airplane/vessel.	Increase the accuracy in positional predictions by reducing the standard deviation of the positional error in prediction.

Task 2.2 Prediction components

1. **Future Location Prediction (FLP)**:
   - **Short-term (FLP-S)**: Online, based only on recent positions (time series), look-ahead time is a few minutes at most.
   - **Medium/Long-term (FLP-L)**: Online, based on recent positions and history (routes network), look-ahead time is up to entire trips (end-to-end).
2. **Trajectory Prediction (TP)**:
   - *Long-term, Unconstrained (TP-U): "Only the starting point is specified"*

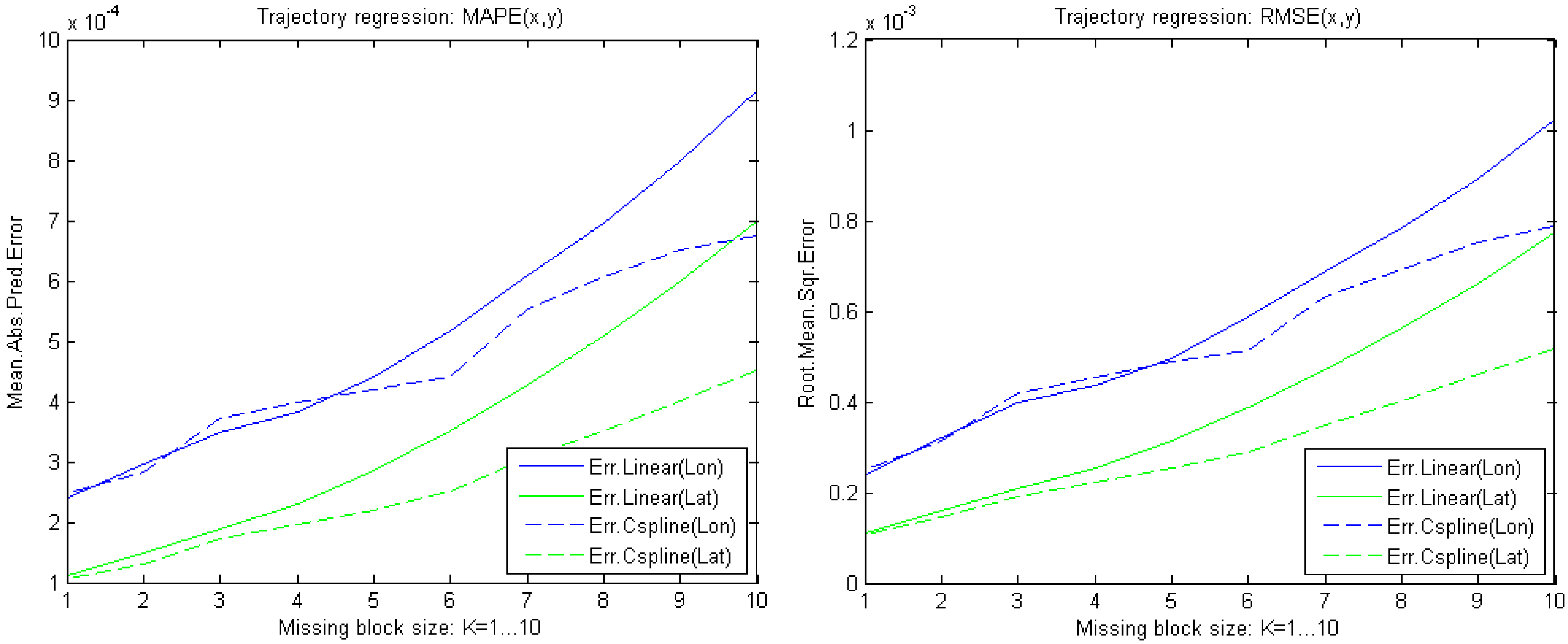
     ⇒ Addressed via FLP-L (routes network).
   - *Long-term, Destination-only (TP-D): "Staring and ending points are specified"*

     ⇒ Addressed via FLP-L (routes network).
   - **Long-term, Constraints-based (TP-C)**: *"Several reference points are specified"*

     ⇒ Offline/batch-only, exploiting complete & enriched flight plans.

Task 2.2 Prediction components

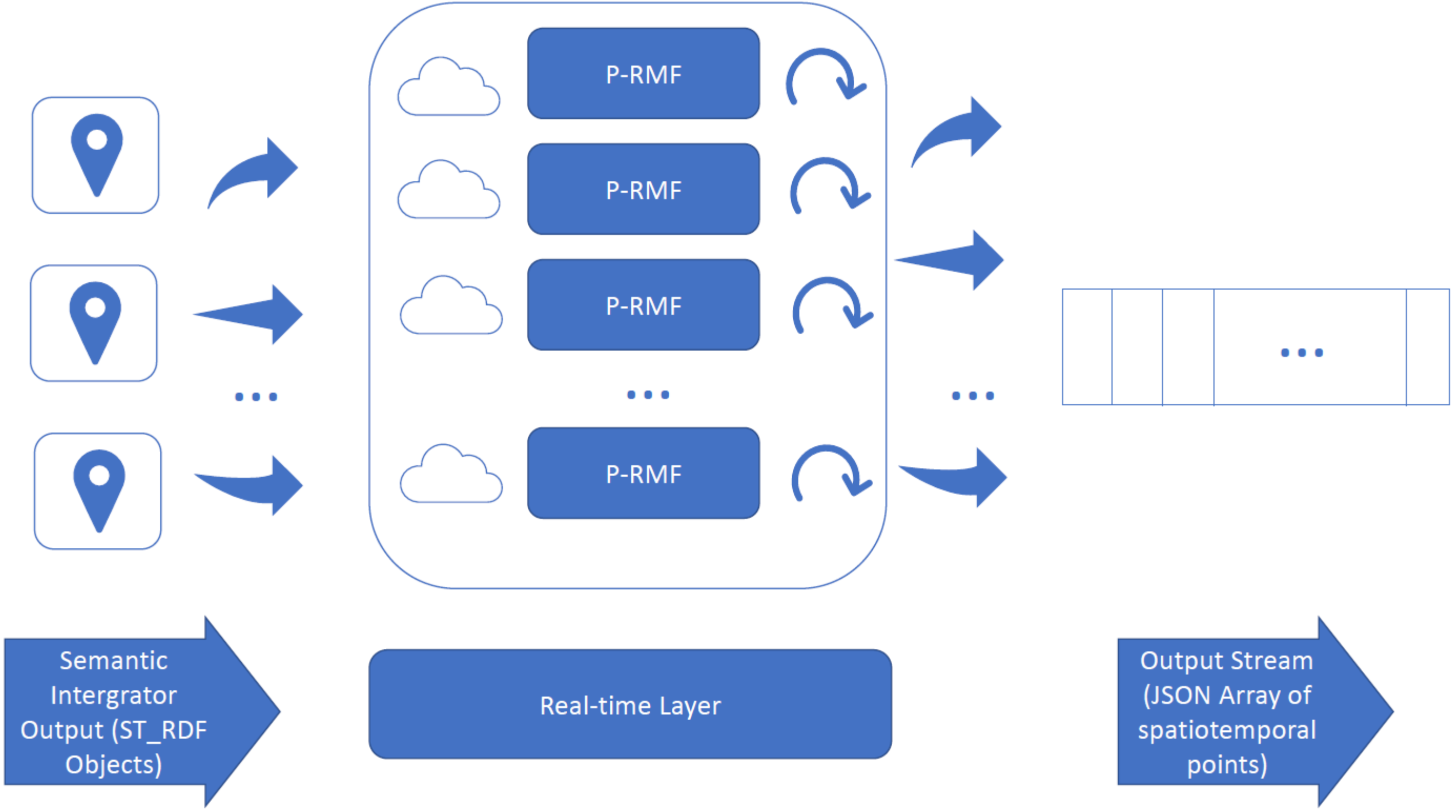
1. **Distributed online FLP (FLP-S):** Extension of the RMF algorithm to a distributed P-RMF* for short-term FLP *(Petrou et.al. 2018)*
   - Using surveillance data + predefined motion patterns
   - Improvement: Bypass the problems of original RMF on real data
2. **Distributed online Medium/Long-term (FLP-L)**
   - Designed and at a good level of implementation
3. **Semantically-driven TP (TP-C):** Novel hybrid clustering/LR-based method for TP *(Georgiou et al. 2018)*
   - Using enriched trajectories + prior constraints (e.g. flight plans) for use on TP-C tasks
   - Improvement: HMMs replaced by more efficient Linear Regressors



FLS-S: Close-up view of extrapolation errors, horizontal-only

datAcron

European Commission

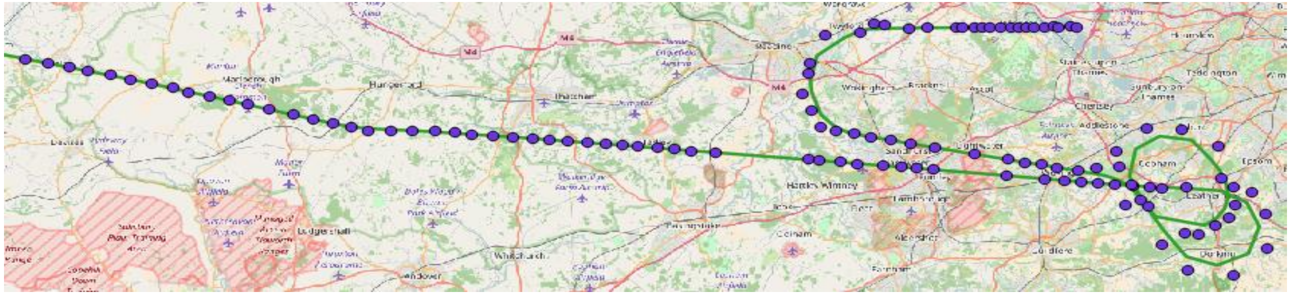


FLP-S (a.k.a P-RMF*) architecture

P-RMF* in Action (aviation, London landing)

Flight Aware data (London): Id= 10, Sampling rate= 8sec, Horizon= 8

a) actual flight landing

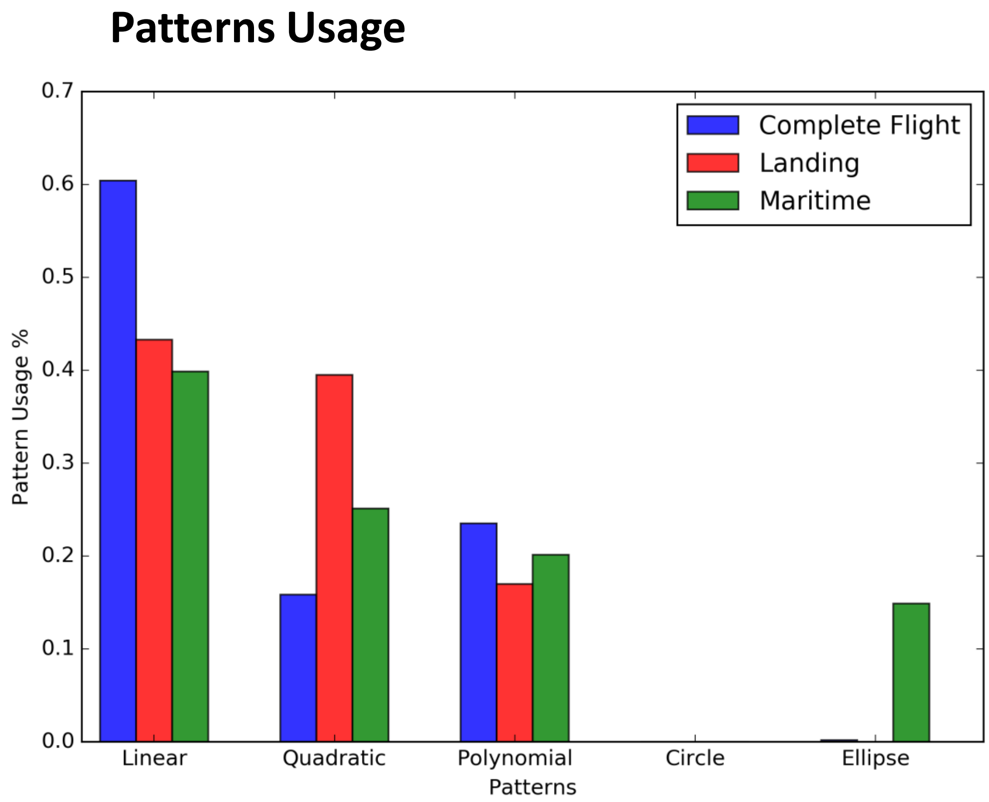


b) predicted future point (H=8)

datAcron

European Commission

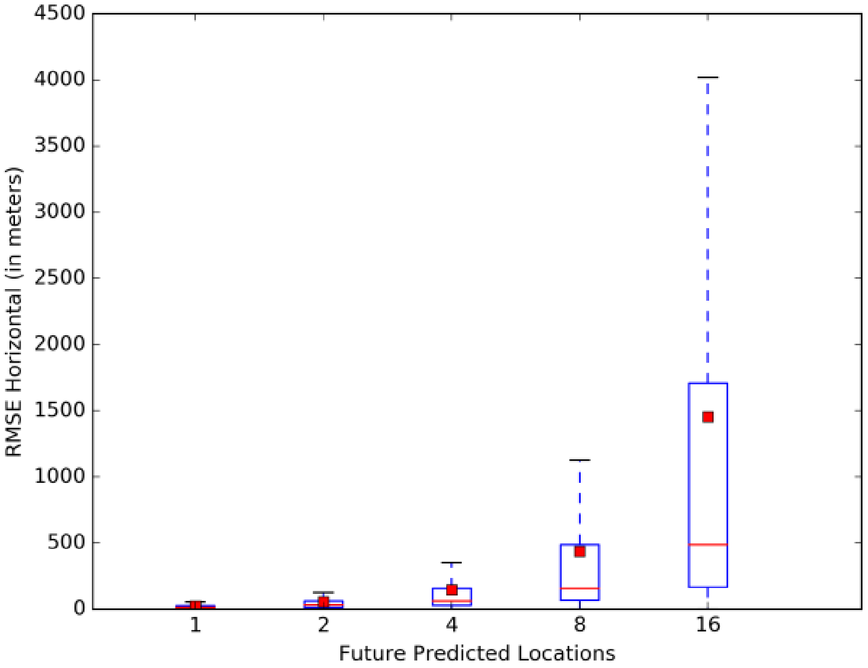
P-RMF* in Action (aviation, maritime)

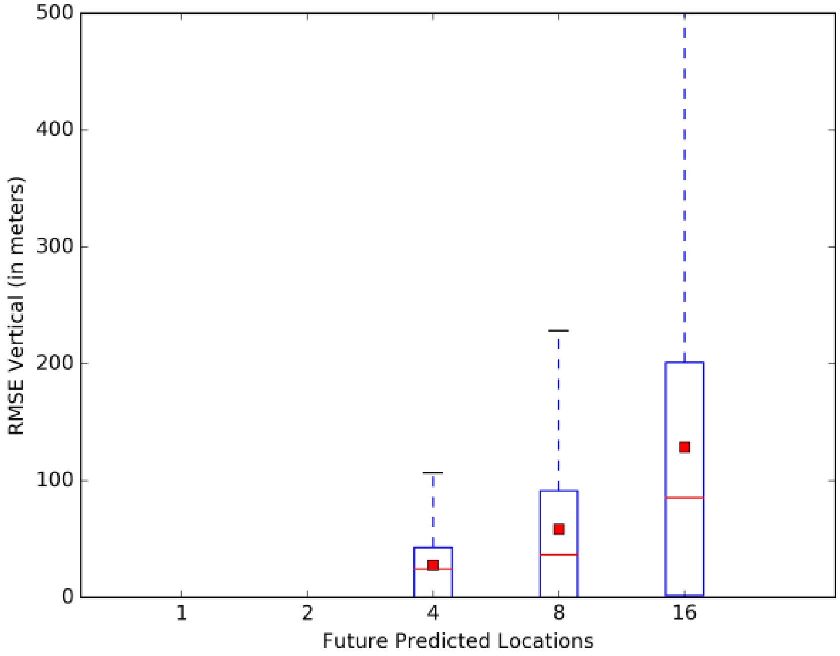


datAcron

European Commission

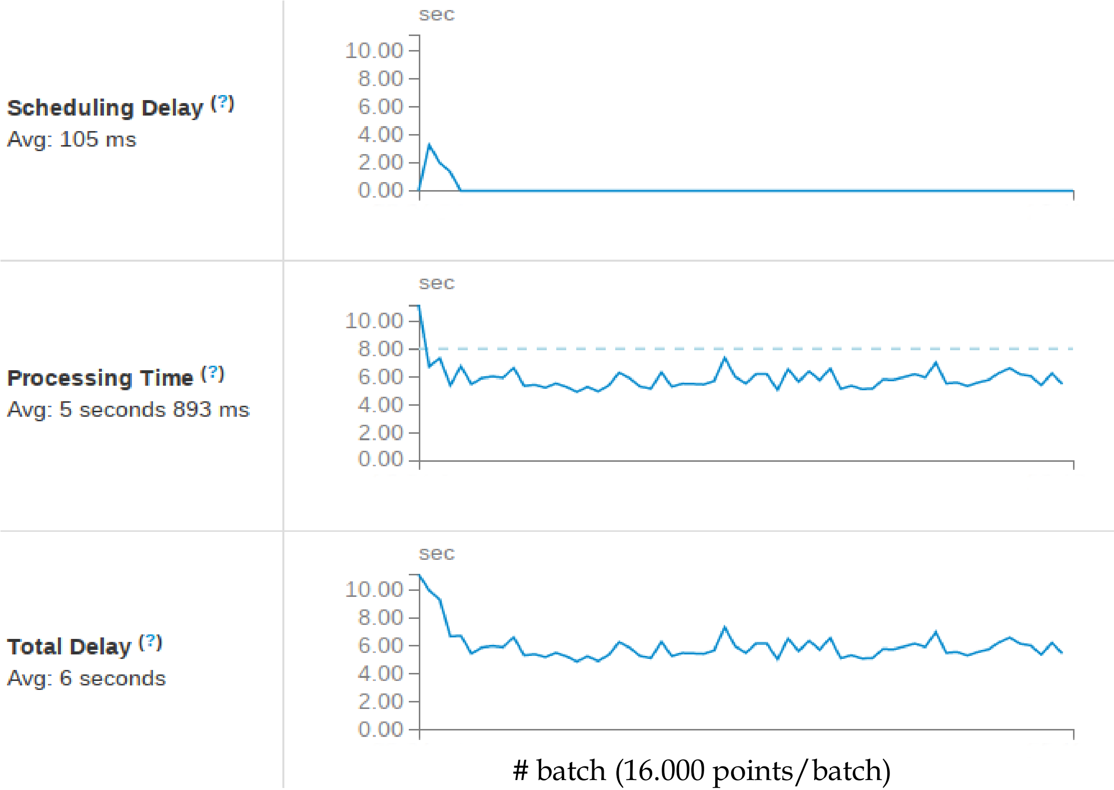
FLP-S: LEMD/LEBL (IFS radar)





FLP-S: Horizontal & vertical error for **complete flights** with sampling rate 4 seconds

FLP-S: Task Execution Time Based on Batch Interval

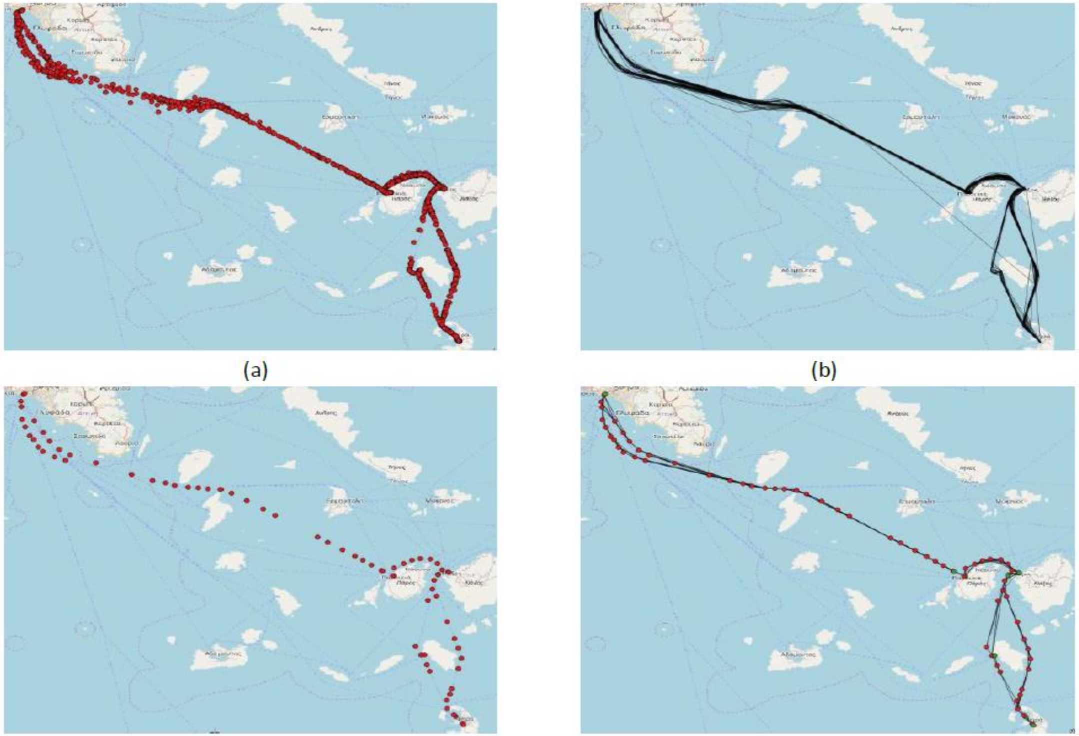


batch (16.000 points/batch)

FLP-S: Performance metrics for $18 \cdot 10^6$ points, 4000 points/sec processing in 8 secs

datAcron

European Commission



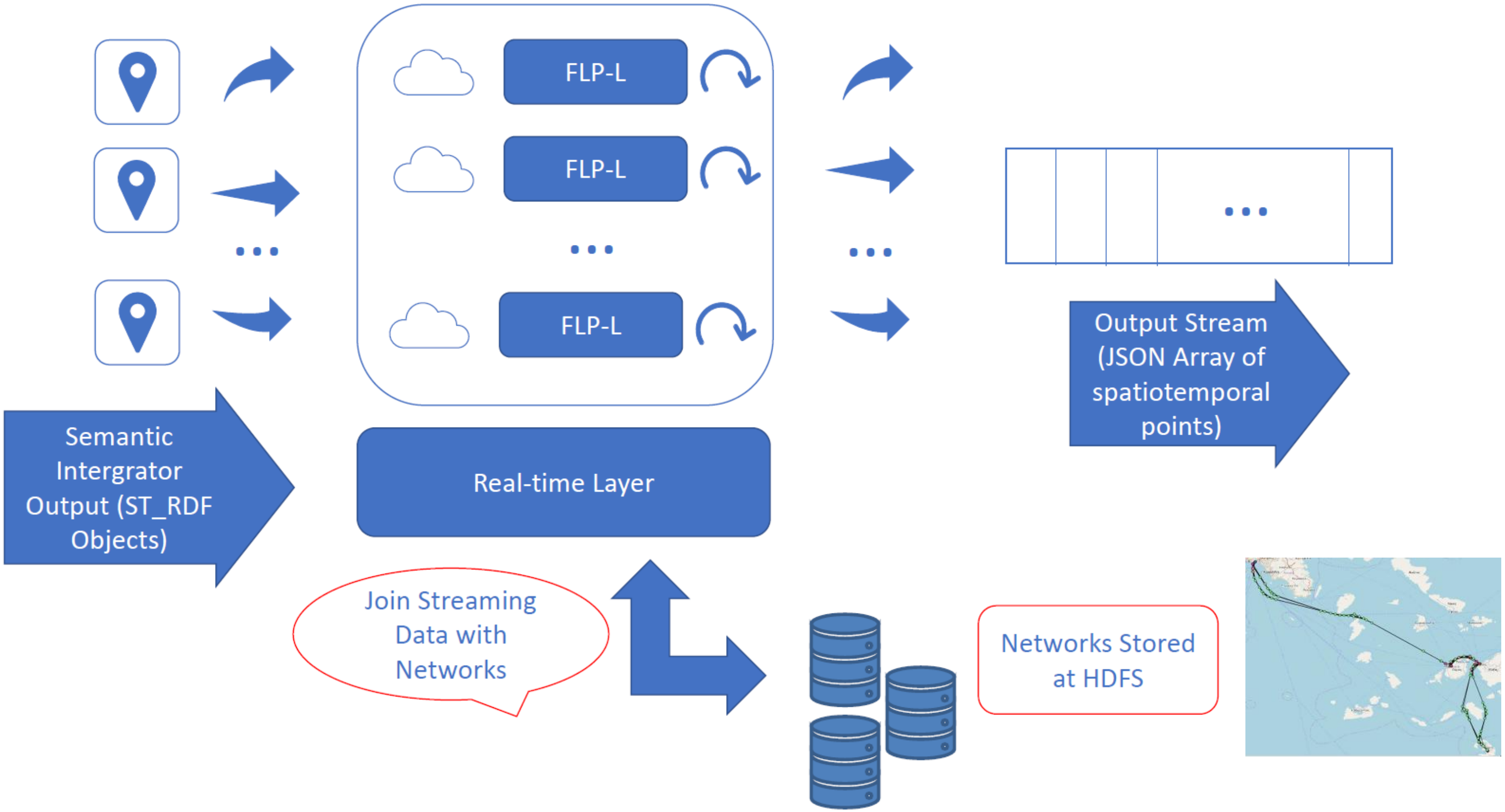
(c) (d)

Figure 27 Overview of Semantic-aware Network inference solution in maritime domain: (a) all enriched points from input, (b) semantic trajectories formed from enriched points (c) semantic nodes extraction (d) semantic path discovery.

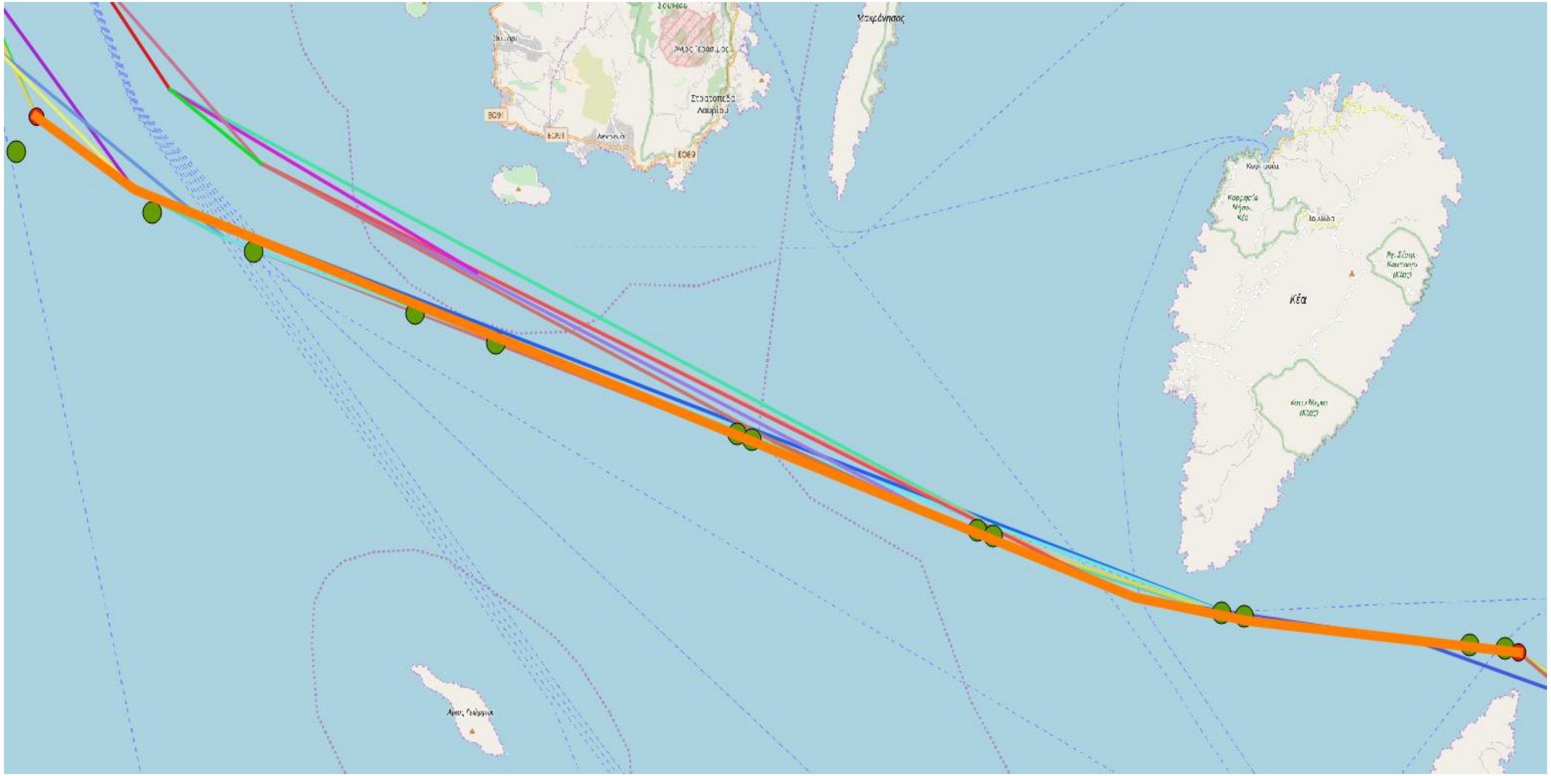
FLP-L: Routes discovery & network construction

datAcron

European Commission

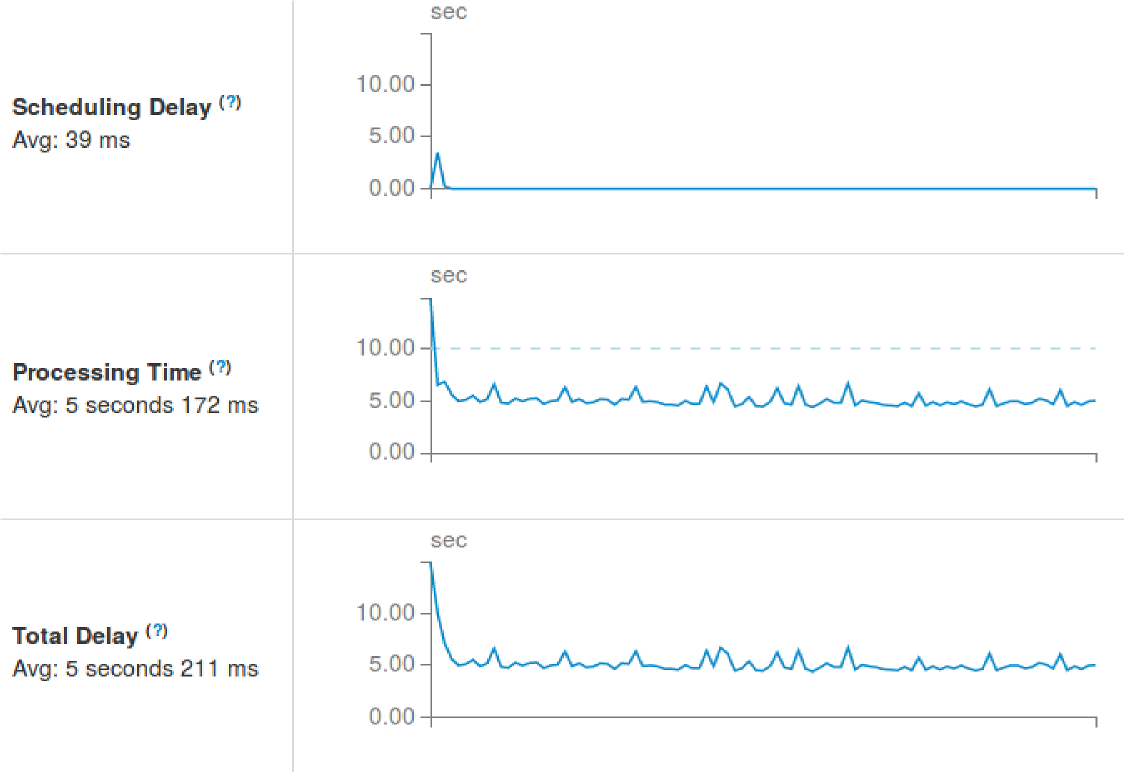


FLP-L architecture

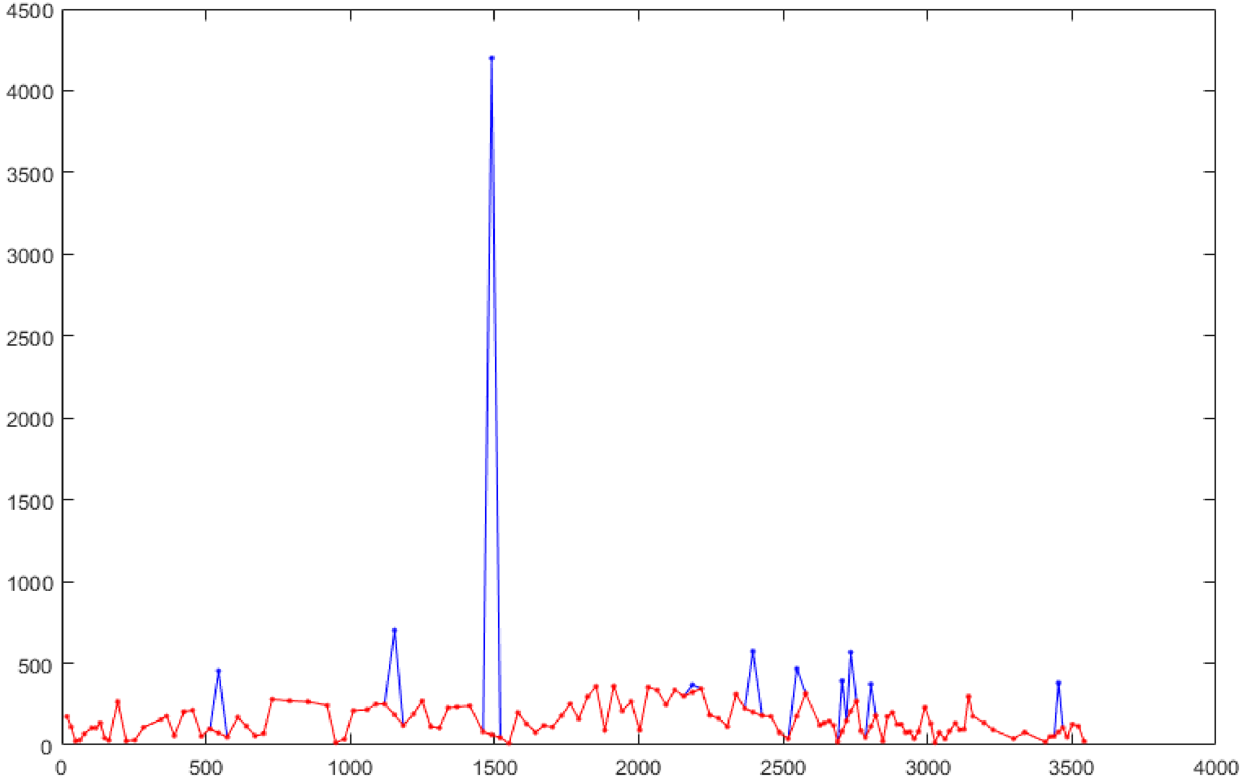


FLP-L: Red left = knn-based “start”, Red right = knn-based “end”, orange line = matched path

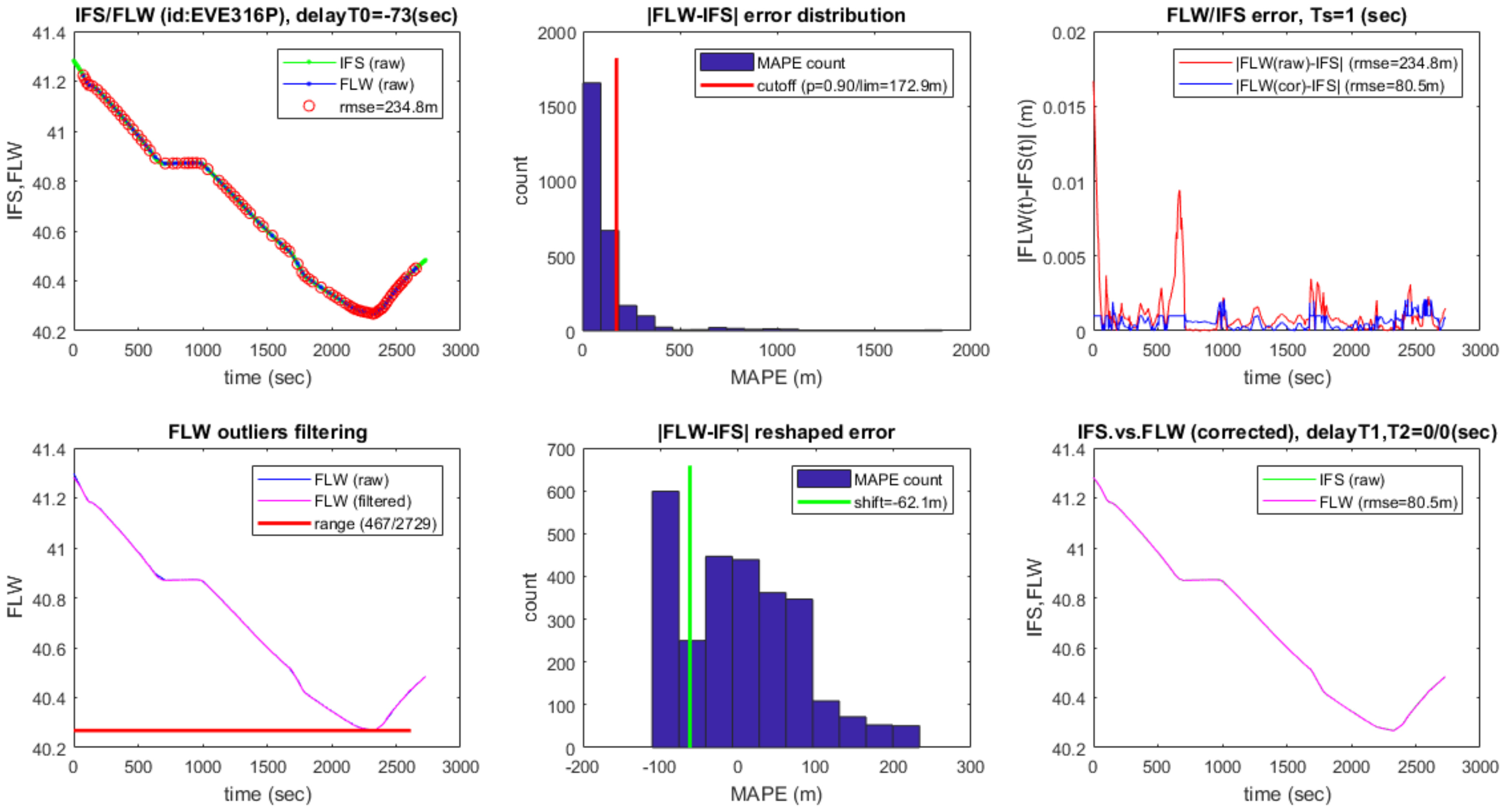
FLP-L: Task Execution Time Based on Batch Interval



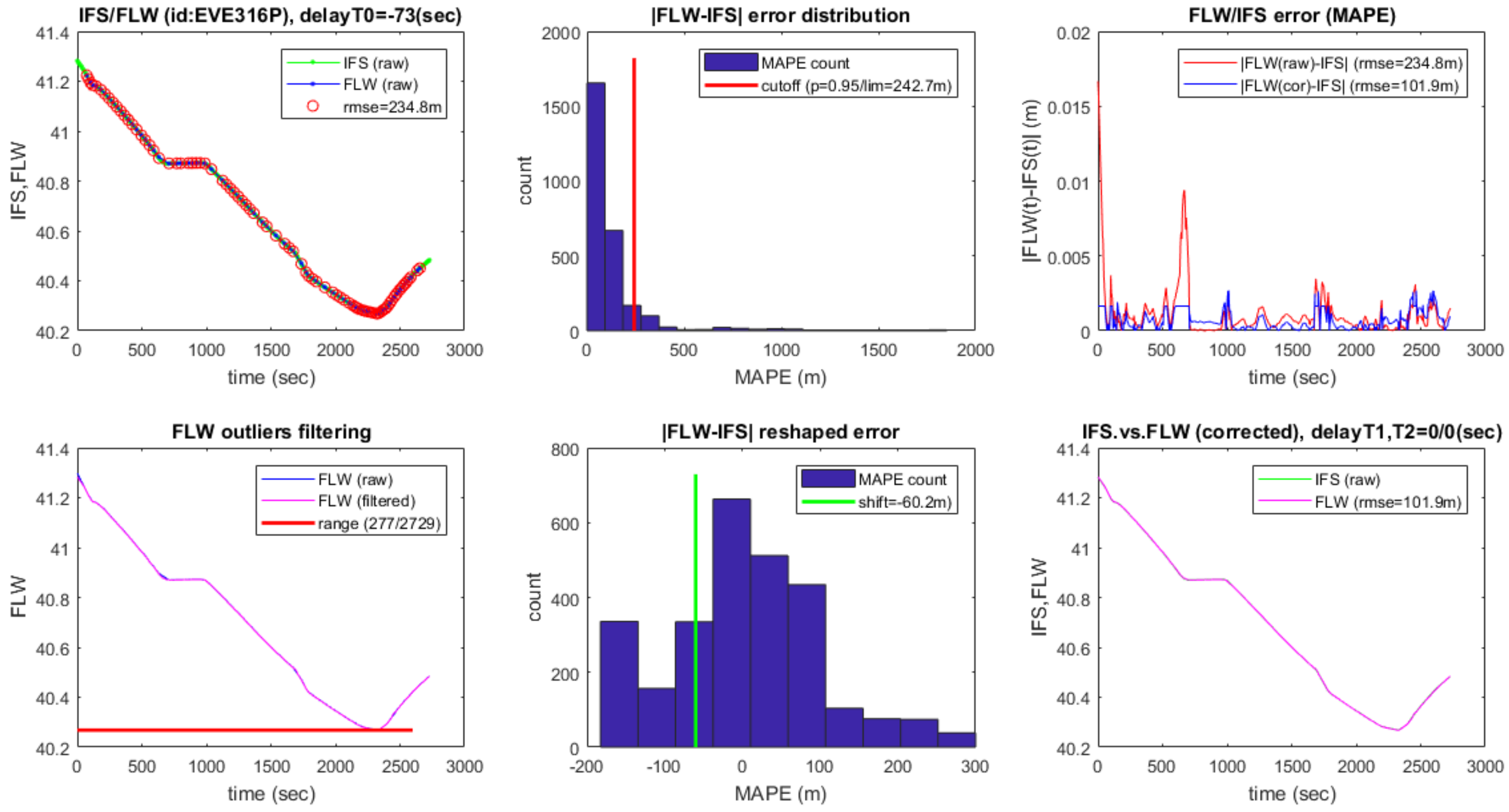
FLP-L: Performance metrics for $25 \cdot 10^6$ points, 6000 points/sec processing in 10 secs



Cross-streaming: IFS/FLW deviations example (blue=orig./red=filtered)



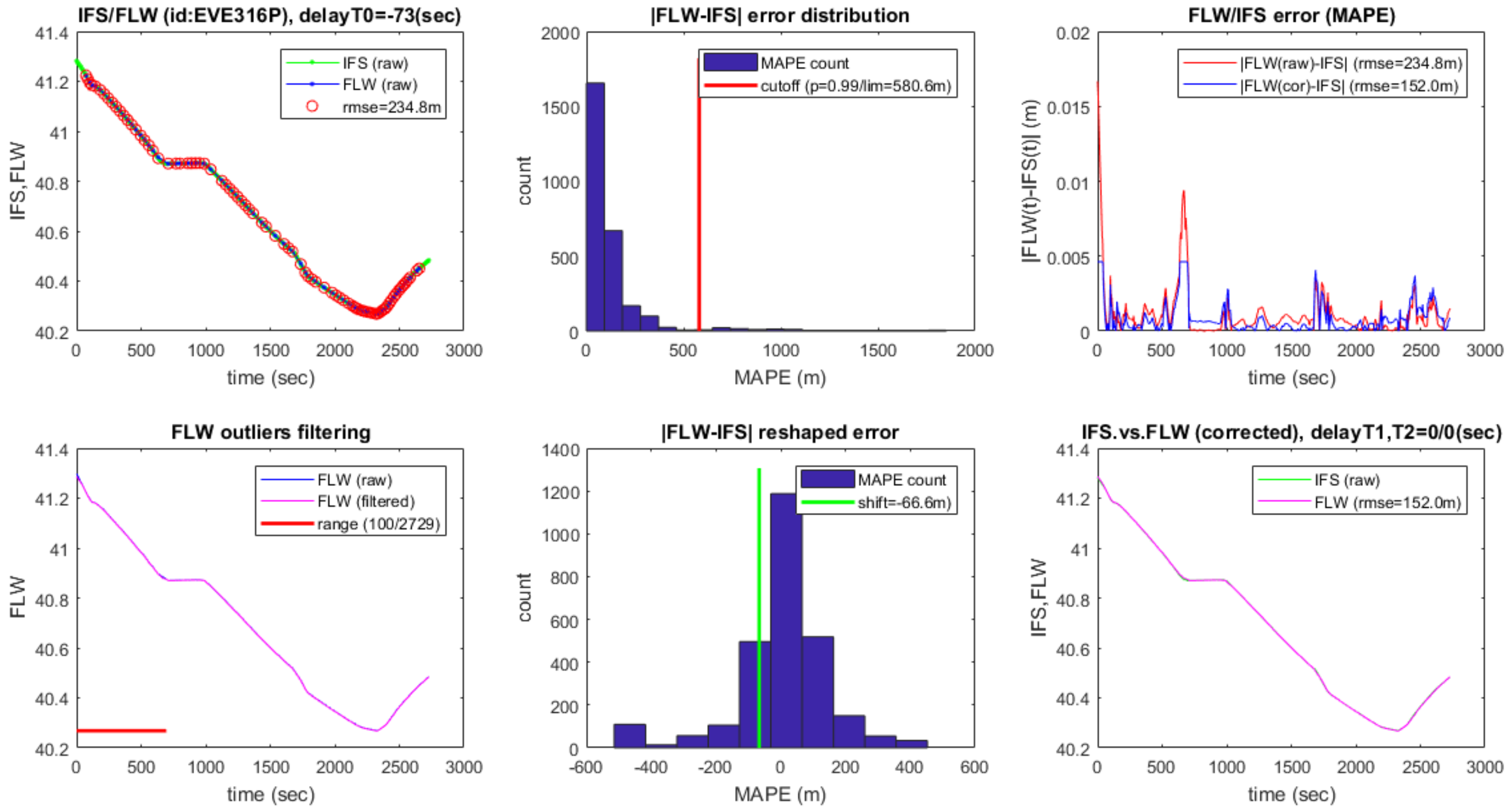
Cross-streaming: IFS/FLW statistical filter design (“strict”, p=0.90)



Cross-streaming: IFS/FLW statistical filter design ("medium", p=0.95)

datAcron

European Commission



Cross-streaming: IFS/FLW statistical filter design ("relaxed", p=0.99)

datAcron

European Commission

Data Science Lab. @ University of Piraeus

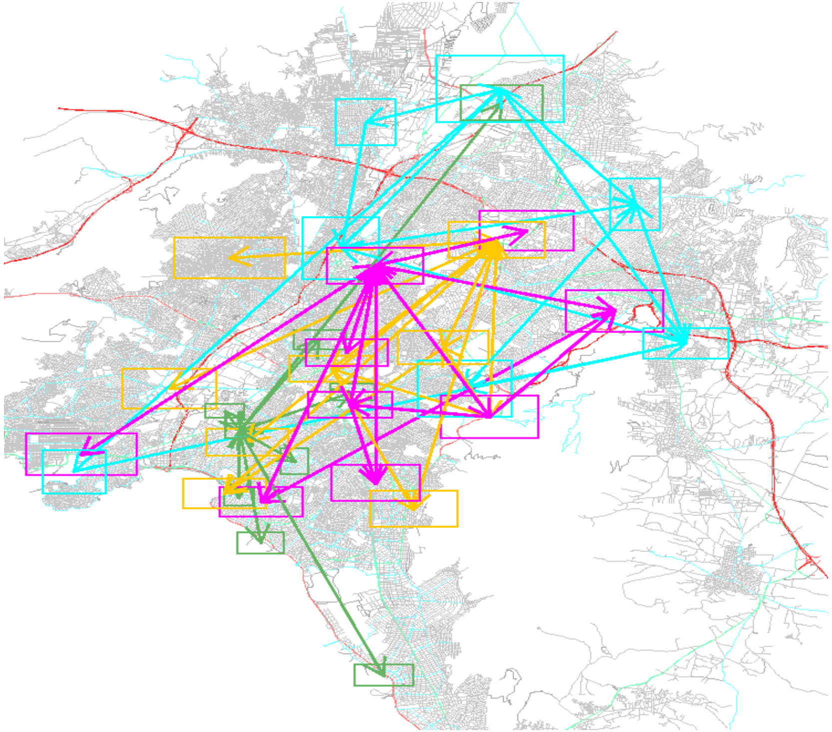
This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement No 780754.

DATA STORIES
Data Science Lab. @ University of Piraeus

UNIVERSITY OF PIRAEUS RESEARCH CENTER

Track & Know

- Funded by the European Union's Horizon 2020
- Consortium of 14 partners from 9 different countries
- Research, develop and exploit a new software framework to increase efficiency of Big Data
- Applications in transport, mobility, motor insurance and health sectors
- Develop user-friendly toolboxes that will be readily applicable
- Validation in real-world pilots



Track & Know

Big Data KPIs

Performance KPI	*Target Value*
Time-to-realization reduction	40% - 60%
Query time reduction	>15%
Data Load balancing time reduction for distributed query execution	>5%
Data Load balancing volume reduction for distributed query execution	>25%
Unstructured Data processing time reduction	>10%
Structured Data processing throughput improvement	>20%
Big Mobility Data pattern detection improvement	>30%
Big Mobility Data forecasting accuracy improvement	>40%
Complex Event Recognition processing improvement	>8%
Visualisation processing time reduction for interactive Mobility Data	>15%
Visualisation processing time reduction for aggregated Mobility Data	>10%

UPRC – WP4 planned work (M10+)

T4.1: “Analytics for mobility patterns detection for forecasting”

- Analysis methods & tools over Big Mobility Data
  - Cluster analysis → two types of clustering algorithms
  - Motion pattern detection → pattern-based short-term FLP, routes network for long-term FLP
  - Exploit enriched & integrated data sources → clustering & routes networks will exploit data enrichments
- Short/Long forecasting
  - of routes, flows, concentration nodes → routes networks, hot-spot analysis
  - contextual properties, outliers detection → special predictors (turns, node-ETA), “enriched” trajectory synopses

⇒ Also, from toy scenarios: driving profiles, fuel consumption, etc

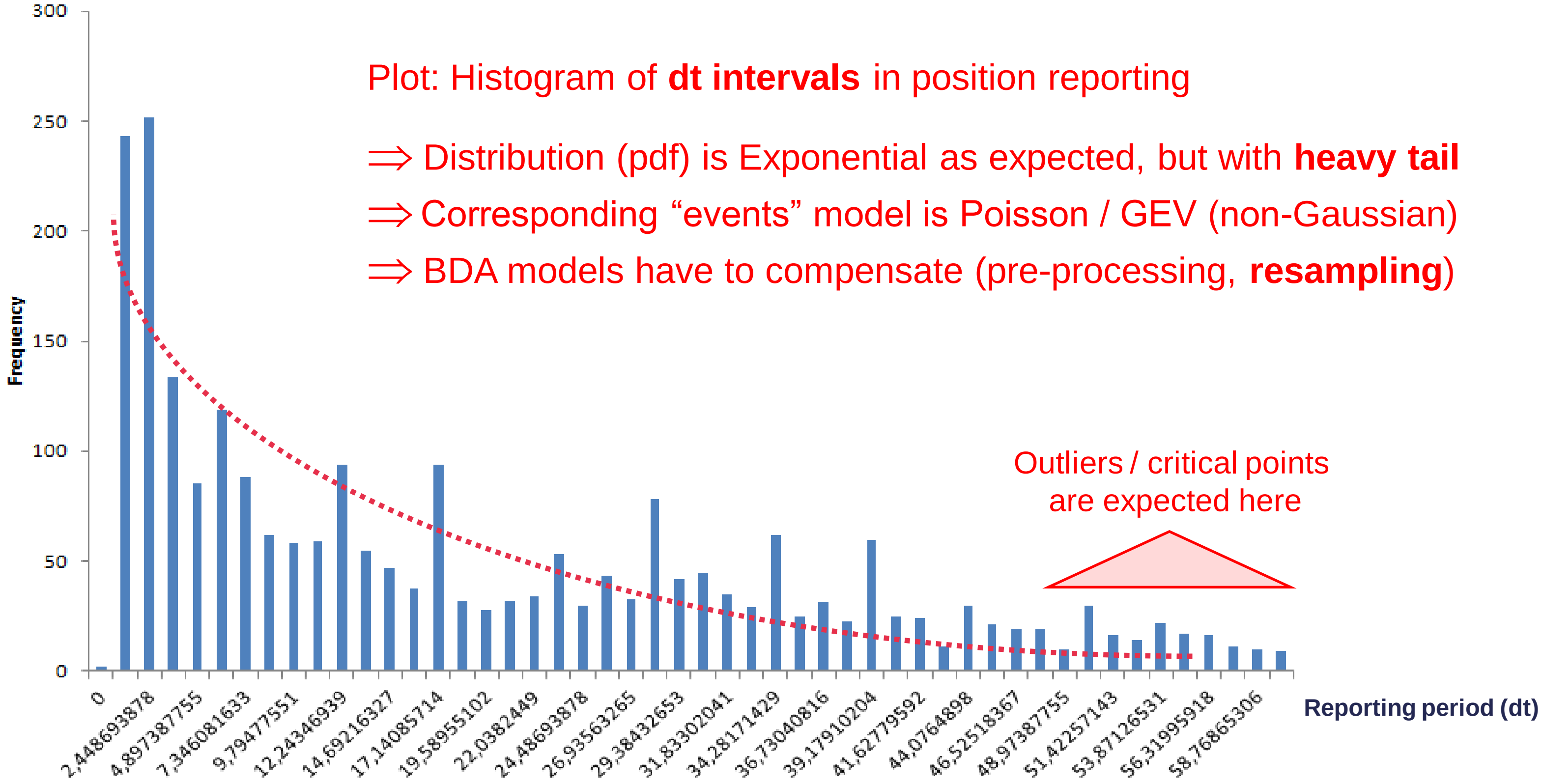
UPRC – WP4 status update (M11)

Plot: Histogram of **dt intervals** in position reporting

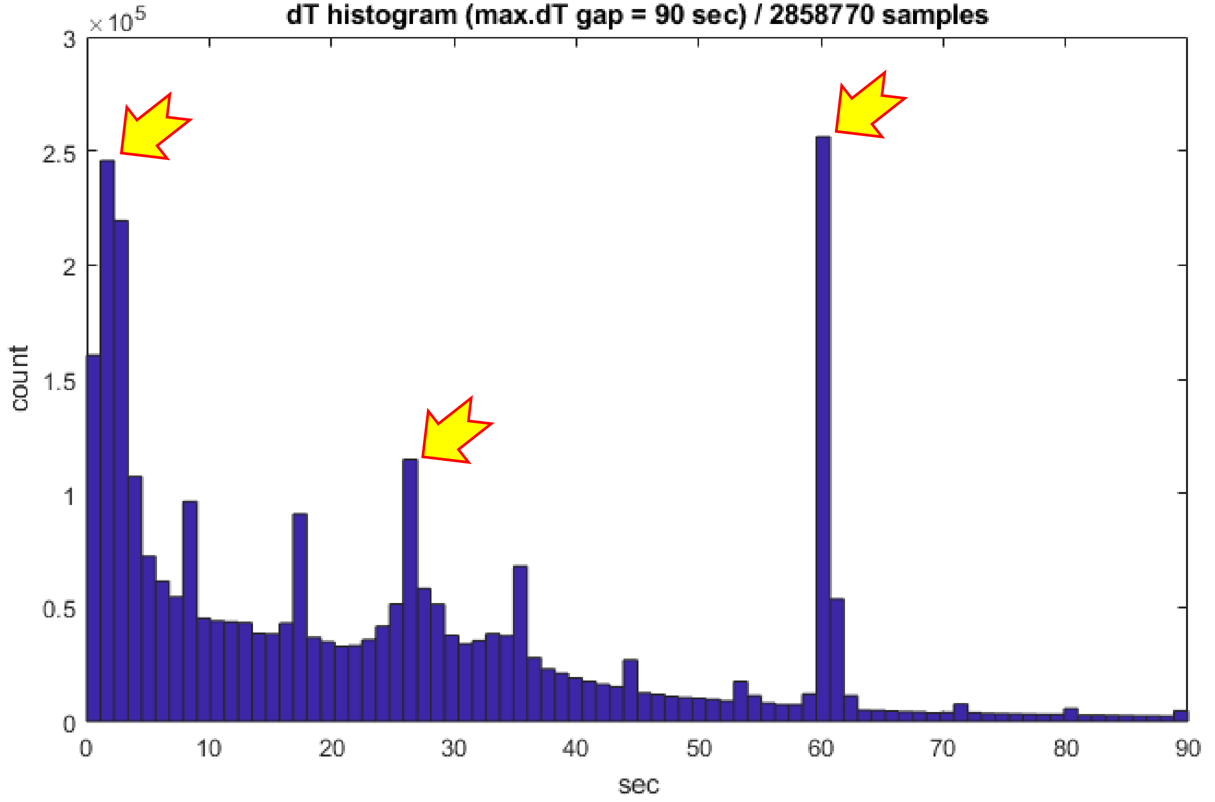
⇒ Distribution (pdf) is Exponential as expected, but with **heavy tail**

⇒ Corresponding "events" model is Poisson / GEV (non-Gaussian)

⇒ BDA models have to compensate (pre-processing, **resampling**)

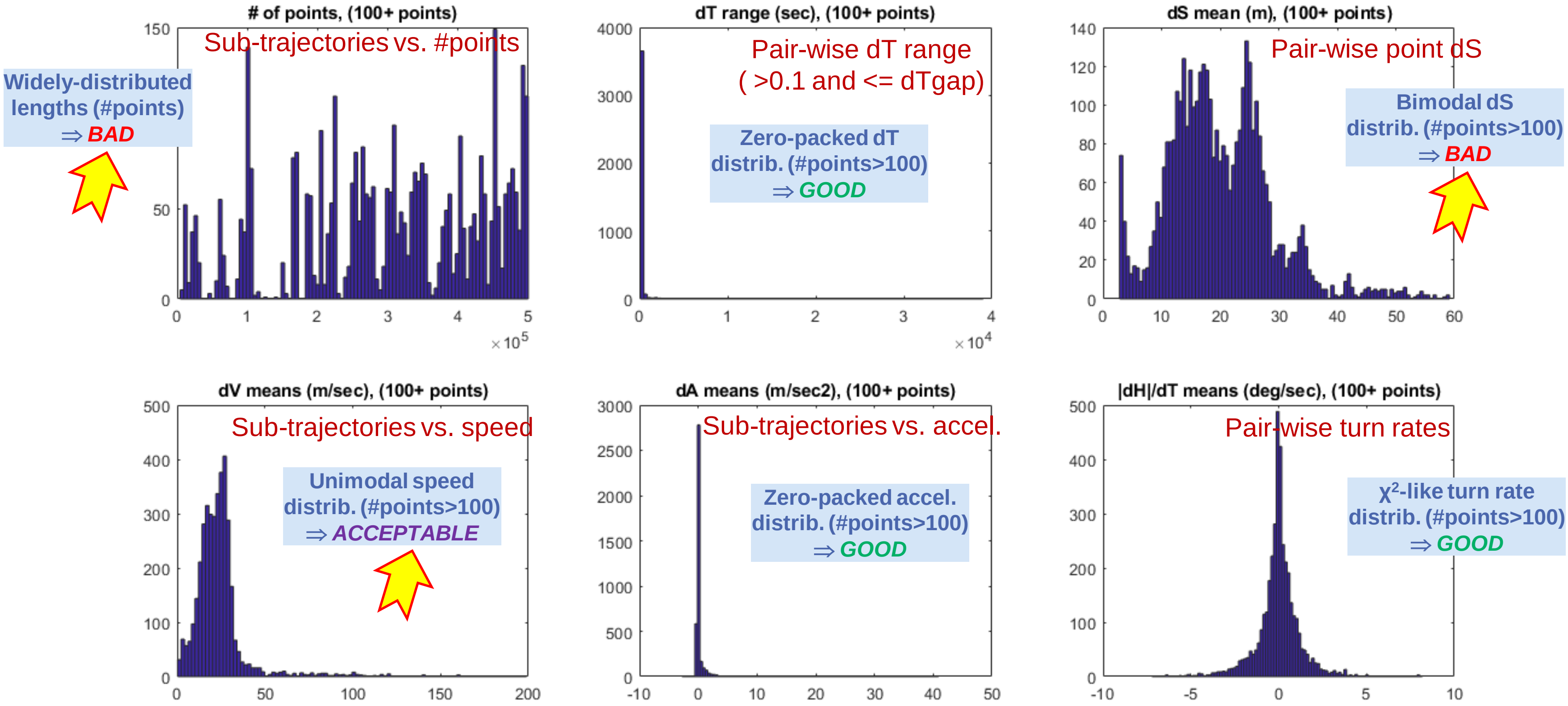


UPRC – WP4 status update (M11)



Multiple peaks in the dT distribution (right tail trimmed) ⇒ ***BAD***

UPRC – WP4 status update (M11)



Data Science Lab. @ University of Piraeus

- EU- H2020 ICT/ **BigDataStack**:
  High-performance Data-centric Stack for Big Data Applications and Operations [www.bigdatastack.eu/], 2018-20
- EU-H2020 ICT/**Track and Know:**
  Big Data for Mobility Tracking Knowledge Extraction in Urban Areas
  [trackandknowproject.eu/], 2018-20
- EU- H2020 ICT/ **datAcron**:
  Big Data Analytics for Time Critical Mobility Forecasting [datacron-project.eu/], 2016-18
- EU- H2020 SESAR/ **DART**:
  Data-driven Aircraft Trajectory Prediction Research [dart-research.eu], 2016-18
- GR/ **RoadRunner**:
  Scalable and Efficient Analytics for Big Data [platforms.gr/roadrunner/], 2014-15
- EU- FP7 Marie Curie/ **SEEK**:
  Semantic Enrichment of Trajectory Knowledge Discovery [www.seek-project.eu], 2012-15
- EU- FP7 ICT/ **DATASIM**:
  Data Science for Simulating the Era of Electric Vehicles [www.uhasselt.be/datasim], 2011-14
- EU- FP7 Marie Curie/ **CloudIX**:
  Cloud-based Indexing and Query Processing [research.idi.ntnu.no/cloudix/], 2011-13

Data Science Lab. @ University of Piraeus

Big (spatial-temporal) data management & query processing

- P. Tampakis et al. (2018) Distributed trajectory join processing using MapReduce. Submitted.
- C. Doulkeridis et al. (2017) Parallel and distributed processing of spatial preference queries using keywords. Proceedings of EDBT.
- F. Gryllakis et al. (2017) Searching for spatio-temporal-keyword patterns in semantic trajectories. Proceedings of IDA.
- M. Saouk et al. (2016) Efficient processing of top-k joins in MapReduce. Proceedings of Big Data.
- S. Sideridis et al. (2016) On querying and mining semantic-aware mobility timelines. Int. J. Data Science and Analytics, 2(1).
- D. Pertesis & C. Doulkeridis (2015) Efficient skyline query processing in SpatialHadoop. Information Systems, 54(C).
- C. Doulkeridis & K. Nørvåg (2014) A survey of large-scale analytical query processing in MapReduce. VLDB Journal, 23(3).

Big (spatial-temporal) data analytics & mining

- H. Georgiou et al. (2018) Predicting the next steps of moving objects: a survey. Under preparation.
- H. Georgiou et al. (2018) Semantic-aware aircraft trajectory prediction using flight plans. Submitted.
- G.A. Vouros et al. (2018) Big data analytics for time critical mobility forecasting: recent progress and research challenges. Proceedings of EDBT.
- N. Pelekis et al. (2017) On temporal-constrained sub-trajectory cluster analysis. Data Mining and Knowledge Discovery, 31(5).
- P. Nikitopoulos et al. (2016) BigCAB: Distributed hot spot analysis over big spatio-temporal data using Apache Spark (GIS Cup). Proceedings of ACM SIGSPATIAL - GIS.
- N. Pelekis et al. (2016) Simulating our LifeSteps by example. ACM Transactions on Spatial Algorithms and Systems, 2(3).

Data Science Lab @ Univ. of Piraeus

URL: http://datastories.org
Facebook: @DataStories
Twitter: @UnipiDataSciLab

"…Our goal is to address the challenging problems related to the wealth data, by advancing research and producing solutions to real world problems related to efficient and scalable management of Big Data, including gathering and cleansing data, storing and indexing data, analyzing and mining data."

“Assurance”: R&D in fMRI (post-doc, ΕΚΠΑ, 2013-2015)

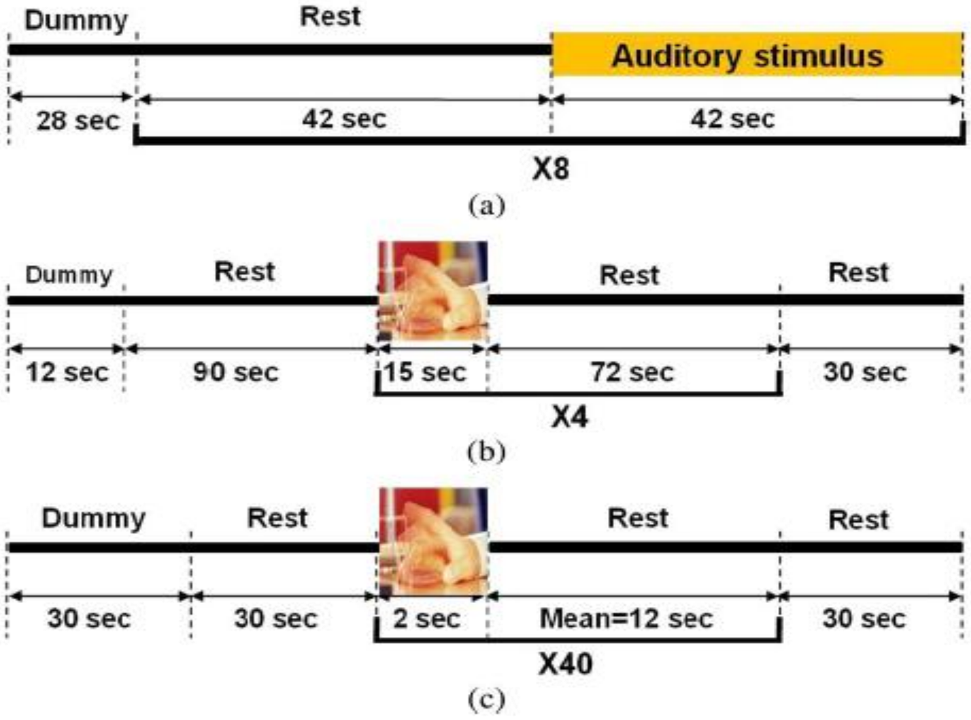
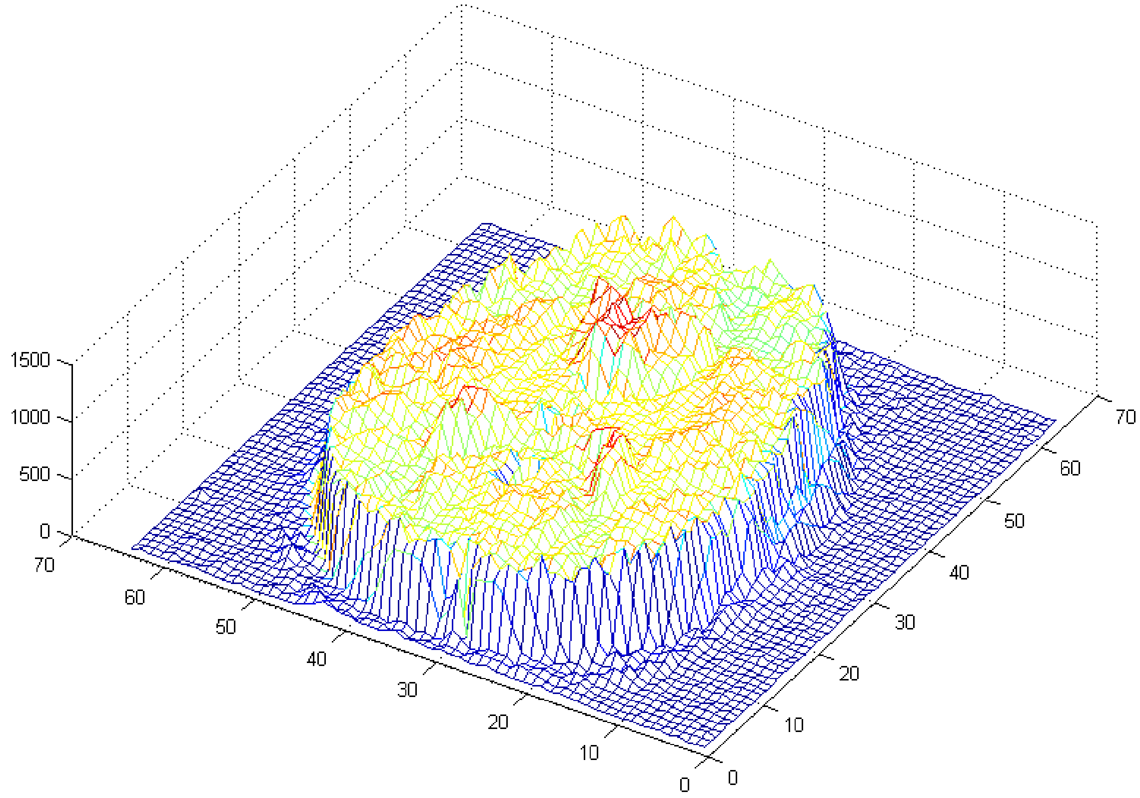
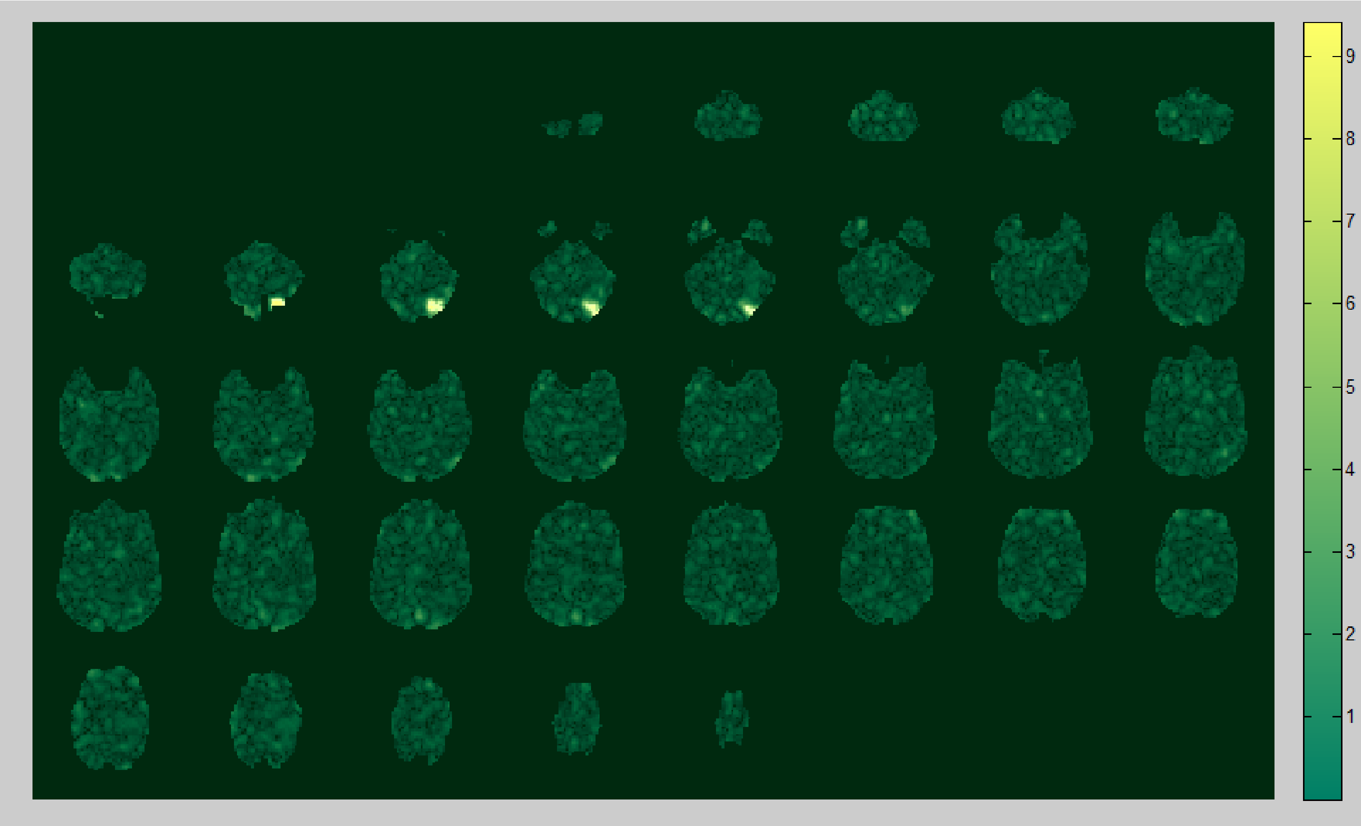


Fig. 3. Experimental paradigms for (a) auditory stimulus tasks, (b) block paradigm right finger tapping tasks, and (c) event-related right finger tapping tasks.

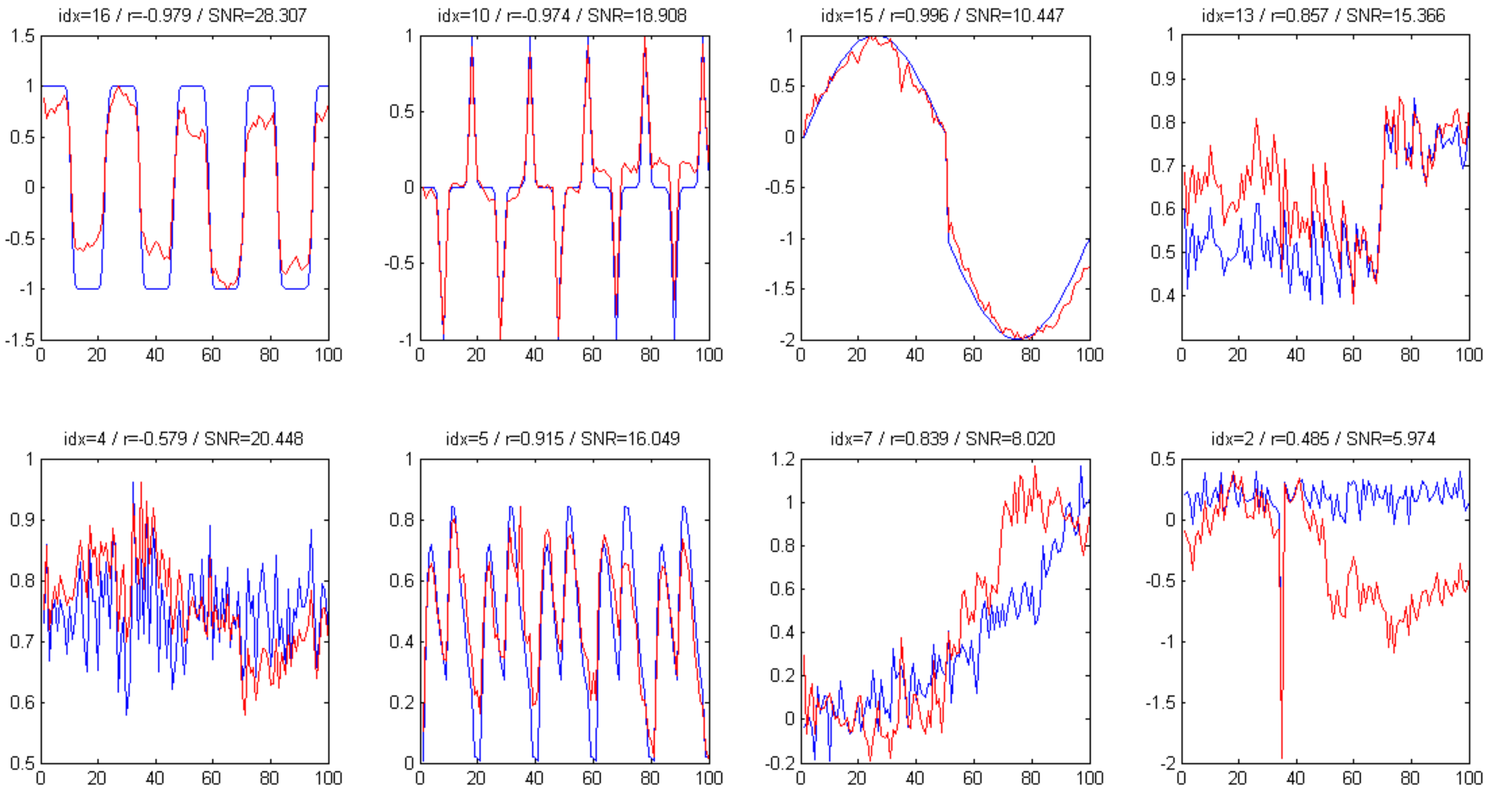
“Assurance”: R&D in fMRI (post-doc, ΕΚΠΑ, 2013-2015)



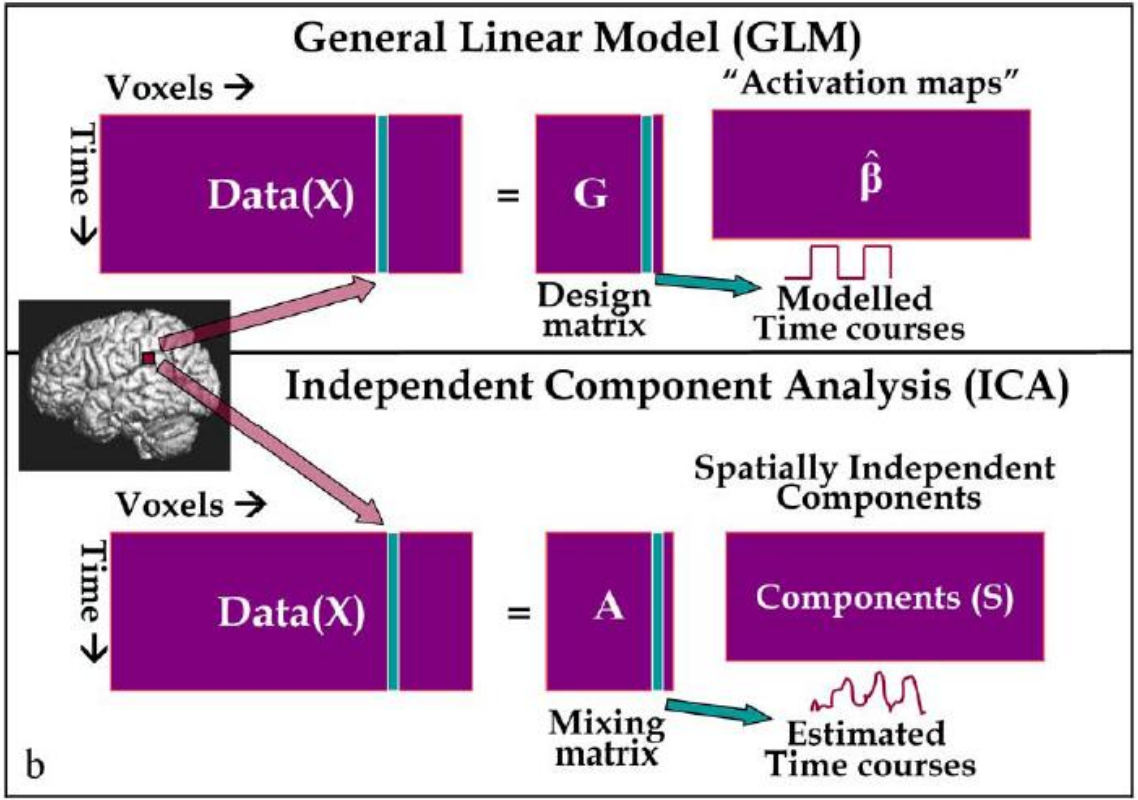
“Assurance”: R&D in fMRI (post-doc, ΕΚΠΑ, 2013-2015)



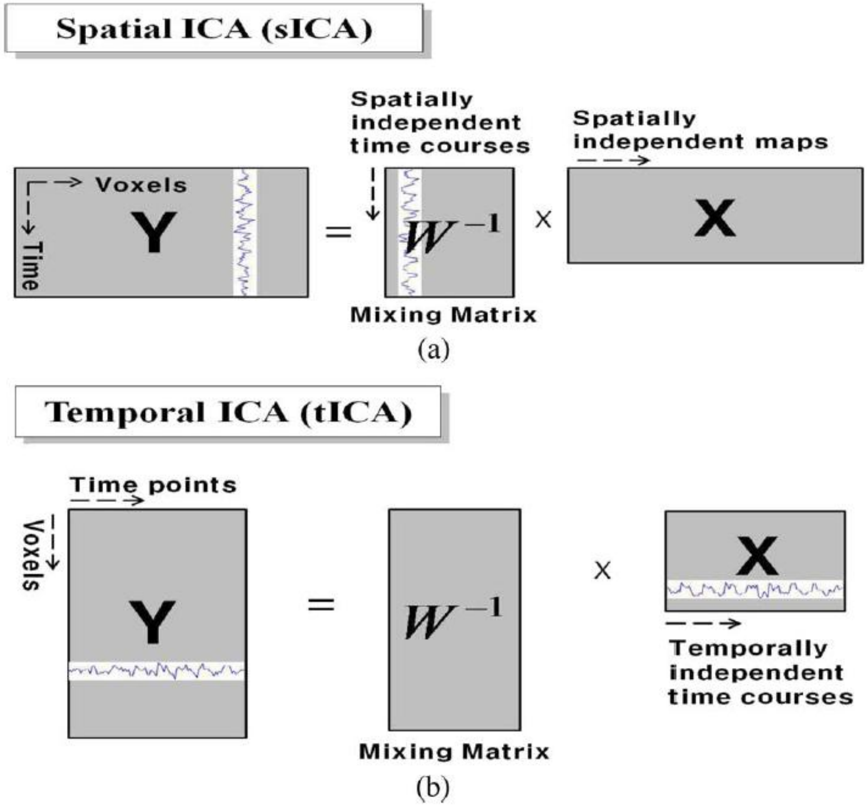
“Assurance”: R&D in fMRI (post-doc, ΕΚΠΑ, 2013-2015)



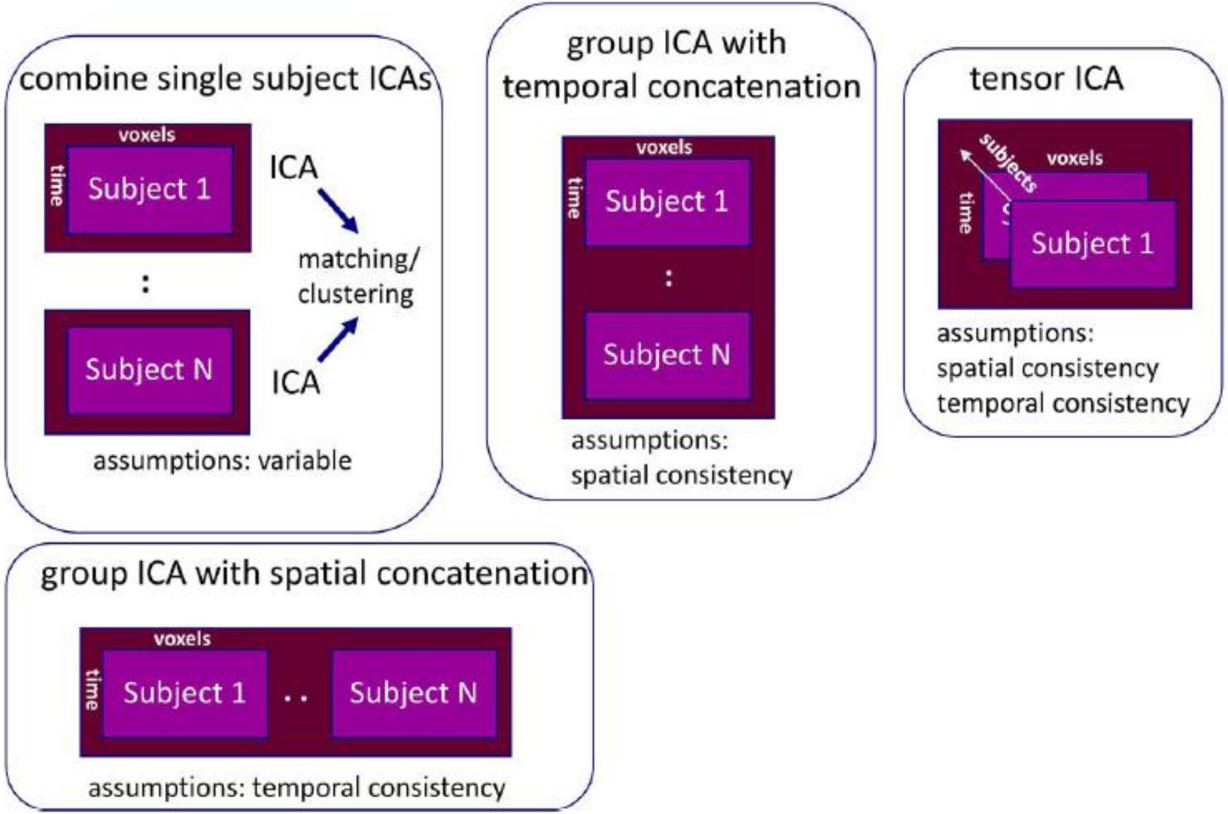
“Assurance”: R&D in fMRI (post-doc, ΕΚΠΑ, 2013-2015)



“Assurance”: R&D in fMRI (post-doc, ΕΚΠΑ, 2013-2015)



"Assurance": R&D in fMRI (post-doc, ΕΚΠΑ, 2013-2015)



“Assurance”: R&D in fMRI (post-doc, ΕΚΠΑ, 2013-2015)

Dsize=12	Ca	Cm	Cam
fastICA	0.776 (0.917)	0.849 (0.971)	0.812 (0.944)
FIDL	0.833 (0.961)	0.801 (0.921)	0.817 (0.941)
sparse GLM	0.702 (0.764)	0.714 (0.800)	0.708 (0.782)
Proposed	0.846 (0.950)	0.853 (0.950)	0.850 (0.950)

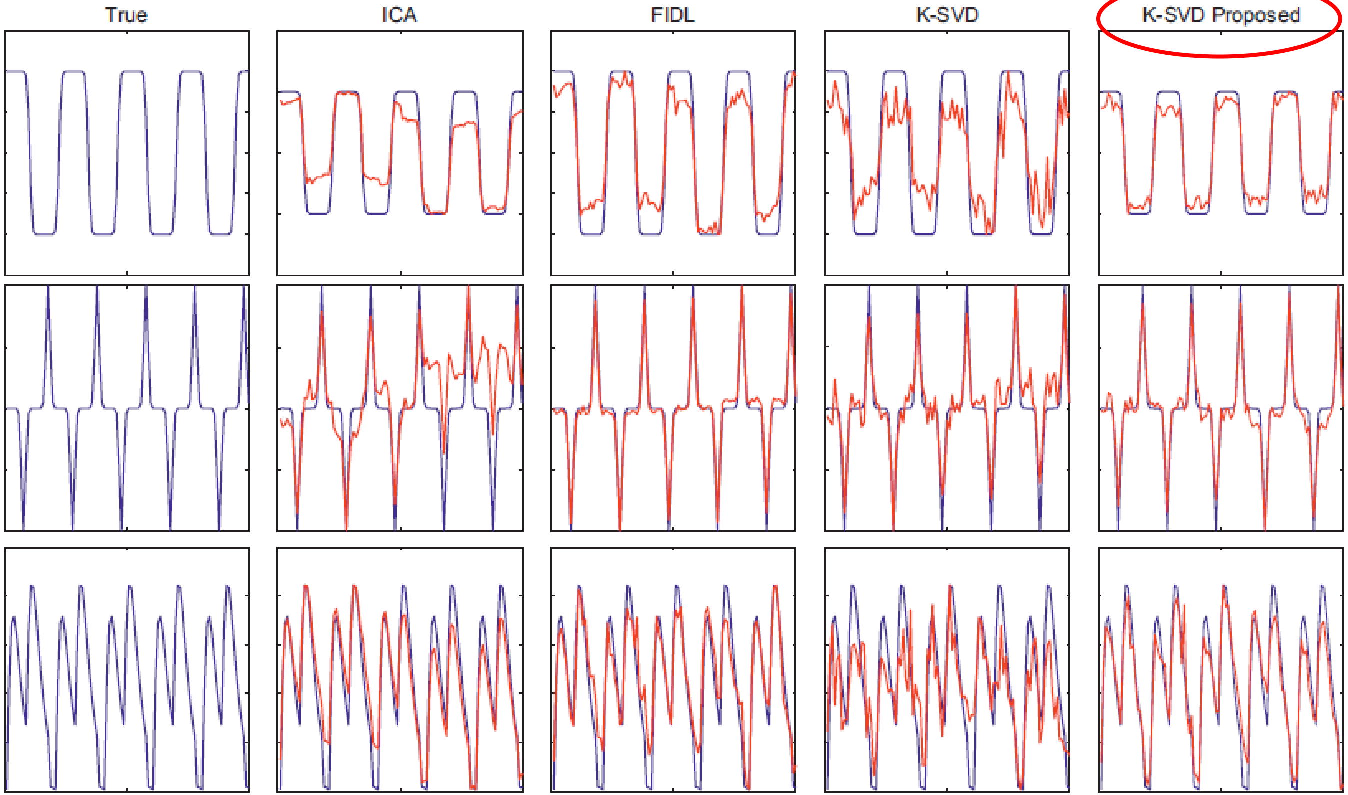
Table 1. Performance with Dictionary size 12

Dsize=20	Ca	Cm	Cam
fastICA	0.776 (0.917)	0.849 (0.971)	0.812 (0.944)
FIDL	0.836 (0.960)	0.792 (0.895)	0.814 (0.927)
sparse GLM	0.798 (0.914)	0.697 (0.793)	0.747 (0.853)
Proposed	0.862 (0.972)	0.861 (0.952)	0.861 (0.962)

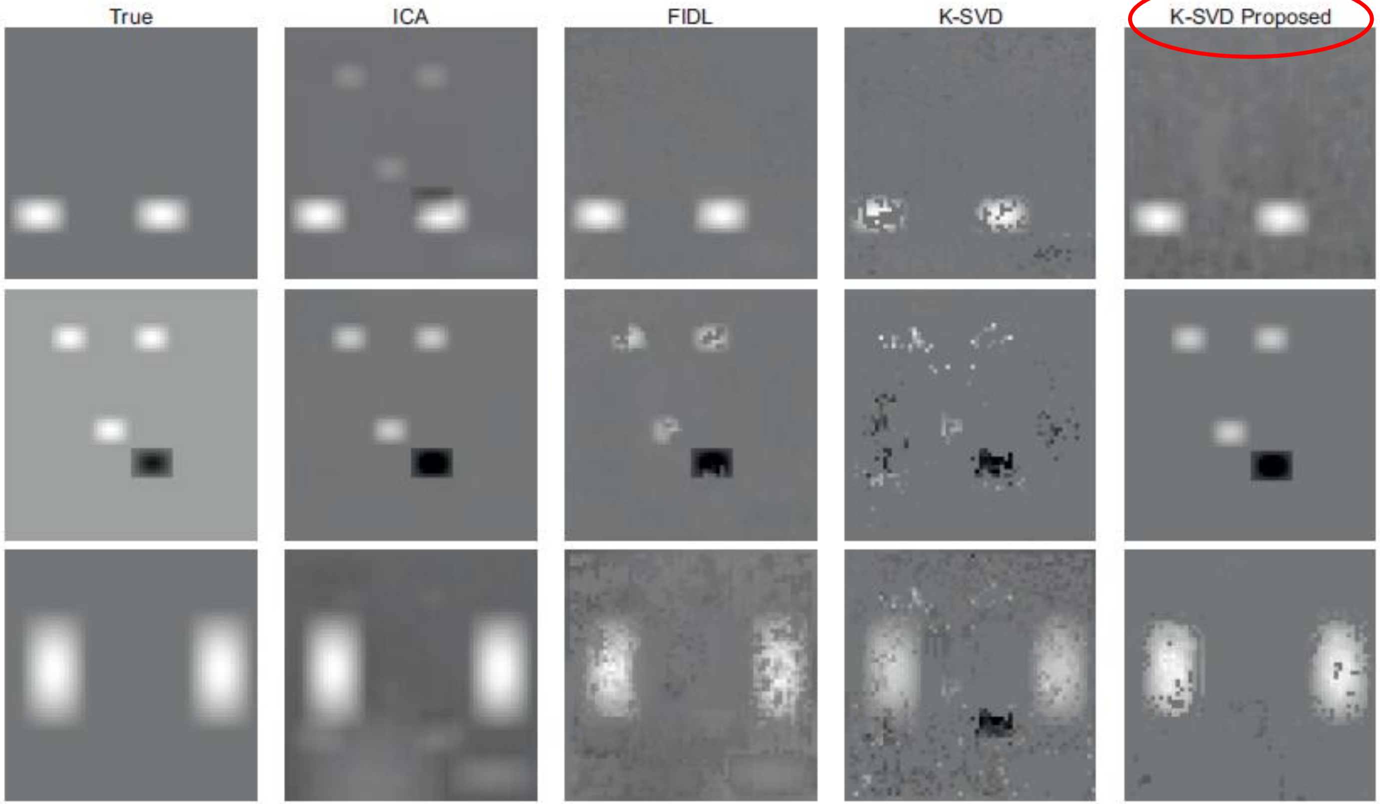
Table 2. Performance with Dictionary size 20

** See: Y. Kopsinis, H. Georgiou, S. Theodoridis (2014). “fMRI Unmixing Via Properly Adjusted Dictionary Learning”, 22nd European Signal Processing Conference (EUSIPCO 2014), 1-5 Sept 2014 @ Lisbon, Portugal.*

“Assurance”: R&D in fMRI (post-doc, ΕΚΠΑ, 2013-2015)

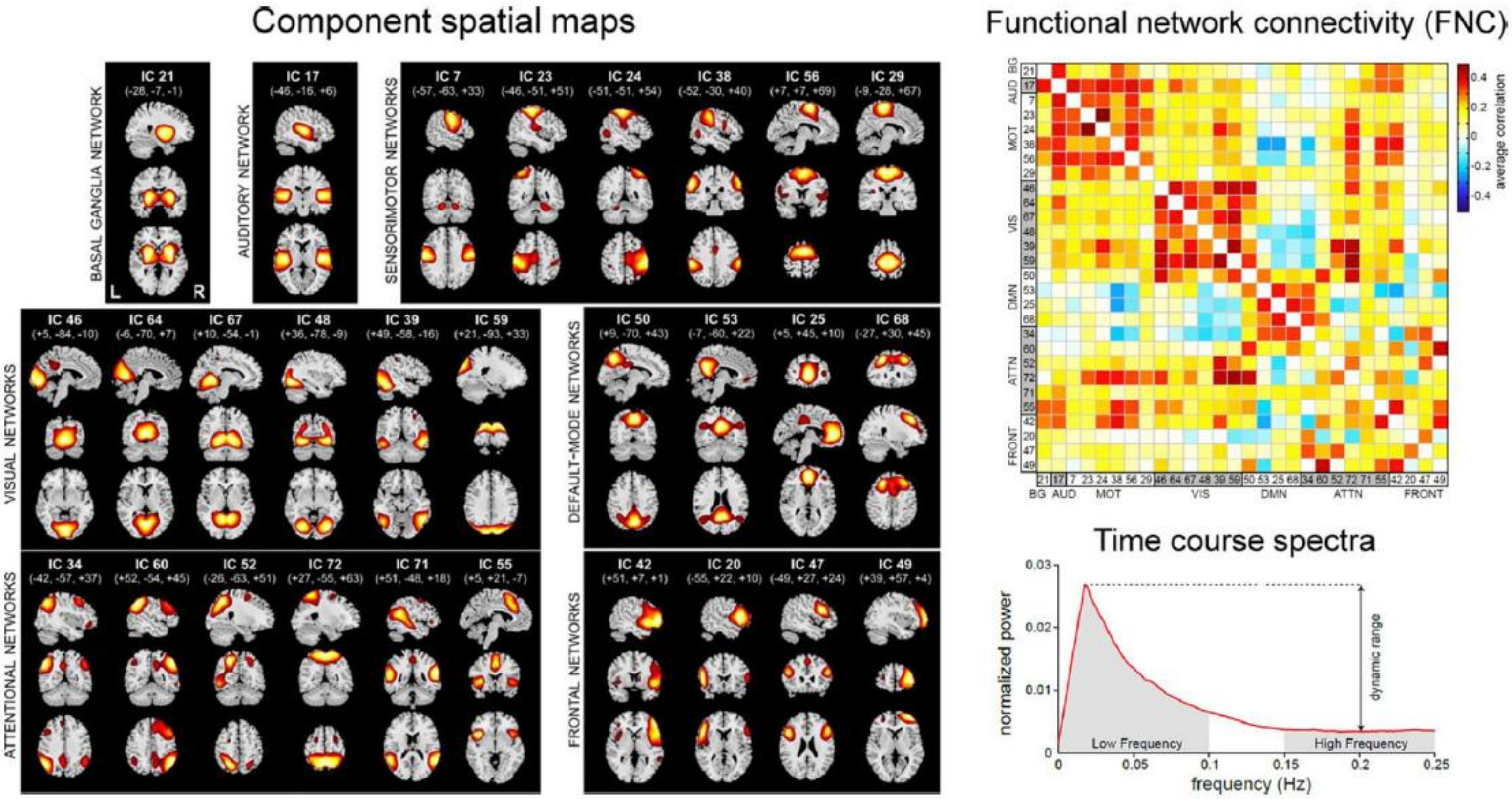


“Assurance”: R&D in fMRI (post-doc, ΕΚΠΑ, 2013-2015)

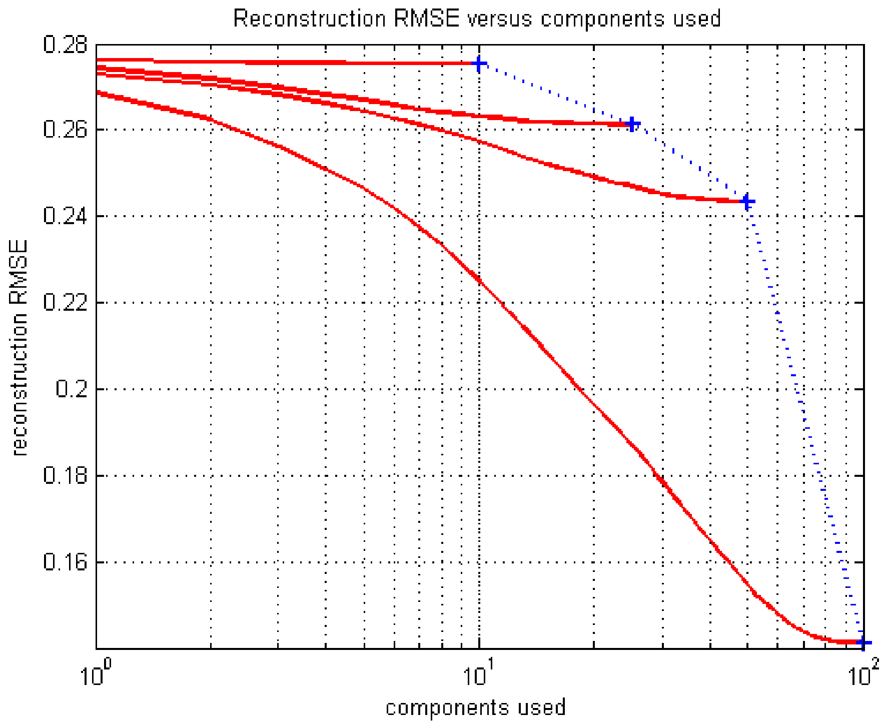


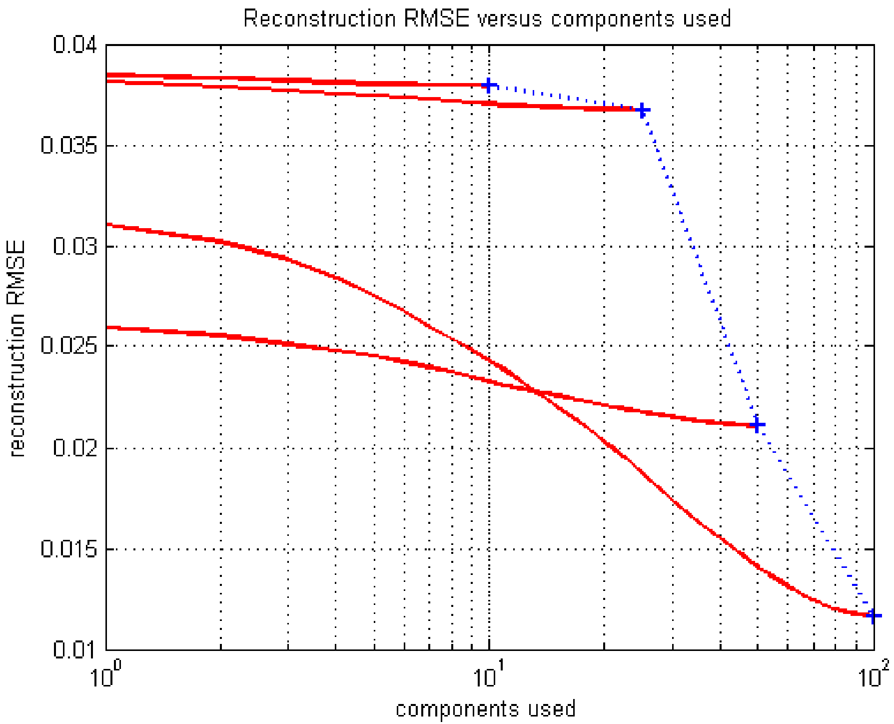
* See: Y. Kopsinis, H. Georgiou, S. Theodoridis (2014). “fMRI Unmixing Via Properly Adjusted Dictionary Learning”, 22nd European Signal Processing Conference (EUSIPCO 2014), 1-5 Sept 2014 @ Lisbon, Portugal.

“Assurance”: R&D in fMRI (post-doc, ΕΚΠΑ, 2013-2015)



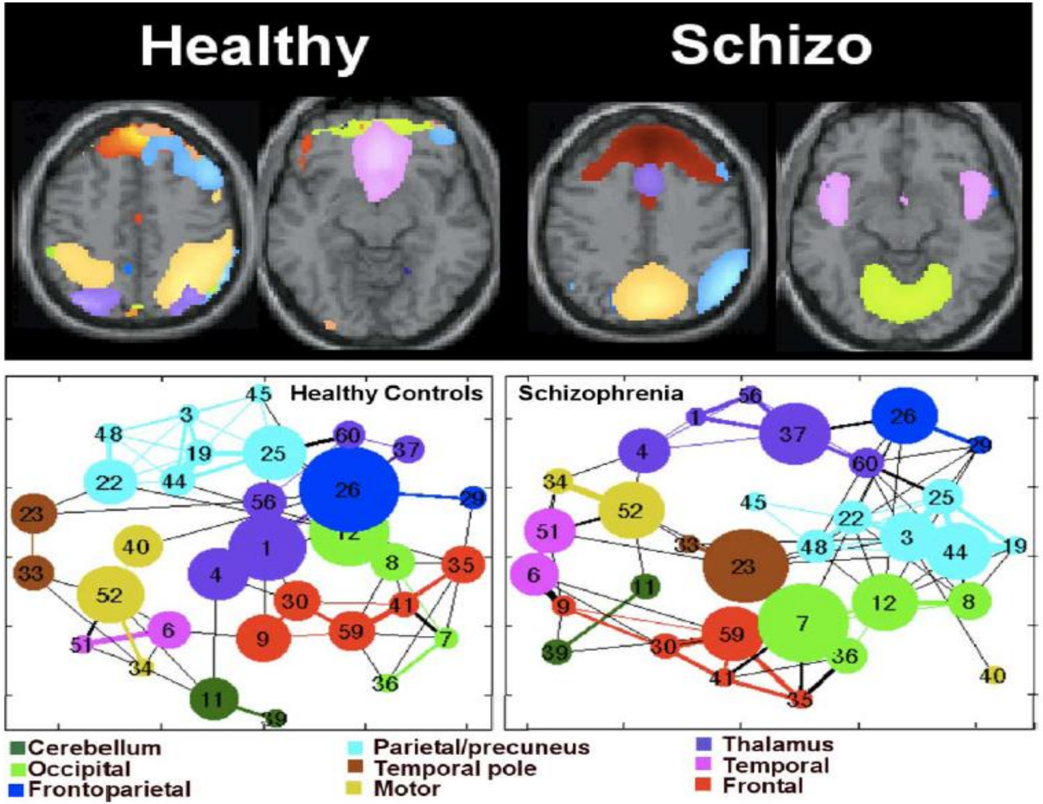
“Assurance”: R&D in fMRI (post-doc, ΕΚΠΑ, 2013-2015)





** See: H. Georgiou (2017). “Intrinsic dimension estimation of the fMRI space via sparsity-promoting matrix factorization”, 21st Panhelenic Conference in Informatics (PCI 2017), 28-30 Sept 2017 @ Larisa, Greece (ACM).*

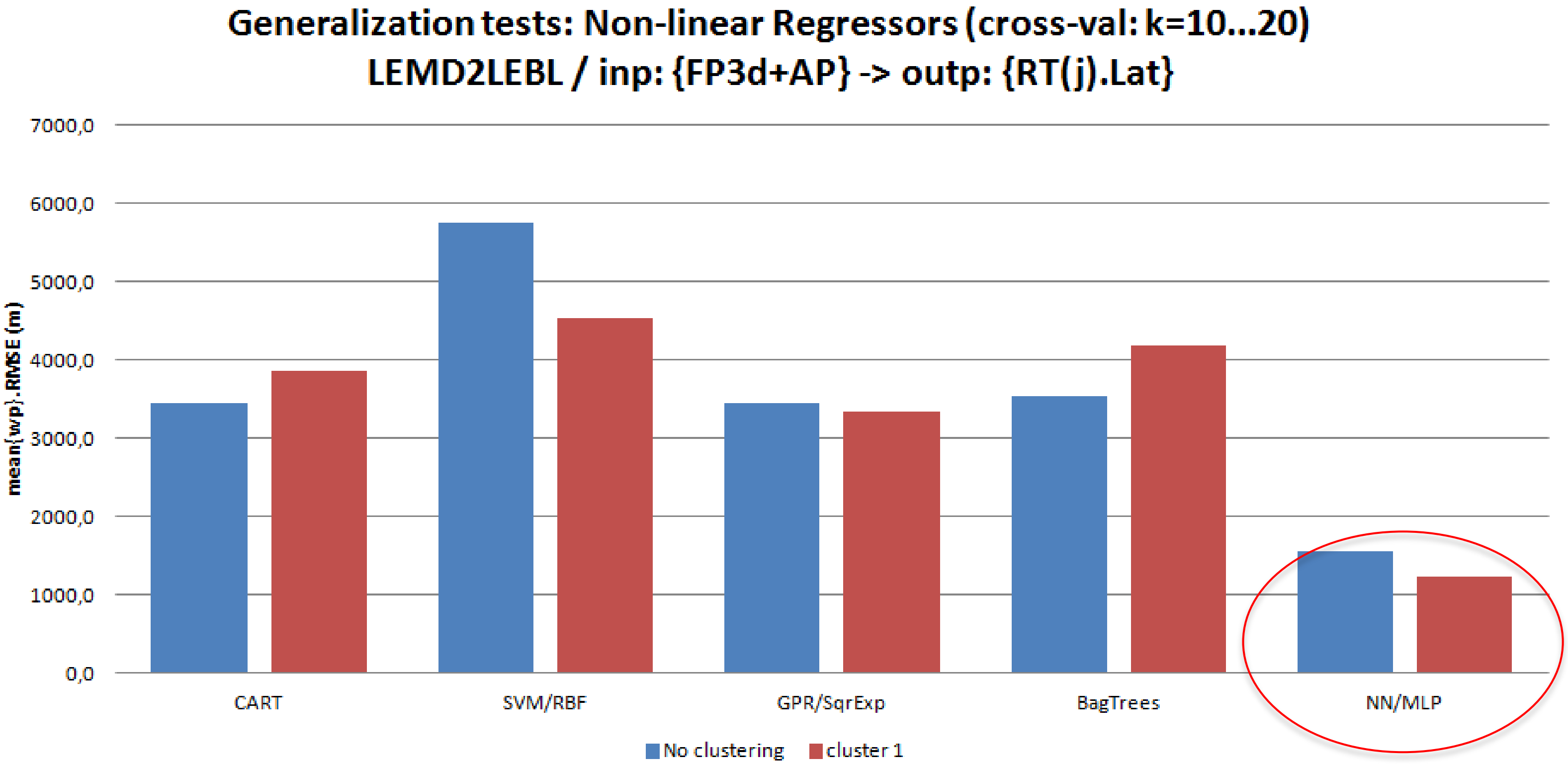
“Assurance”: R&D in fMRI (post-doc, ΕΚΠΑ, 2013-2015)



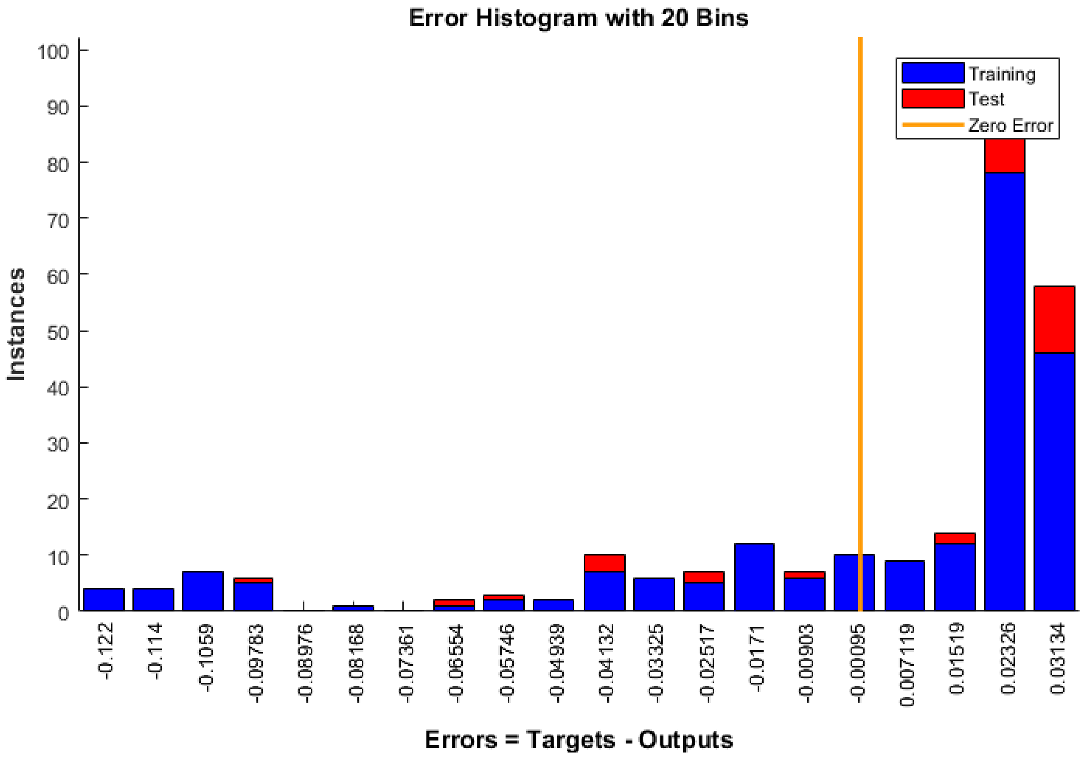
Questions

Χάρης Γεωργίου (MSc,PhD)
Data Science Lab @ Πανεπ. Πειραιά
Email: harris@xgeorgio.info

Non-linear Regressors (experiments)

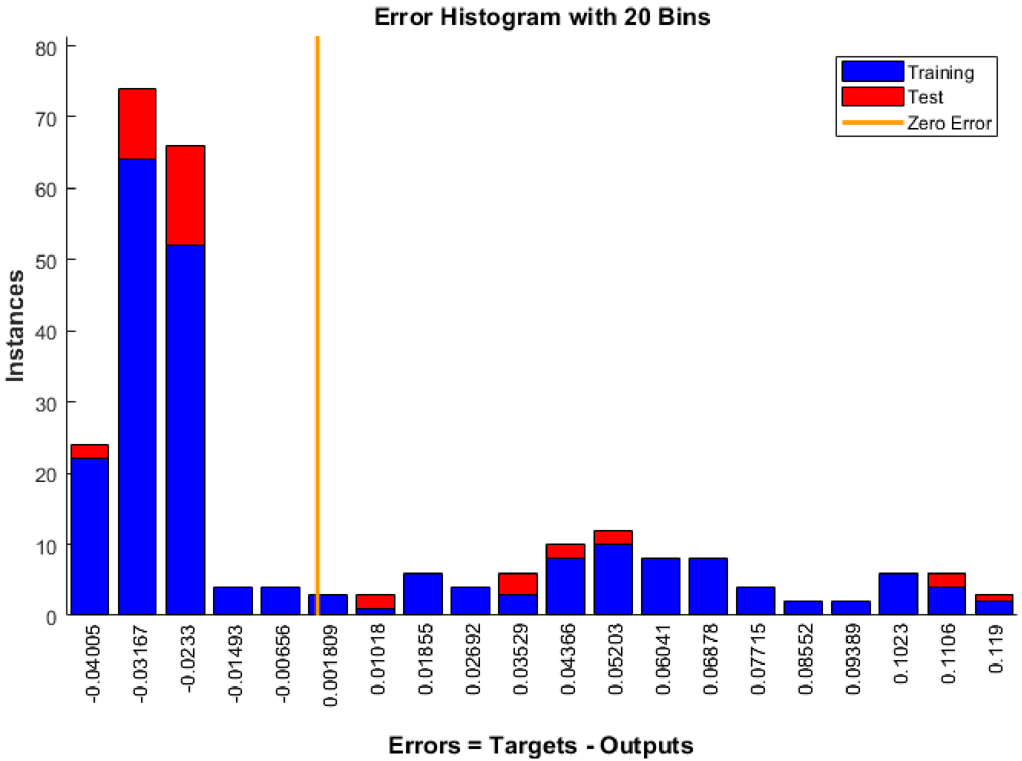


Non-linear Regressors (NN)



LEMD2LEBL / cluster 1 / wp8 (tst=15%): input=FP(3D)+AP → output=Lat (NN)

Non-linear Regressors (NN)



LEMD2LEBL / cluster 1 / wp10 (tst=15%): input=FP(3D)+AP → output=Lat (NN)

Hybrid Clustering-Pred.Model (HMM, LR, DT, NN)

DART

SESAR JOINT UNDERTAKING

- Stage-2: HMMs vs. Linear Regressors:
  - ➢ LR improves accuracy (9-34%) vs. HMM across all configurations
  - ➢ clustering/LR better (up to 16,5%) than clustering/HMM

- Input dimension: FP(1D) vs. full FP(1D) vs. full FP(3D) vs. full FP(3D)+AP:
  - ➢ LR/FP(3D) provides x3-x4 improvement of accuracy than LR/FP(1D) over HMM
  - ➢ LR/FP(4D), i.e. LR/FP(3D) + AP, improves accuracy only marginally (1-3%)

- Non-clustered vs. Clustered dataset (HMM, LR, CART):
  - ➢ Clustering improves accuracy across all models & dimensionality (14-26,5%)
  - ➢ Misaligned samples in clustering produces instability in stage-2 pred. models

** See: H. Georgiou et al. (2018), Semantic-aware aircraft trajectory prediction using flight plans. (submitted)*

Hybrid Clustering-Pred.Model (HMM, LR, DT, NN)

DART

SESAR JOINT UNDERTAKING

- LR (primary choice) vs. CART (as alternative):
  - LR: (+) more resilient to "noise" in training, (-) less robust in generalization
  - CART: (-) less resilient to "noise" in training, (+) more robust in generalization
  - Complementary behavior, hints for combining them per-dimension

- Per-dimension comments, "active window" (LR), dataset properties:
  - Lat. is inherently much more difficult to predict, possibly due to E/W orientation
  - Some LR use only "local neighbor" of ref. points (small coefficients elsewhere)

** See: H. Georgiou et al. (2018), Semantic-aware aircraft trajectory prediction using flight plans. (submitted)*

Hybrid Clustering-Pred.Model (HMM, LR, DT, NN)

DART

SESAR JOINT UNDERTAKING

- Advanced non-linear regressors, generalization (vs. LR, CART):
  - ➢ More resilient to “noise” in training, better generalization & stability
  - ➢ Clustering (stage-1) becomes less important with more robust regressors
  - ➢ NN: seems the best tradeoff between complexity vs. performance

- NN/MLP full experimental assessment:
  - ➢ Lat. is still much more difficult to predict, possibly due to E/W orientation
  - ➢ Multi-linear regression (3D output) seems much more difficult to achieve

** See: H. Georgiou et al. (2018), Semantic-aware aircraft trajectory prediction using flight plans. (submitted)*

WP2-Single Trajectory Prediction

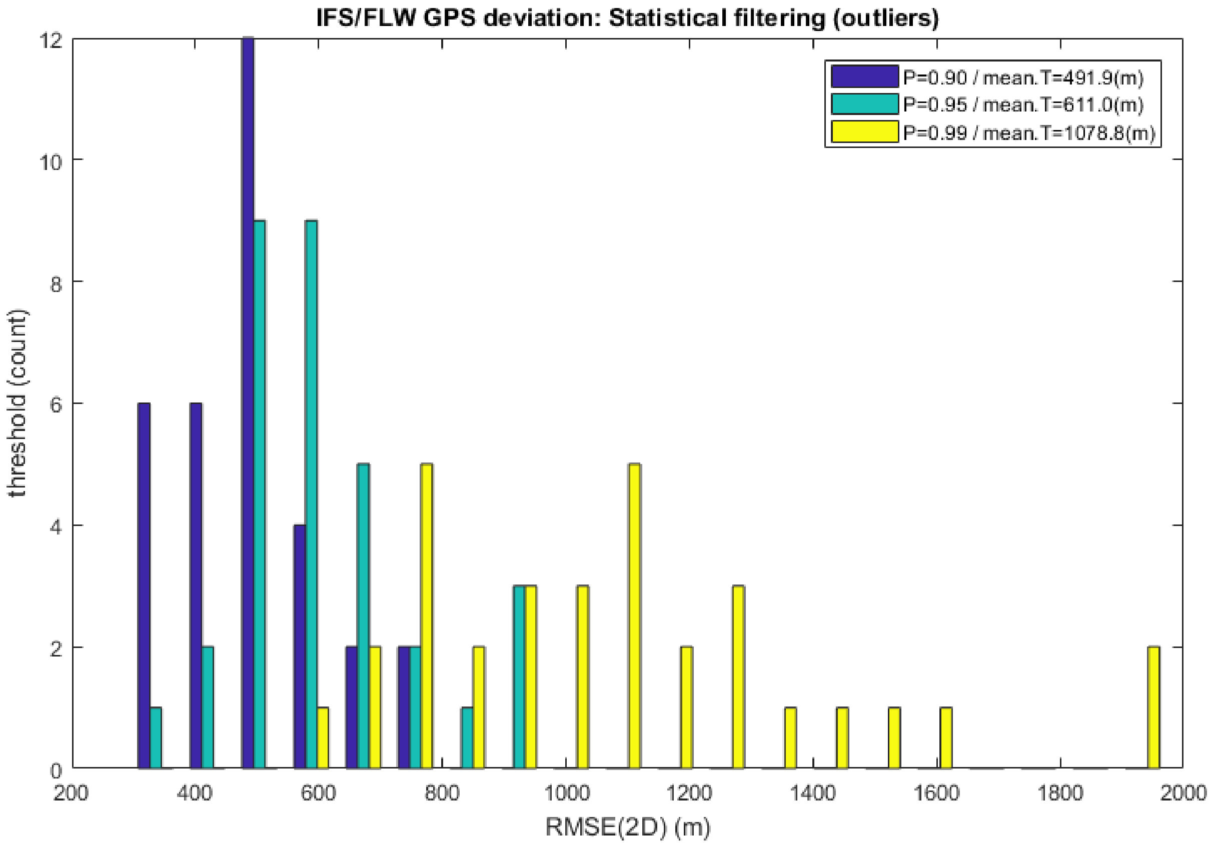
Issues, risks & contingencies:

- FP/RP "enriched" ref. points provide only a coarse trajectory as time series
- Many large deviations between "intended" (FP) and actual route (RT)
- Clustering is always beneficial in linear (HMM, LR) and multi-linear (CART)
- ...but becomes less important when robust non-linear regressors are used
- Error analysis (pred) shows few large peaks ("heavy" right tail in pdf)
- Dimensionality analysis: full 3-D FP input can be reduced (d=5...13 << 48)

✓ *Full-resolution (raw) IFS flight route may be combined with FP "constraints"*

✓ *Complexity vs. Performance tradeoff in regressors is a design-time decision*

** See: H. Georgiou et al. (2018), Semantic-aware aircraft trajectory prediction using flight plans. (submitted)*



Cross-streaming: IFS/FLW outliers distribution vs. *p*-thresholds

datAcron

European Commission

"Assurance": R&D in fMRI (post-doc, ΕΚΠΑ, 2013-2015)

General Task

- Factor the data matrix as $Y \approx AB$ aiming at getting $A \approx T$ and $B \approx S$
- There is an infinite number of such factorizations
- A priori information information about the characteristics of B and/or A need to get imposed

fMRI unmixing approaches

- **fMRI unmixing via Dictionary Learning:** Factor $Y \approx AB$ under the constraint that the columns of B are **sparse**.

$$\min_{A,B} \|Y - AB\|_F^2, \text{ s.t. } \|B_{\cdot,j}\|_0 \leq K, \; j = 1 \cdots n, \quad (1)$$

- **Fast and Incoherent Dictionary Learning** (FIDL), [Abolghasemi 2013]:
  1. Incoherence is imposed in A,
  2. relatively low complexity per iteration,
  3. improved performance to both synthetic and real fMRI data.

** See: Y. Kopsinis, H. Georgiou, S. Theodoridis (2014). "fMRI Unmixing Via Properly Adjusted Dictionary Learning", 22nd European Signal Processing Conference (EUSIPCO 2014), 1-5 Sept 2014 @ Lisbon, Portugal.*

"Assurance": R&D in fMRI (post-doc, ΕΚΠΑ, 2013-2015)

Detect split atoms and merge them back

- Split maps corresponding to the same FBN are functionally associated, therefor should exhibit similar activation time-courses
- **Spatial Maps Merging:**
  1. Construct, $\boldsymbol{m}$, the vector with indices of a set of highly correlated time courses.
  2. Set, $\boldsymbol{\beta}$ the union of the supports of the spatial maps indexed in $\boldsymbol{m}$.
  3. Solve for $\boldsymbol{a}$ and $\boldsymbol{b}$ the optimization problem

$$\min_{\boldsymbol{a},\boldsymbol{b}} \|\boldsymbol{A}_{\cdot,\boldsymbol{m}}\boldsymbol{B}_{\boldsymbol{m},\boldsymbol{\beta}} - \boldsymbol{a}\boldsymbol{b}^T\|_{\mathsf{F}}^2.$$

  Rank-1 best approximation (via truncated SVD).

  4. Re-initialize the rest of the atoms indexed in $\boldsymbol{m}$.
- **Time Courses Merging:** It can be done in exactly the same way.
- Both merging tasks need not to be performed in each K-SVD iteration.

* See: Y. Kopsinis, H. Georgiou, S. Theodoridis (2014). "fMRI Unmixing Via Properly Adjusted Dictionary Learning", 22nd European Signal Processing Conference (EUSIPCO 2014), 1-5 Sept 2014 @ Lisbon, Portugal.