# Combining Visual Analytics and Machine Learning for Route Choice Prediction
## Application to Pre-Tactical Traffic Forecast

Rodrigo Marcos
Data Scientist, Nommon Solutions and Technologies

SESAR Innovation Days
Beograd, 29th November 2017

# Scope and Objectives

## Problem:

- ATFCM in the pre-tactical phase

## Current approach:

- Based on similarity
  http://www.eurocontrol.int/articles/ddr-pre-tactical-traffic-forecast

## Objectives:

- Use visual analytics to extract route choice determinants
- Model behaviour of airlines regarding route choice between airport pairs using machine learning techniques
- Evaluate pre-tactical prediction power

# State of the Art
# Airline Route Choice Behaviour

Abundant research on tactical trajectory prediction:

- Prediction of arrival time

- Conflict detection

- …

Limited research on airline route choice prediction before the availability of flight plans (pre-tactical forecast):

- Luis Delgado (2015) "European route choice determinants"

# Approach

- Data: actual trajectories (M3) from DDR2

- Route clustering per OD

- Visual exploration of route choice determinants

- Train a machine learning model

- Evaluate quality of predictions vs null model

# Case Studies

- ODs:
  - Istanbul to Paris
  - Canary Islands to London

- Multinomial regression

- Candidate variables
  - Route length
  - Charges
  - Time
  - Schedule
  - Congestion

- Temporal scope:
  - Training/exploration: AIRACs 1601-1603
  - Testing: AIRACs 1501, 1502

# Clustering
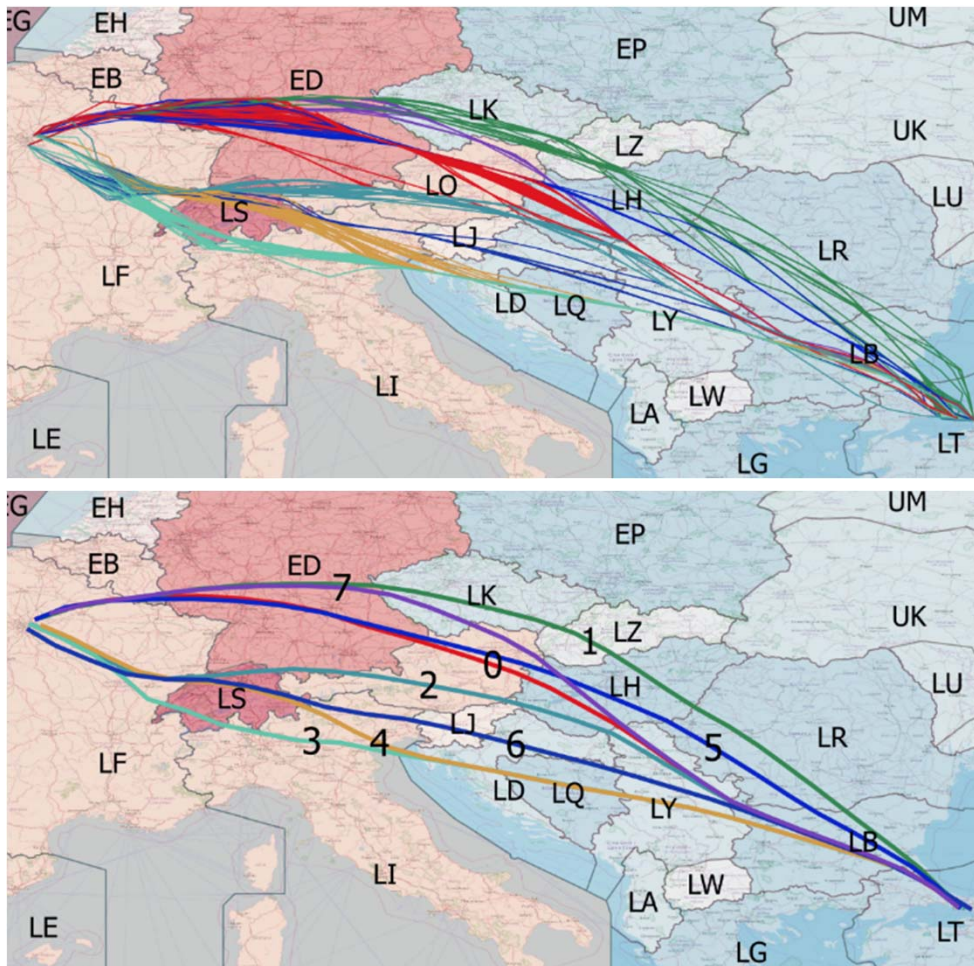


| Cluster | No of flights |
|---------|---------------|
| 0       | 139           |
| 1       | 110           |
| 2       | 190           |
| 3       | 218           |
| 4       | 117           |
| 5       | 73            |
| 6       | 29            |
| 7       | 24            |

Clustered with DBScan
Metric:
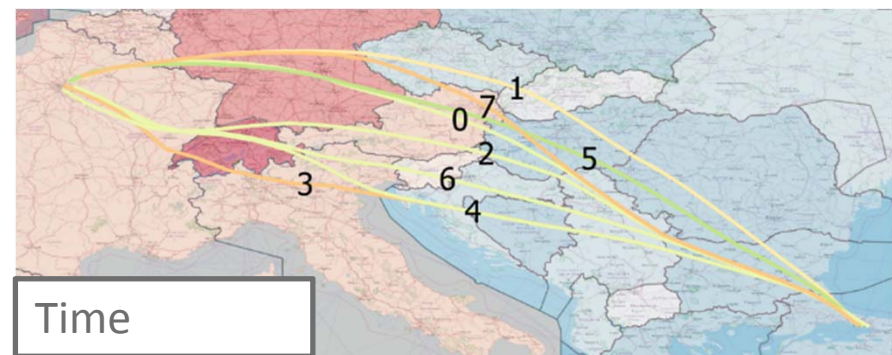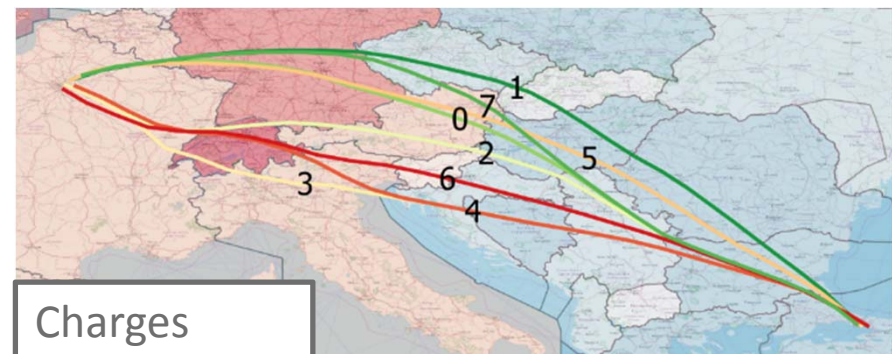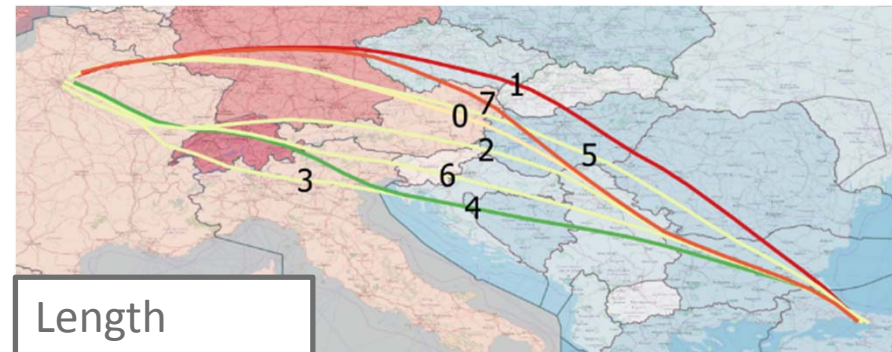Flown kilometres per ANSP

# Visual Exploration
# Cost-worthiness

2 variables considered

- Average route length
- Average route charges

1 variable discarded

- Average flight time

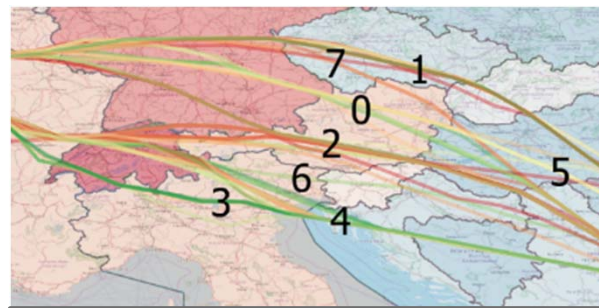| Cluster | No of flights |
|---------|---------------|
| 0 | 139 |
| 1 | 110 |
| 2 | 190 |
| 3 | 218 |
| 4 | 117 |
| 5 | 73 |
| 6 | 29 |
| 7 | 24 |



Length

Charges

Time

# Visual Exploration
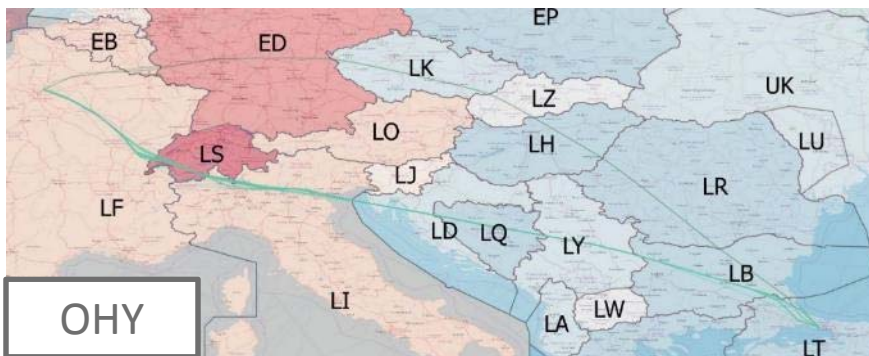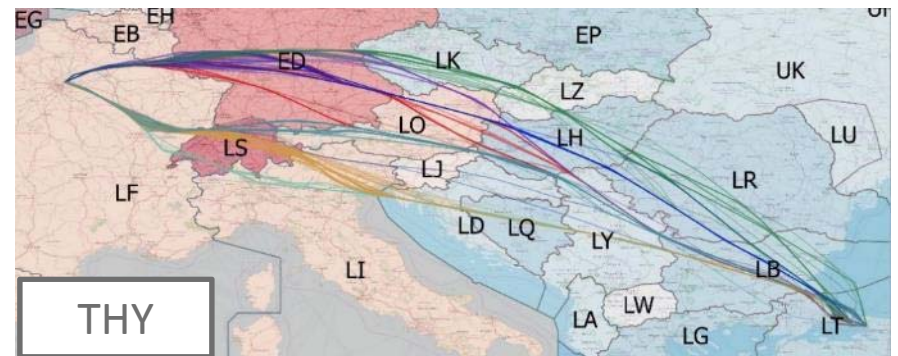# Airline Behaviour

2 variables considered

- Arrival time
- Airline



20:00-22:00 (all airlines)
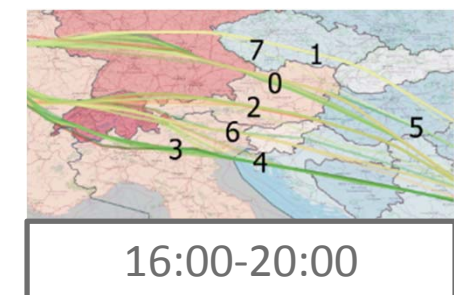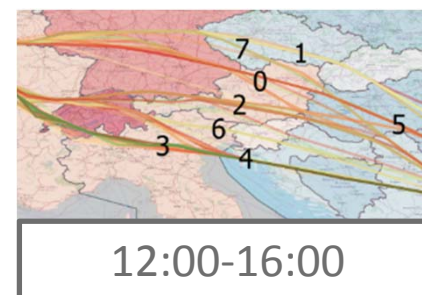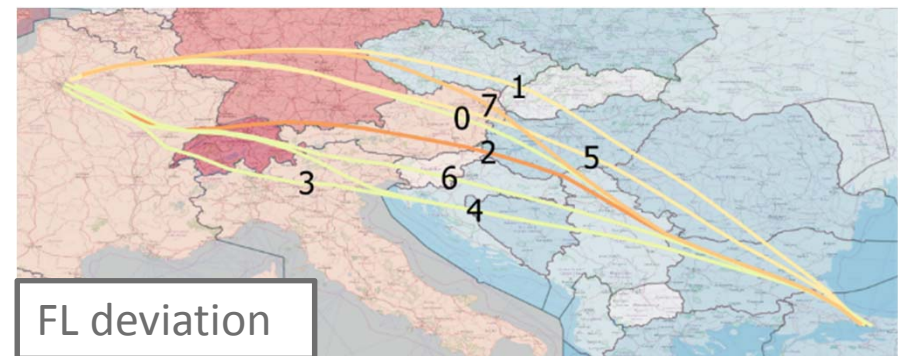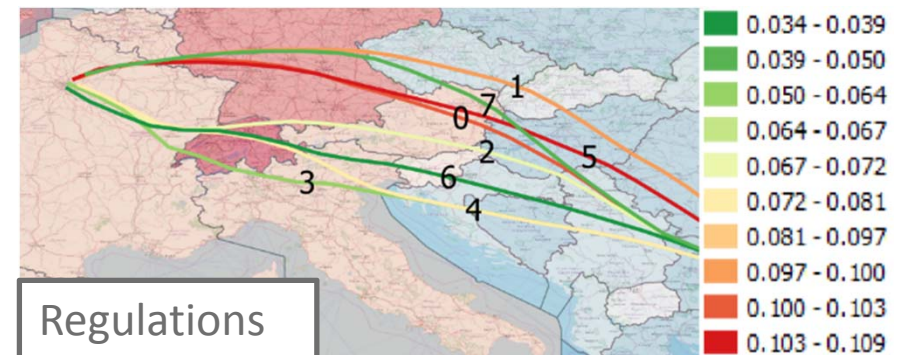
22:00-00:00 (all airlines)

OHY

THY

# Visual Exploration Congestion

1 variable considered

- Average number of regulated flights

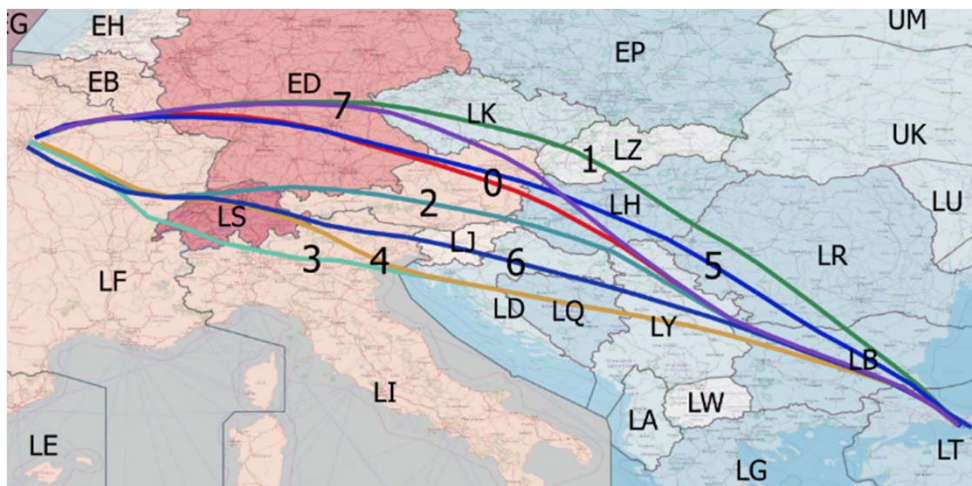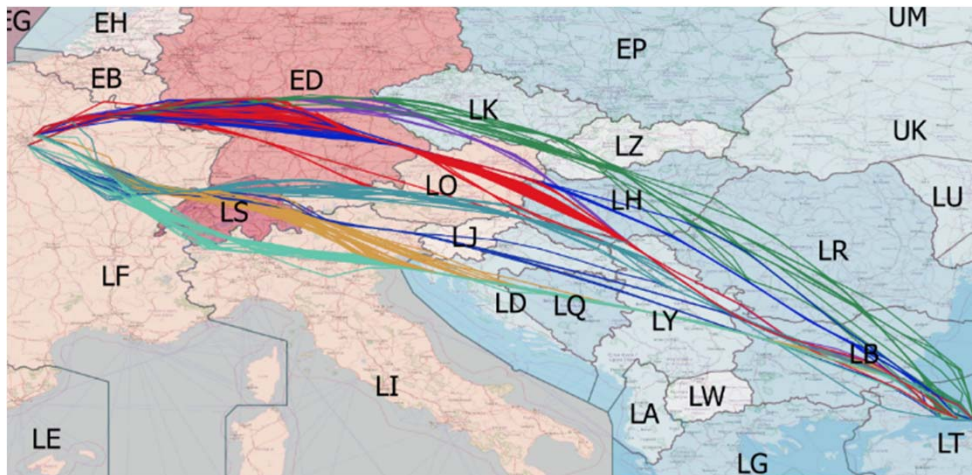1 variable discarded

- Average standard deviation of en-route FL with respect to RFL
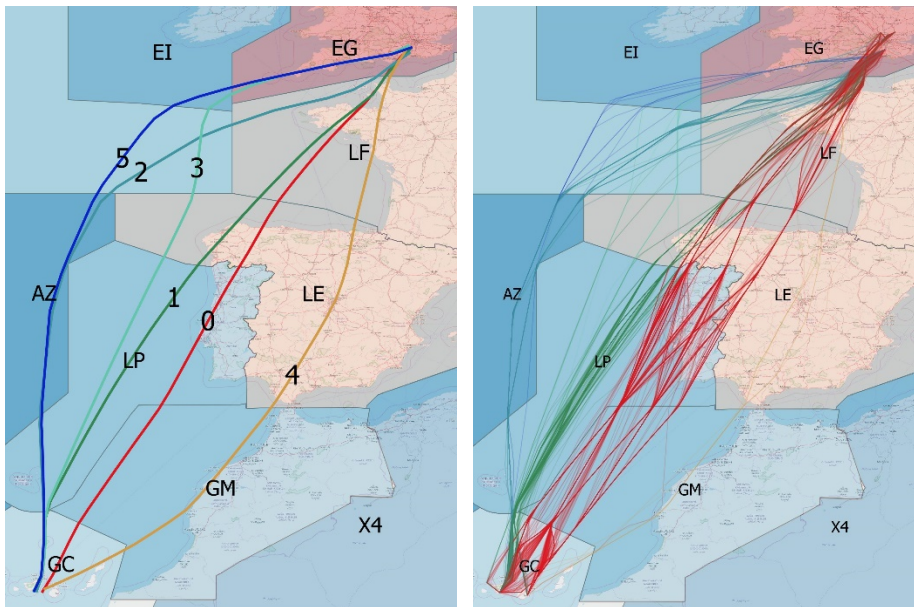


Regulations

FL deviation

12:00-16:00

16:00-20:00

Legend:
- 0.034 - 0.039
- 0.039 - 0.050
- 0.050 - 0.064
- 0.064 - 0.067
- 0.067 - 0.072
- 0.072 - 0.081
- 0.081 - 0.097
- 0.097 - 0.100
- 0.100 - 0.103
- 0.103 - 0.109

# Visual Exploration
# Cluster Properties



## Istanbul - Paris

| Cluster | No of flights | Average length (NM) | Average charges (EUR) | Regulations per flight |
|---------|---------------|---------------------|-----------------------|------------------------|
| 0 | 139 | 1277 | 1188 | 0.15 |
| 1 | 110 | 1314 | 1144 | 0.11 |
| 2 | 190 | 1273 | 1199 | 0.06 |
| 3 | 218 | 1274 | 1203 | 0.06 |
| 4 | 117 | 1256 | 1207 | 0.07 |
| 5 | 73 | 1274 | 1204 | 0.1 |
| 6 | 29 | 1271 | 1229 | 0.03 |
| 7 | 24 | 1304 | 1152 | 0.04 |

# Visual Exploration Cluster Properties



## Canary Islands - London

| Cluster | No of flights | Average length (NM) | Average charges (EUR) | Regulations per flight |
|---------|---------------|---------------------|-----------------------|------------------------|
| 0 | 659 | 1620 | 1653 | 0.18 |
| 1 | 238 | 1638 | 1676 | 0.13 |
| 2 | 68 | 1740 | 1051 | 0.13 |
| 3 | 13 | 1732 | 1582 | 0.46 |
| 4 | 7 | 1724 | 1893 | 0.42 |
| 5 | 10 | 1780 | 1165 | 0 |

# Approach Parameters

**Route parameters
(used for modelling):**

- Cost-worthiness:
    - Average route charges
    - Average route length

- Congestion:
    - Rate of regulated flights

**Flight parameters
(used for segmentation):**

- Airline (CASK)
- Arrival time

# Modelling Approach
# Multinomial Regression Model

$$P(Y_i = j) = \frac{e^{\beta_i \cdot X_j + \alpha_j}}{1 + \sum_{k=1}^{J} e^{\bar{\beta}_i \cdot \bar{X}_k + \alpha_k}}$$
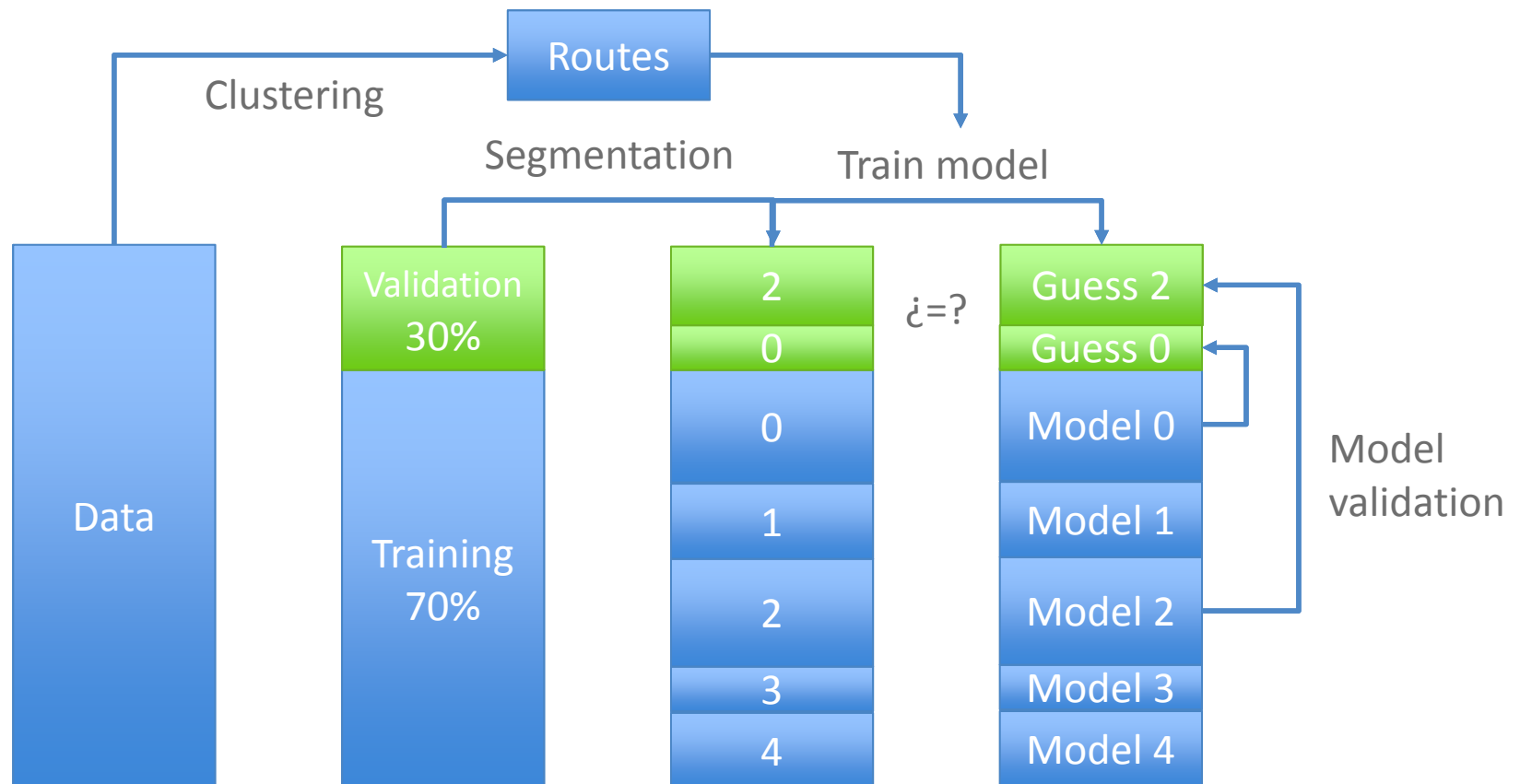
Model of class i and cluster j
- $X_j$ vector of parameters of cluster j
- $\beta_i$ vector of constants of model i
- $\alpha_j$ independent constant of cluster j

Variables:

- Cost-worthiness:
  - Average route charges
  - Average route length
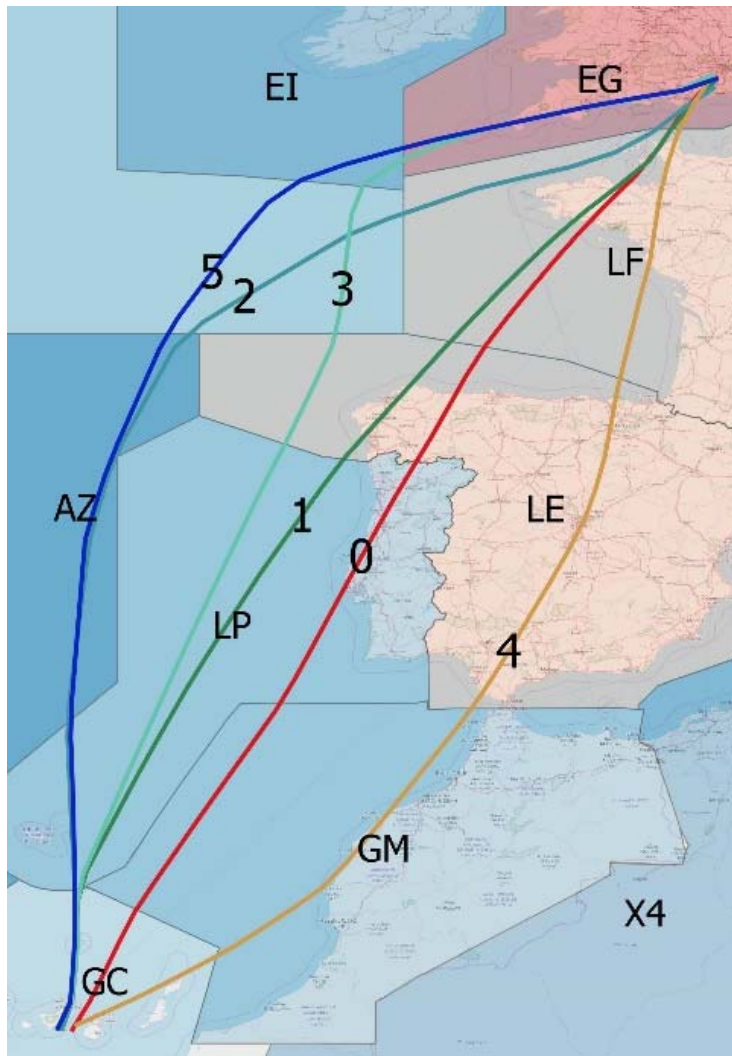- Congestion:
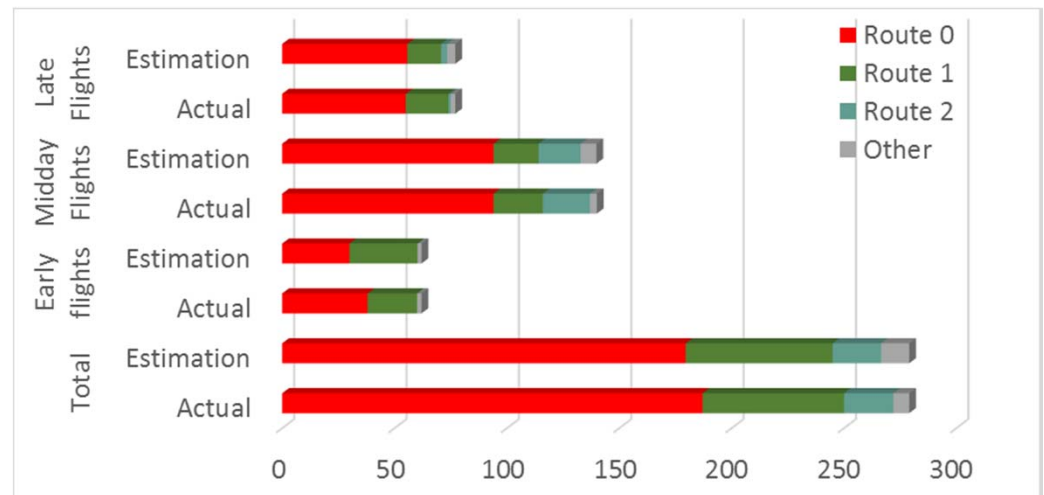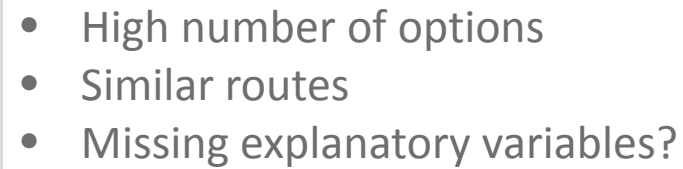  - Rate of regulated flights

# Approach
## Training and Validation

# Validation Results
# Canary Islands-London



- Low number of routes
- Very different
- Well explained

# Validation Results Istanbul-Paris



- High number of options
- Similar routes
- Missing explanatory variables?

| Cluster | No of flights | Average length (NM) | Average charges (EUR) | Regulations per flight |
|---------|---------------|---------------------|-----------------------|------------------------|
| 3 | 218 | 1274 | 1203 | 0.06 |
| 4 | 117 | 1256 | 1207 | 0.07 |

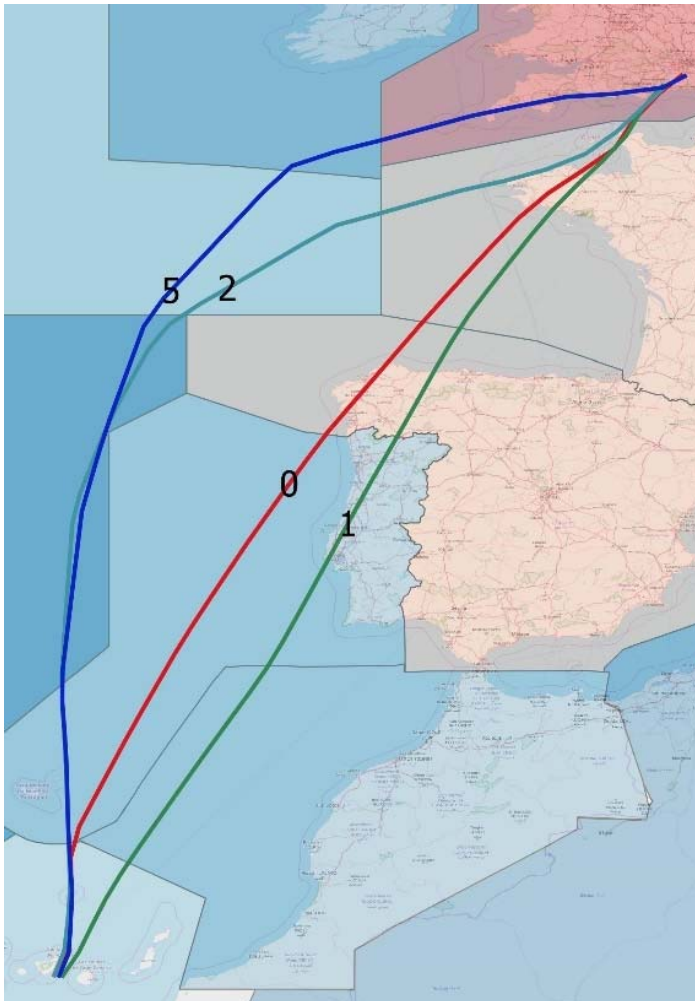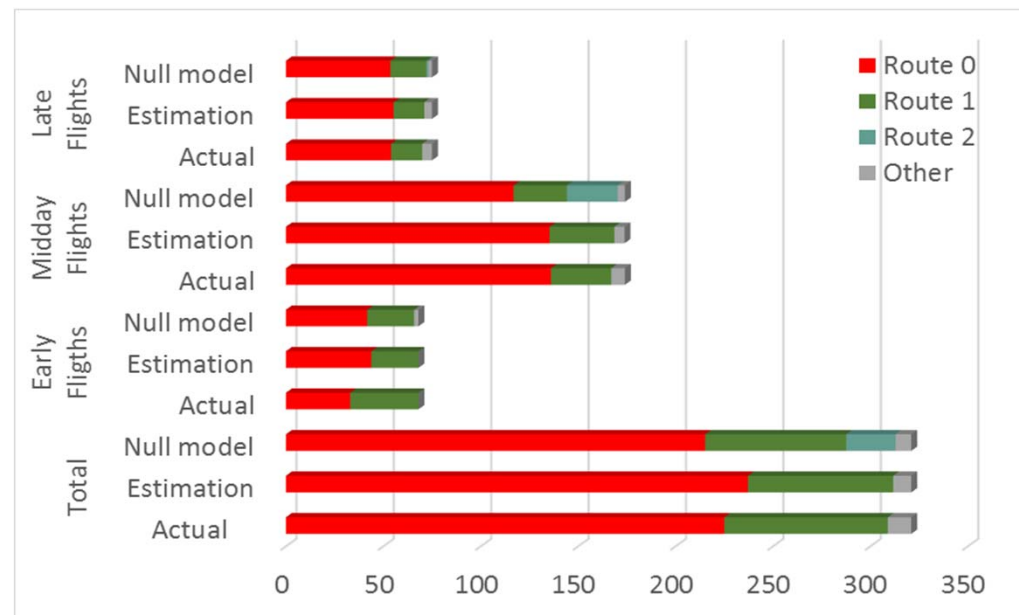| Cluster | No of flights | Average length (NM) | Average charges (EUR) | Regulations per flight |
|---------|---------------|---------------------|-----------------------|------------------------|
| 0 | 139 | 1277 | 1188 | 0.15 |
| 5 | 73 | 1274 | 1204 | 0.11 |
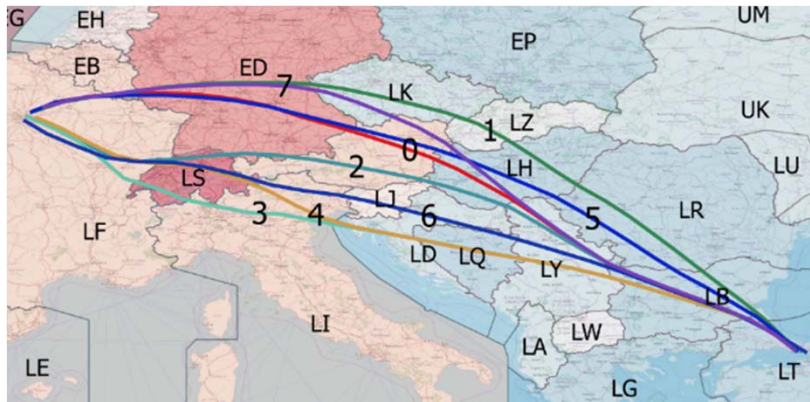
# Approach Testing
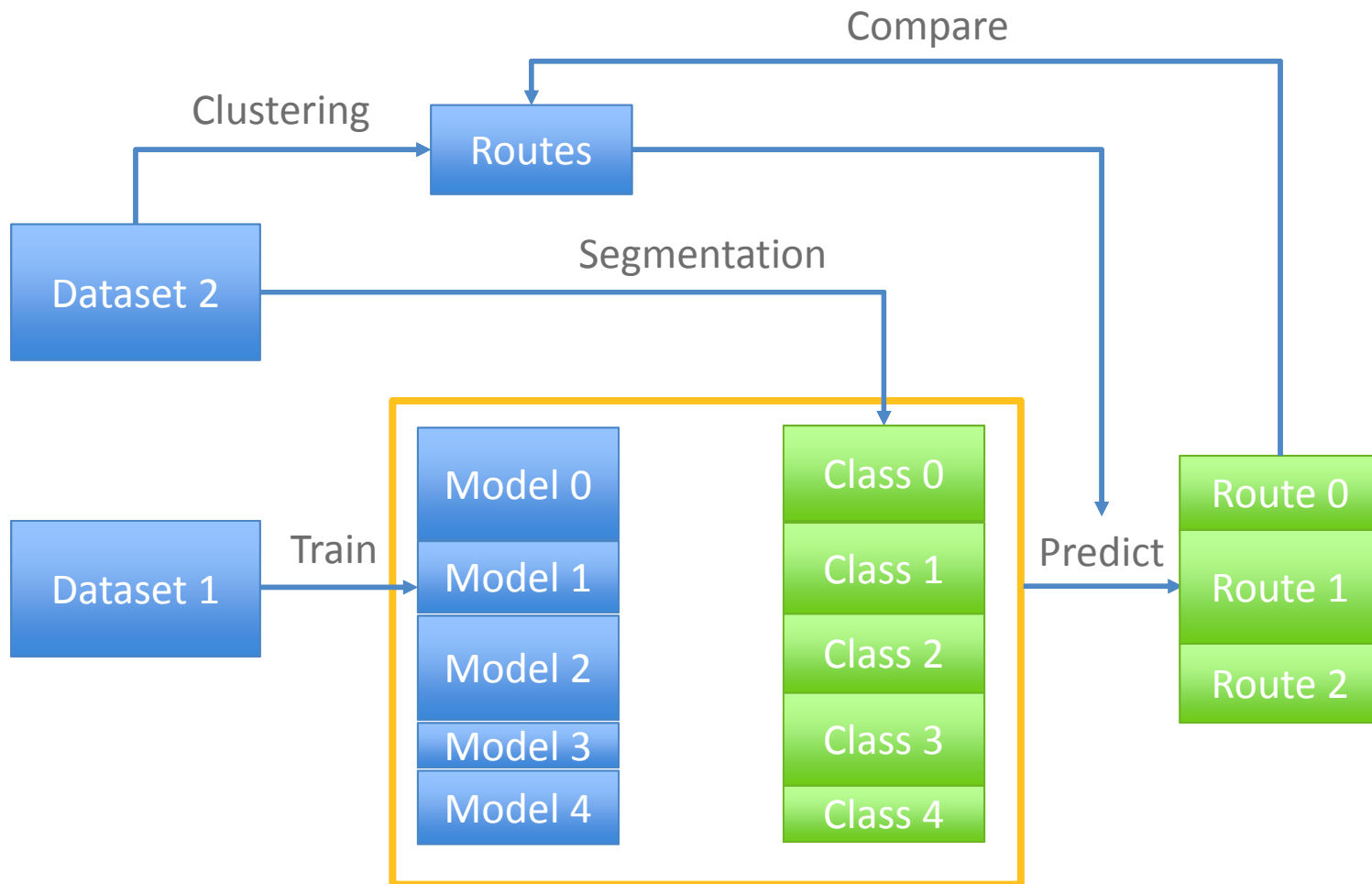
# Testing Results
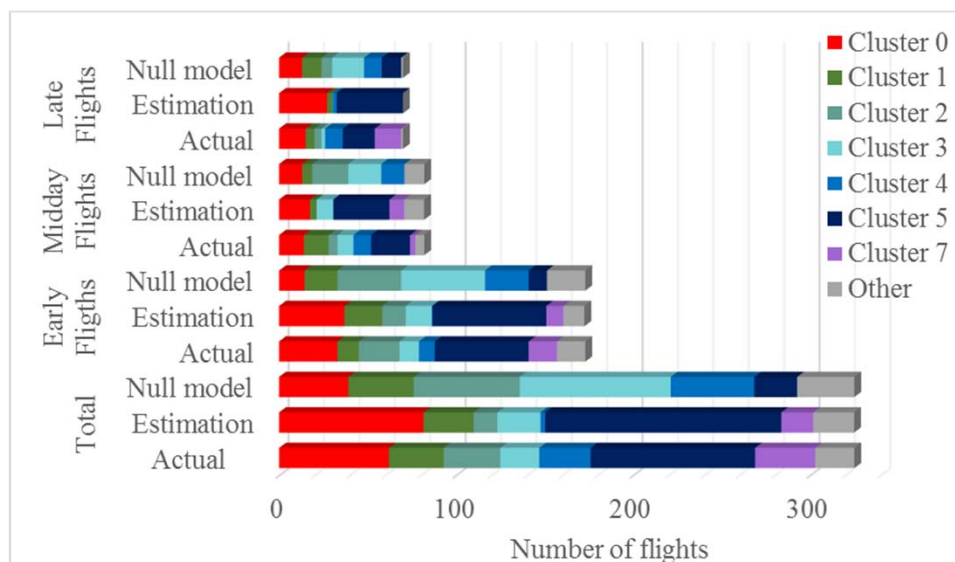# Canary Islands-London



- The model captures:
  - behaviour of new airline (Norwegian)
  - airlines changing route options
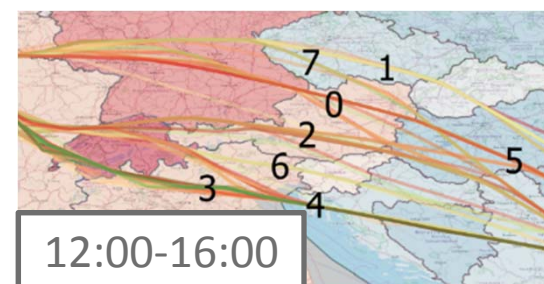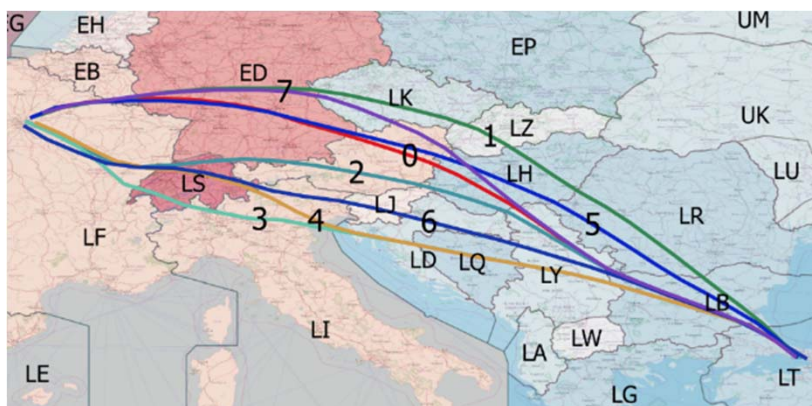- Improvements with respect to null model

# Testing Results
# Istanbul-Paris



The model captures:
- other routes considered (7)
- significant change in charges

Much better than null model

| Cluster | Charges (train) | Charges (testing) | Regulations (train) | Regulations (testing) |
|---|---|---|---|---|
| 0 | 1188 | 1305 | 0.15 | 0.04 |
| 3 | 1204 | 1260 | 0.07 | 0.02 |



12:00-16:00

# Applicability

- Potential for pre-tactical demand forecast

- Range of applicability needs to be clearly identified:

  - Training data requirements

  - Prediction error measurement

  - Generalisation to other ODs

# Future Research Directions

- Better explanatory variables

    - Other indicators

    - Congestion as a function of time

    - Other flight inputs: wind, type of regulation, route availability...

- Training with several years' data

- Continuous training/prediction (automatic adaptive training data)

- Combination with model-based approaches (cost optimisation)

SIDs, Beograd, 29th November 2017

Combining Visual Analytics and Machine Learning

for Route Choice Prediction

Application to Pre-Tactical Traffic Forecast

# Thank you very much
# for your attention!

SESAR
JOINT UNDERTAKING

Founding Members

EUROPEAN UNION    EUROCONTROL