

NOMMON



Unveiling airline preferences for pre-tactical route forecast through machine learning

NOMMON: Manuel Mateos, Ignacio Martín, Rubén Alcolea, Ricardo Herranz and Oliva G. Cantú-Ros
UPC: Xavier Prats

 11th SESAR Innovation Days



founding members



01

INTRODUCTION

02

ROUTE PROBABILITY MODEL

03

EXPERIMENTS & RESULTS

04

CONCLUSIONS & NEXT STEPS

11th SESAR Innovation Days





Introduction



11th SESAR Innovation Days



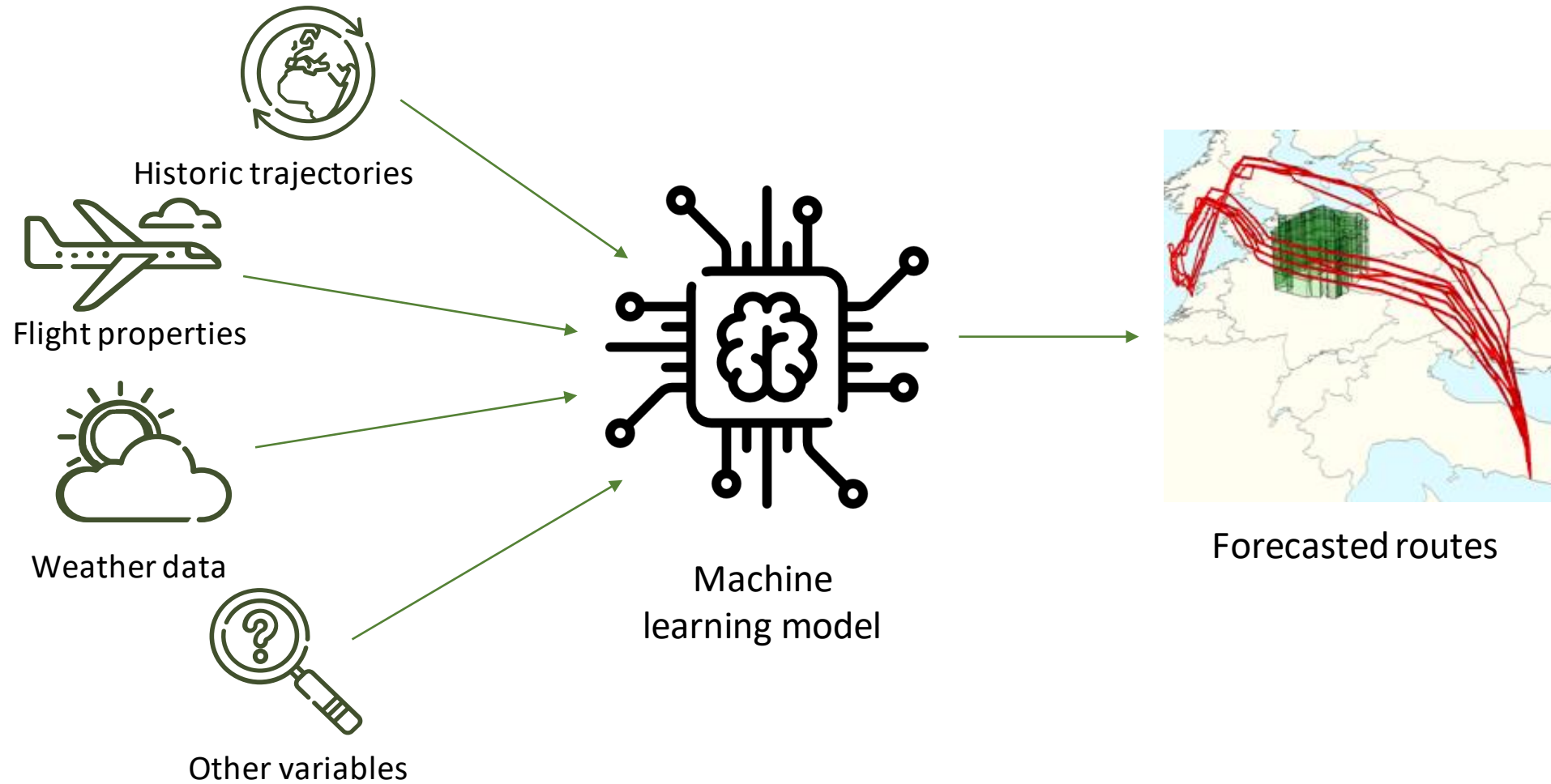
founding members



Background and motivation

1 Need to be addressed	2 Demand forecast potential benefits	3 Challenge	4 Opportunities
Nowadays, the Network Manager needs flight plans from Airlines prior the day of operations to detect demand and capacity imbalances . As flight plans are not available up to a few hours before, the current solution (PREDICT) use the flight plan from the previous week, which is a suboptimal approach.	Accurate demand forecast is one of the key enablers of ATFCM service provision Better resource allocation from the Network Manager would potentially translate into fewer regulations in the day of operations	Traditional models that predict trajectories, often require sensitive information that Airlines are not willing to share, such as TOW, Cost Index, etc.	Using historic data and Machine Learning , it is possible to predict the route in the pre-tactical phase (before day of operations) and this way reduce the regulations.

Prediction problem overview



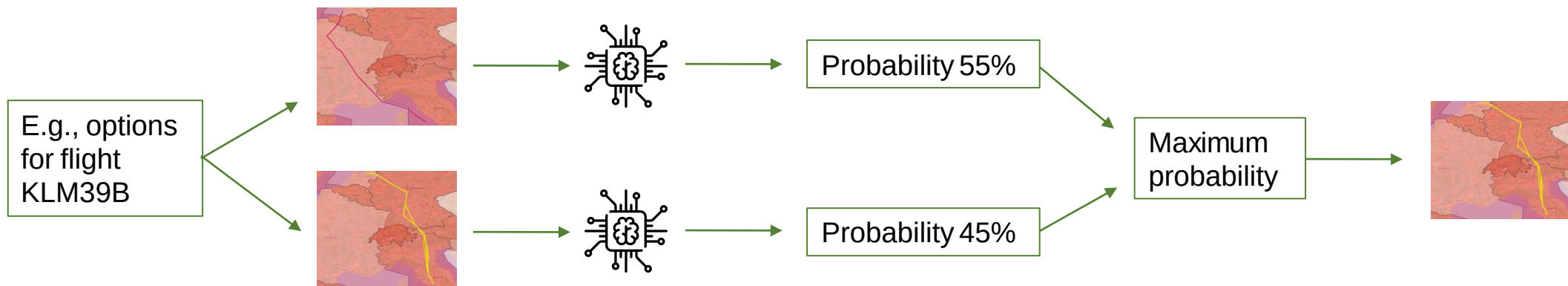
Route Probability Model

 11th SESAR Innovation Days



Probability model approach

- Emulates the **airline flight planning** (One model per airline)
- Predicts the **probability to choose each available route separately** for each flight
- The observation of a route that was **valid and not selected** is also a valid observation
- Features selected are expected to explain the selection/not selection of any route
- All route-related features considered relative to the available routes for a given OD and AIRAC
- Non observed routes can be considered just by computing the **features associated to them**



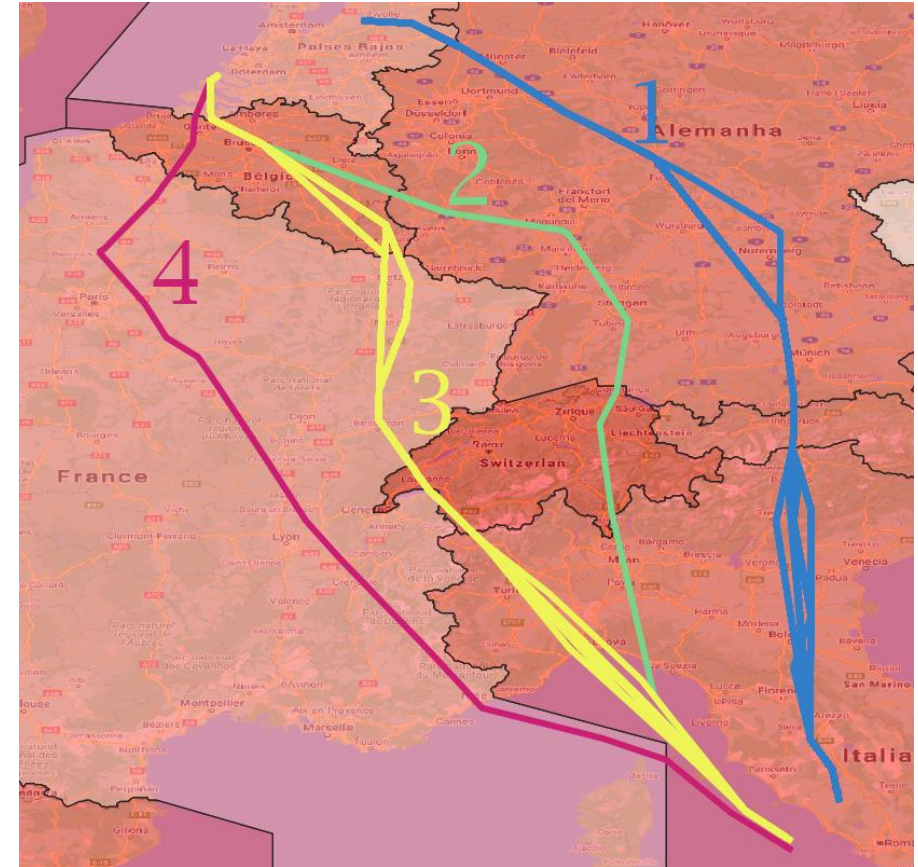
Probability model approach

- Emulates the **airline flight planning** (One model per airline)
- Predicts the **probability to choose each available route separately** for each flight
- The observation of a route that was **valid and not selected** is also a valid observation
- Features selected are expected to explain the selection/not selection of any route
- All route-related features considered relative to the available routes for a given OD and AIRAC
- Non observed routes can be considered just by computing the **features associated to them**



Route characterisation: clustering

- Routes are objects composed by hundreds of points. Changing just a point will result in a different route, but **small differences does not affect traffic demand**.
- Clustering helps to simplify the problem by classifying the **routes in similar groups**.
- Clustering algorithms typically require three different key elements:
 - **Input data:** the routes
 - **Distance metric:** area between routes
 - **Aggregation algorithm:** DBSCAN

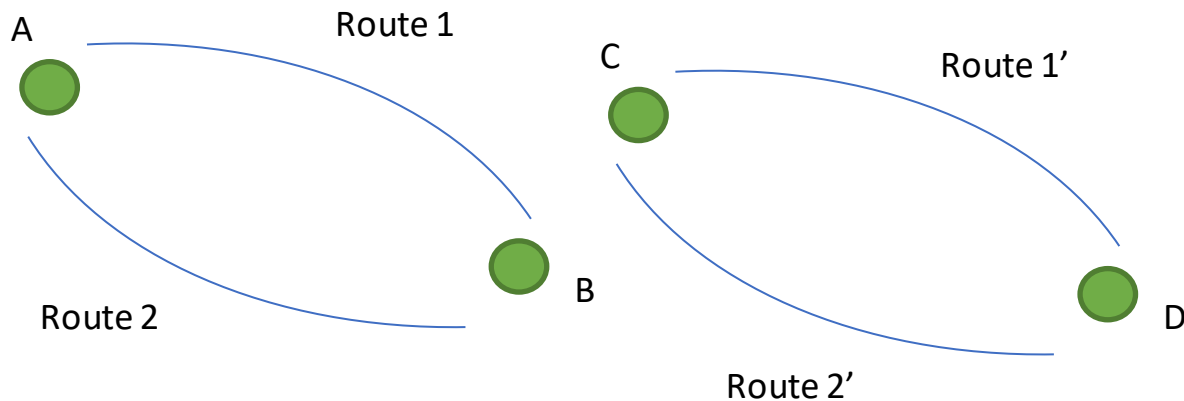


Clustering example for pair LIRF-EHAM

Feature assignment example

Simplified example

- Consider 2 OD pairs with only 2 possible routes each
- Simplification of features:
 - Effective length (including wind)
 - Route charges



CLASS. MODEL		Route 1/1'		Route 2/2'		Cluster used
OD Pair	Flight ID	Length 1	Charges 1	Length 2	Charges 2	
A-B	1	1000	330	1030	400	1
A-B	2	1020	330	960	400	2
C-D	3	1020	345	980	350	2'
...	...	...	...	...	...	...

Differences calculation

Obs. id	Flight ID	Cluster	Length diff	Charges diff	Flown?
1	1	1	0	0	1
2	1	2	30	70	0
3	2	1	60	0	0
4	2	2	0	70	1
5	3	1'	40	0	0
6	3	2'	0	5	1
...	...	...	...	...	...

Route Features

Variable Name	Description
Route length	The length in kilometres of a given route
Wind length	Length of the route in kilometres adjusting the effect of the along wind
Charges	The charges paid for the current route for a given aircraft
Fuel cost	Estimation of the cost of fuel for each given route
Direct costs	Sum of the fuel and charges costs
CAPE	The Convective Available Potential Energy is variable that tends to be high when thunderstorms occur (It is therefore used as a storm proxy)
K-index	K-index is a weather metric that approximates the probability of a thunderstorm to happen
Humidity	The relative humidity observed along the route, that is a requisite for thunderstorms to occur
Local wind at origin/destination	Variable that measures how aligned and in what value local wind at the airport is
Military zones	The route crosses a closed military zone, not use as a feature but to discard routes This information is not available so we used a proxy: consider active military zones during periods of very low traffic

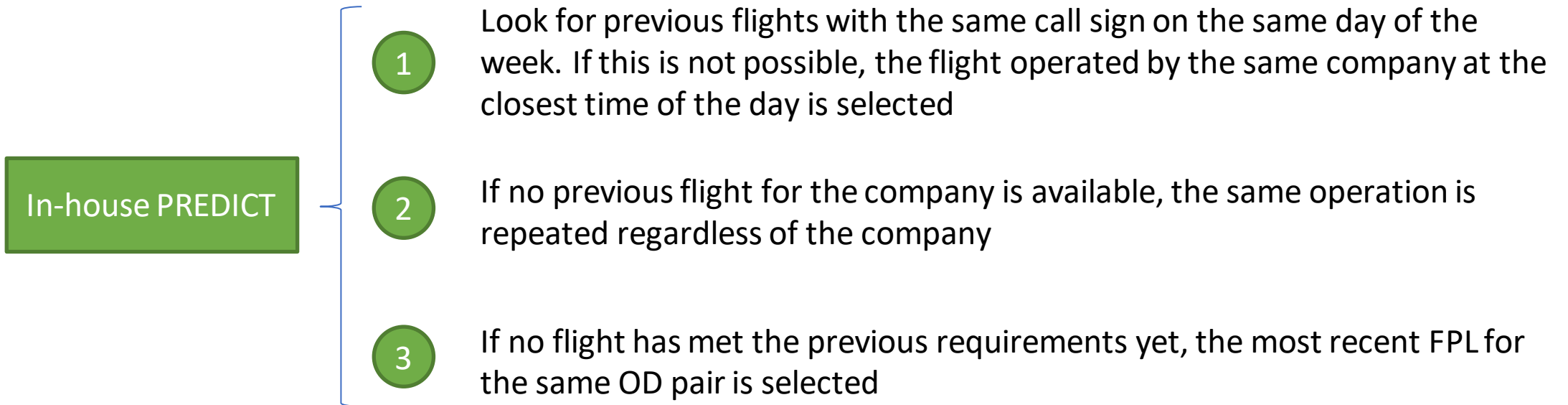
General Features

Variable Name	Description
DoW	The day of week of the flight codified accordingly
Time of flight	The ETOT hour of the flight
DoY	The day of year the flight takes place in
Od pair	Variable for each od_pair in each dataset
Longitude diff	Geodesic longitudinal separation between origin and destination
Latitude diff	Geodesic latitudinal separation between origin and destination
MTOW	Maximum take-off weight of the aircraft used
Airport population	Population density of the Origin/Destination surroundings areas
Airport GDP	Gross Domestic product of the Origin/Destination surroundings areas
Daily flights	Number of flights for each od pair and day
Daily share	Airline's flight share for each od pair and day

Benchmarking exercise: PREDICT

- The models accuracy has been measured and compared against that of PREDICT, the tool currently used by EUROCONTROL
- An in-house implementation of the tool has been done following the information available from the Network Manager documentation

Our implementation takes the following approach for each flight:



Benchmarking exercise: Enhanced model

The Enhanced model is the predecessor of the probability model. It is based on the training of a machine learning multiclass classification model by OD pair

- The model is trained with a **fixed number of classes** (route clusters), it does not allow new routes
- The output of the predictions can only be one of the routes observed in the training
- The inclusion of variables that depend on the route (e.g., existence of a convective event) implies one extra feature for each one of the routes observed
- This approach **requires** conceptually **a different model for each OD (OD pair-based)**, indeed the number of features will be different in OD pairs with different number of observed routes
- The Enhance model limitations have motivated the development of the probability model, therefore it has also been used for benchmarking purposes.



Experiments & Results

 11th SESAR Innovation Days



Case study: KLM Royal Dutch Airlines

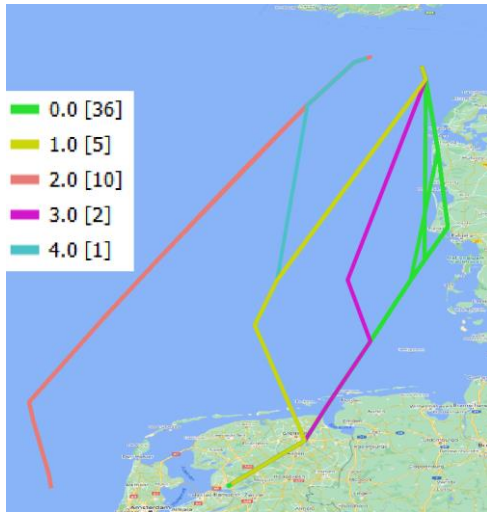
- Airline (KLM), assumed to have a uniform decision making process
- Principal European airports covered
- Over **half million observations** considered (train 1802-1812 and validation 1813)
- OD **pairs over 5,000 Km have been discarded**
- Machine learning algorithm is a **decision tree**, it provides the probabilities and the most probable route is selected
- A flight is considered as correctly predicted when the predicted route matches the observed one
- Different training strategies has been tested. It seems like it is important to train the model with data from the **same season** (winter/summer)

Model ID	Training AIRACs	Validation AIRAC	Accuracy
1	1812	1813	0.814
2	1810-1812	1813	0.831
3	1807-1812	1813	0.834
4	1802-1812	1813	0.852
5	1802,1811,1812	1813	0.849
6	1802,1803,1811,1812	1813	0.854
7	1802,1803,1804,1810, 1811,1812	1813	0.844
8	1802,1803,1804,1811,1812	1813	0.860

Benchmark		
PREDICT	Enhanced model	Probability
81.5%	82.5%	86.0%



Ad-hoc analysis, non-observed routes



OD pair Kristiansand (Norway) - Amsterdam (ENCN - EHAM) presents a new route in AIRAC 1813 (route 3 in purple):

- Routes appears twice in this AIRAC.
- The probability model **predicts both cases**
- The Enhanced model and PREDICT are not able to do it

ID	Route observations	PREDICT		Enhanced		Prob. model	
		Correct assignments	Wrong assignments	Correct assignments	Wrong assignments	Correct assignments	Wrong assignments
0	36	23	12	27	10	27	2
1	5	1	4	1	0	3	1
2	10	4	8	6	10	9	10
3*	2	0	0	0	0	2	0
4	1	0	0	0	0	0	0
Noise	0	0	2	0	0	0	0
Total	54	28	26	34	20	41	14
Perc.	-	51.9%	48.1%	63.0%	37.0%	75.9%	24.1%



Conclusions & Next Steps



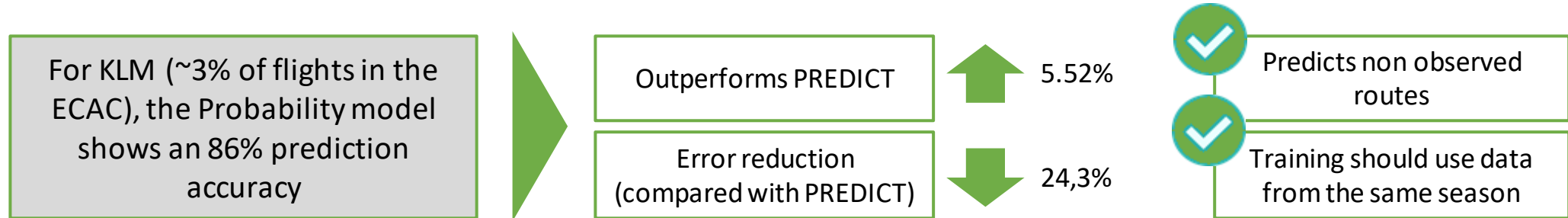
11th SESAR Innovation Days



founding members



Conclusions & Next steps



Next steps:

- Extend the evaluation of the proposed methodology to the main airlines operating in the ECAC area
- Explore other **segmentation** options (od pair clustering, airline clustering, etc.)
- Test more sophisticated machine learning models

Acknowledgments

- Manuel Mateos' PhD is funded by the 1st SESAR ENGAGE KTN Call for PhDs, developed in collaboration between Nommon and the Technical University of Catalonia. This PhD study has received funding from the SESAR Joint Undertaking under the European Union's Horizon 2020 research and innovation programme under grant agreement No 783287. The opinions expressed herein reflect the authors' view only. Under no circumstances shall the SESAR Joint Undertaking be responsible for any use that may be made of the information contained herein.
- I would like to acknowledge the support of EUROCONTROL to this work by facilitating data access and expert advice. Particularly, the authors would like to thank Dr. Ramon Dalmau, Ms. Stella Saldana, Mr. Stefan Steurs, Mr. Eric Allard and Mr. Francis Decroly for their advice on the definition of the models and the design of the evaluation experiments.