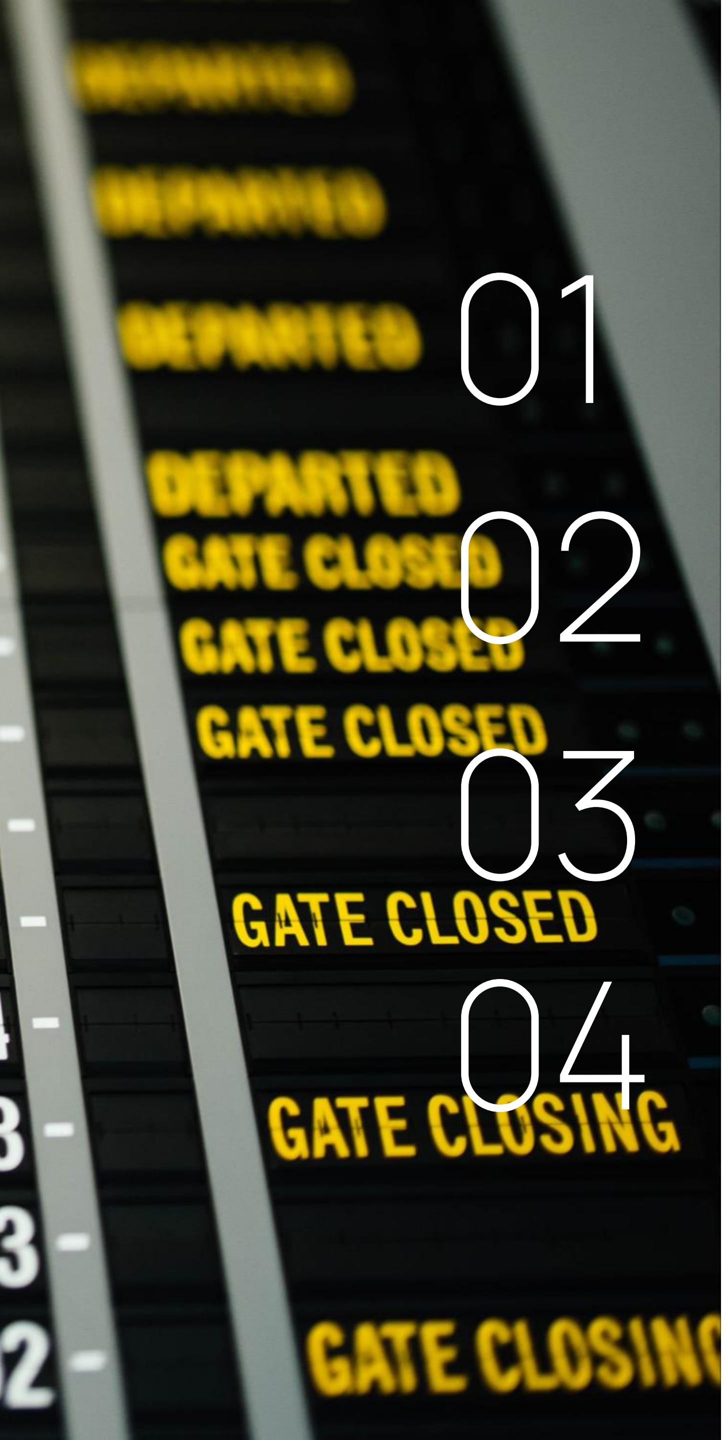
# ENGAGE PhD: Data-based Pre-tactical Trajectory Prediction

The present research is part of Manuel Mateos' PhD. Which is funded by SESAR's Engage Knowledge Transfer Network, which 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.
The authors would also like to acknowledge the support of the Spanish Centre for Industrial Development (CDTI) through the PRETA project (Grant no. IDI-20190029).




01 PhD INTRODUCTION

02 METHODOLOGY

03 INTERIM RESULTS

04 CONCLUSIONS & NEXT STEPS

# PhD INTRODUCTION

# PhD INTRODUCTION

## 1 CONTEXT

Nowadays, the Network Manager need flight plans from Airlines prior the day of operations to assign resources to each particular sector. They currently use the flight plan (composed by route and RFL) from the previous week.

Therefore demand forecasting methods present a discontinuity on the pre-tactical phase and may have significant margin for improvement

## 2 Potential Benefit

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, therefore, fewer flight delays due to that reason

## 3 Current limitations

Traditional models that predict trajectories, often based on the optimal FPL, require sensitive information that Airlines are not willing to share, such as TOW, Cost Index, etc.

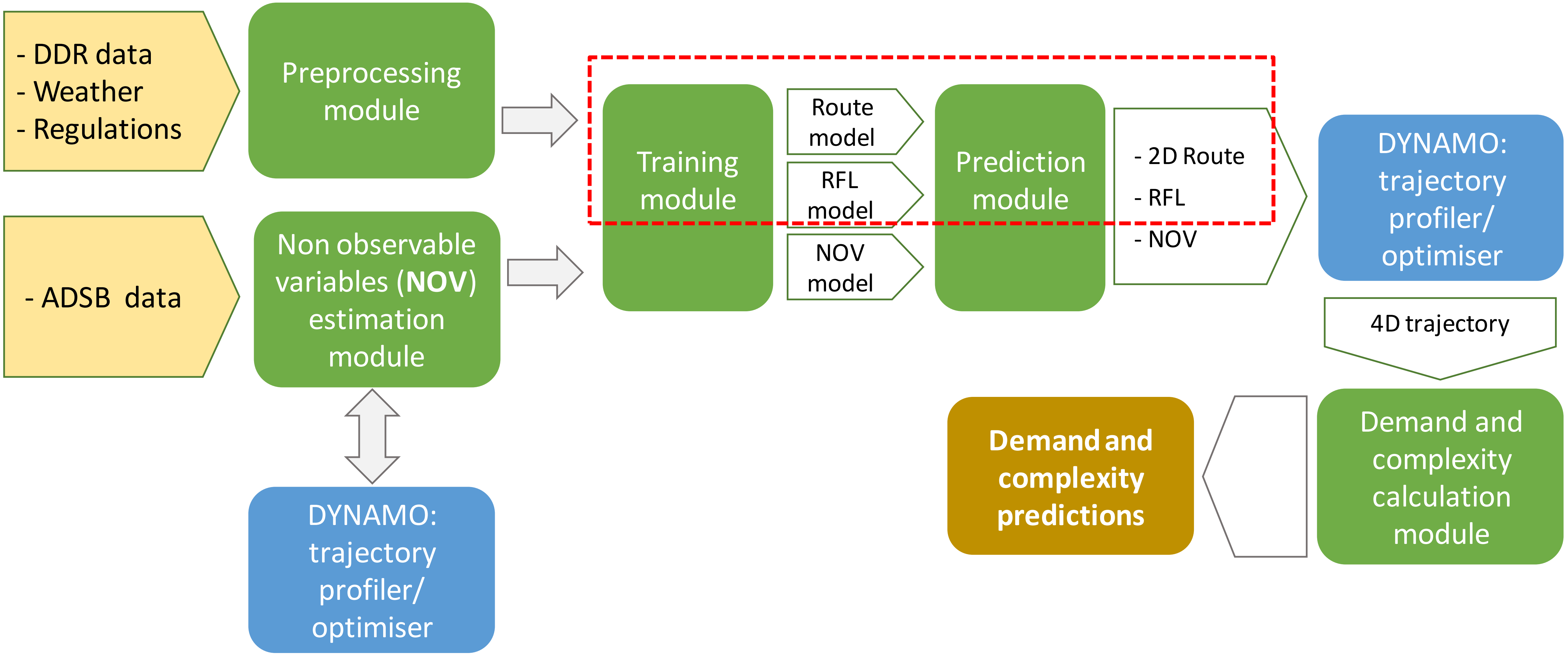
## 4 ML Model for trajectory prediction

Using historic data and Machine Learning, present an opportunity to address these limitations.

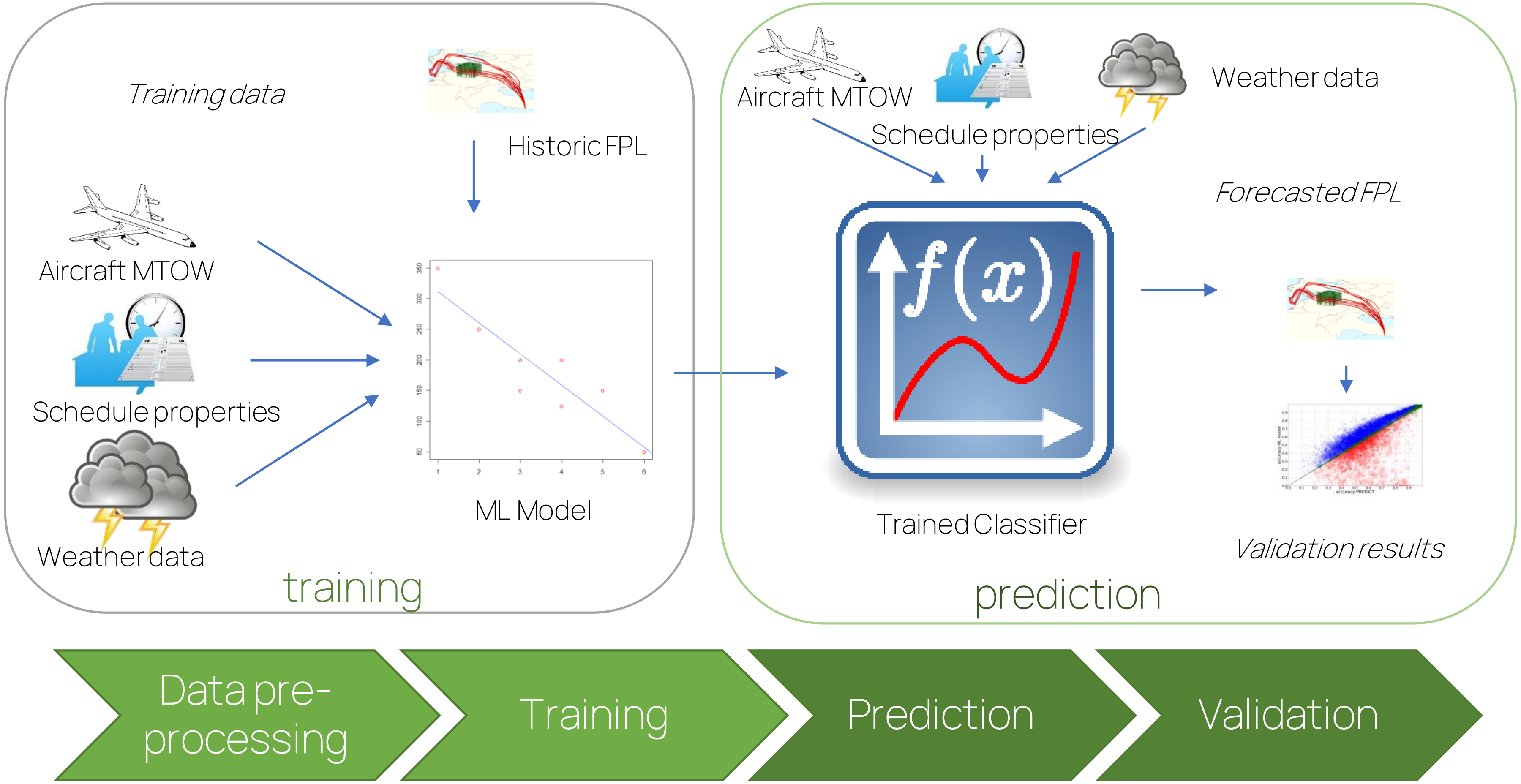
The objective of this PhD is the exploration of different ML techniques to improve current trajectory prediction models.

# Methodology

# General PhD overview



# Trajectory prediction problem overview



# Machine Learning models

- We have developed two different models for route and RFL, with different features, taking an iterative approach, both use the Random Forrest classification algorithm

## Basic Model

- AU: one-hot encoding of the airline ICAO code.
- Day of week: one-hot encoding of the day of the week when the flight takes place
- Hour: sine and cosine of the expected take-off time (ETOT) hour
- Day of the year: sine and cosine of the day of the year
- Aircraft mass: maximum take-off weight (MTOW) of the aircraft model

## Enhanced Model

All features from Basic Model, plus:

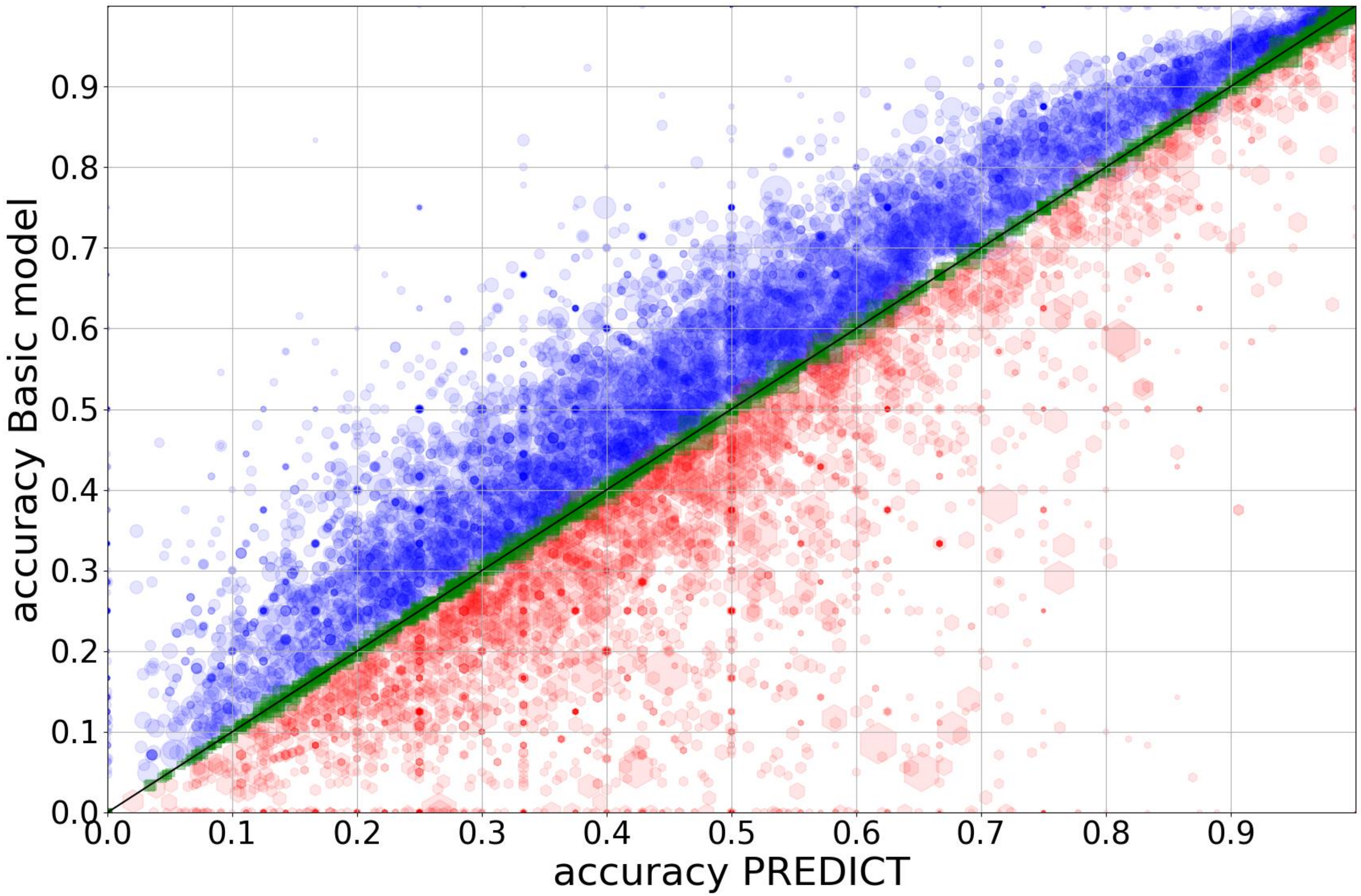
- Wind: wind projection along the flight path, positive (tailwind) or negative (headwind)
- Convective phenomena probability: humidity, CAPE and k-index
- Past regulations: used as an indicator of the expected congestion levels
- Local wind at airports: as a proxy of the airport runway configuration

RFE is used to reduce the number of variables

**ML models are benchmarked against PREDICT**, which is the current EUROCONTROL solution for predicting flight plans. It uses the flight plan provided by the airline the previous week

# Interim Results

# Basic model results




- Tested on a total of 10,807 OD pairs (Full ECAC coverage). Each point represents one OD
- The size of the point represents the number of flights
- Training (AIRACs: 1801-1812), testing (1813). Model for OD pair
- Blue: Basic > PREDICT (48.7%)
- Red: Basic < PREDICT (33.5%)
- Green: Basic = PREDICT (17.8%)

# Enhanced model results

| OD pair | PREDICT | Basic Model | Enhanced Model |
|---|---|---|---|
| EDDF-UUEE | 0.451 | 0.574 | 0.611 |
| EDDK-LTAI | 0.420 | 0.420 | 0.280 |
| EDDT-LEPA | 0.430 | 0.474 | 0.511 |
| EGLL-OMDB | 0.348 | 0.321 | 0.418 |
| EHAM-LIRF | 0.291 | 0.283 | 0.385 |
| LEPA-EDDT | 0.348 | 0.326 | 0.318 |
| LFPG-LGAV | 0.349 | 0.302 | 0.403 |
| LFPO-LPPT | 0.332 | 0.349 | 0.399 |
| LGAV-LFPG | 0.223 | 0.198 | 0.339 |
| LIRF-EHAM | 0.363 | 0.257 | 0.273 |
| LPPT-LFPO | 0.384 | 0.524 | 0.517 |
| LTAI-EDDK | 0.289 | 0.200 | 0.222 |
| OMDB-EGLL | 0.387 | 0.443 | 0.475 |
| UUEE-EDDF | 0.416 | 0.439 | 0.445 |
| Global acc. | 0.360 | 0.376 | 0.420 |

- Due to computation restrictions, only a few pairs were run
- OD pairs selected attending to the availability of significantly large datasets while trying to get some geographical heterogeneity
- Enhanced model outperforms PREDICT by 16.5% in the selected OD pairs, increasing accuracy in most OD pairs
- Results may be biased due to the fact that these pairs present lower PREDICT accuracy than the average
- In general terms, the analysed OD pair show the expected accuracy progression: PREDICT < Basic < Enhanced

# Conclusions & Next Steps

# Conclusions & next steps

Machine Learning models outperform current solution (PREDICT)

| | | |
|---|---|---|
| Basic Model | ↑ | 2.3% |
| Enhanced Model | ↑ | 16.5% (for a pair selection) |

Next steps:

- Investigate the applicability of the Enhanced Model to the entire network
- Explore the inclusion of new features (crosswind, proximity to a severe weather event, etc.)
- Develop automatic algorithm selection for each OD pair
- Explore the development of models covering several OD pairs