FLUGHAFEN ZÜRICH

Predicting the departing runway of a flight based on a classification tree model

Benjamin Perucco
Data Analyst
Flight OPS Analytics

FLUGHAFEN ZÜRICH

Outline

- Facts about Zurich airport
- Who we are and what we do
- Why predicting the departing runway is important
- Building prediction models
- Five prediction models in more detail
- What’s next?
- Questions?

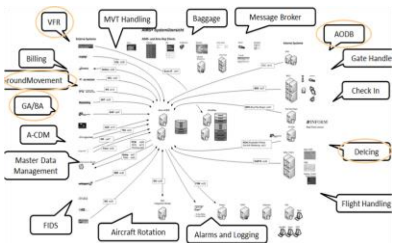
Facts about Zurich airport

- 31 mio passengers
- 280’000 movements
- 760 daily flights in average
- 77 airlines
- 206 destinations
- 3 terminals
- 3 runways
  → 10/28: 2’500m
  → 16/34: 3’700m
  → 14/32: 3’300m

Figures based on year 2018

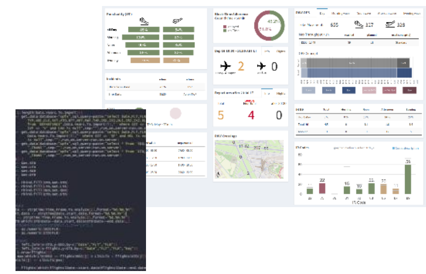
Who we are and what we do

Coordination of IT projects



6 data analysts, 1 trainee

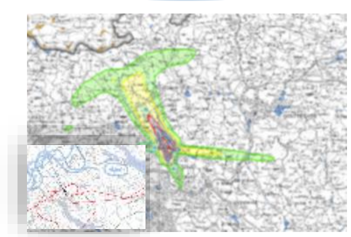
Data analytics



Flight OPS Analytics

Real time simulation

Noise calculation

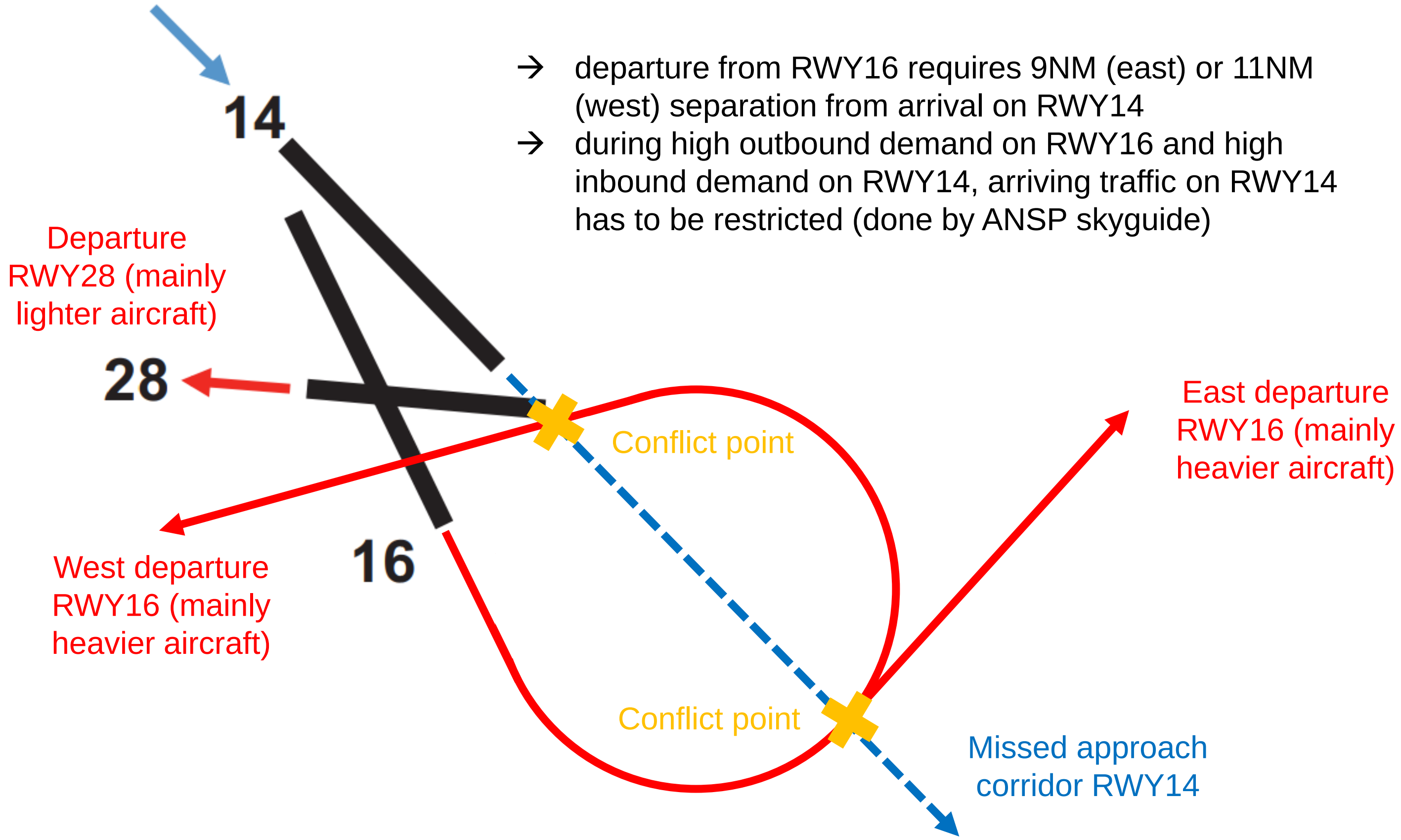


Fast-time simulation

Why predicting the departing RWY is important

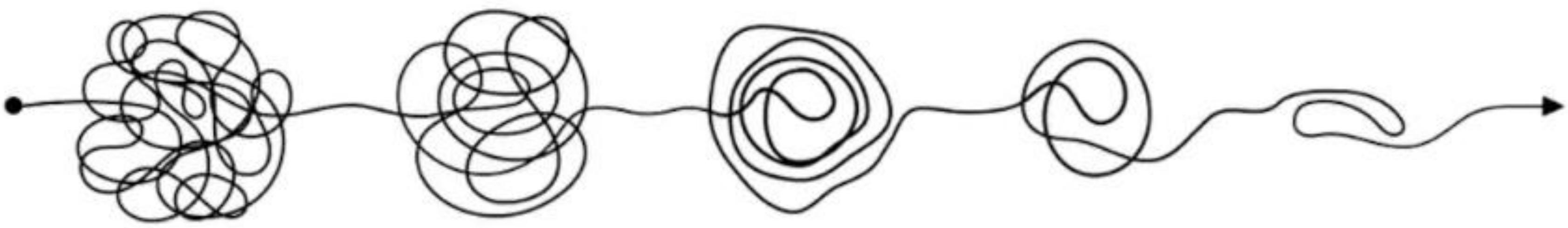
- Runway layout at Zurich airport lead to several crossings on ground and in air

→ departure from RWY16 requires 9NM (east) or 11NM (west) separation from arrival on RWY14
→ during high outbound demand on RWY16 and high inbound demand on RWY14, arriving traffic on RWY14 has to be restricted (done by ANSP skyguide)



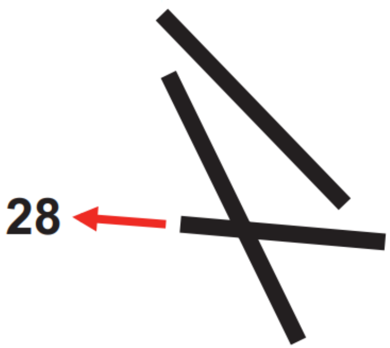
Building prediction models

- Various databases available to build models from
- Airport operational database contains a lot of timestamps and other information, i.e. scheduled time, airline, aircraft type, destination, airborne time, etc.
- Meterological data are received from several stations around the airport in different resolution (i.e. wind information every 3s, temperature and pressure every 30min)
- Starting with simple and progressing to more complex models



1st model
Assumption all aircraft take off from RWY28

- We assume all aircraft take off from the shorter RWY28
- An oversimplification of course



Global confusion matrix for all aircraft types
Model: assumed all takeoffs on RWY28
based on test dataset

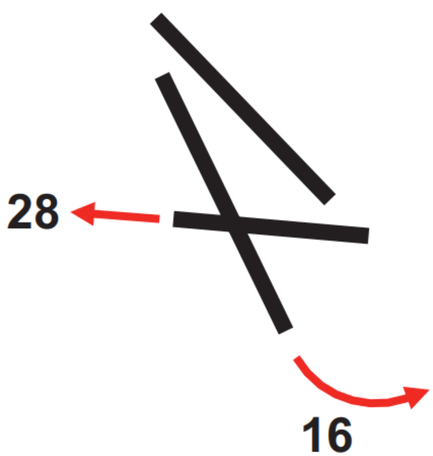
Accuracy = 81 %		Reference		
		RWY16	RWY28	
Prediction	RWY16	TP = 0	FP = 0	PPV = NaN %
	RWY28	FN = 8830	TN = 37696	NPV = 81 %
		TPR = 0 %	TNR = 100 %	

- Still an accuracy of 81% is achieved

only LX and WK flights

2nd model
Heavy aircraft and A321 on RWY16, rest on RWY28

- Heavy aircraft (like A333, B77W, etc.) and A321 take off from the longer RWY16
- Medium and light aircraft take off from the shorter RWY28



Global confusion matrix for all aircraft types
Model: assumed A321 and heavies on RWY16, rest on RWY28
based on test dataset

+8.9%

Accuracy = 89.9 %		Reference		PPV still somewhat low
		RWY16	RWY28	
Prediction	RWY16	TP = 7988	FP = 3875	PPV = 67.3 %
	RWY28	FN = 842	TN = 33821	NPV = 97.6 %
		TPR = 90.5 %	TNR = 89.7 %	

+16.6%

+90.5% -10.3%

only LX and WK flights

3rd model
Predicting the takeoff RWY based on historical data

- Probability for RWY16 or RWY28 departure is based on historical data
- Excerpt is provided below

```
# A tibble: 32,585 x 7
   FLT    CSG     FLC   RWY   ABT                 ICT   IDS  
   <chr>  <chr>   <chr> <chr> <chr>               <chr> <chr>
 1 LX154  SWR154  LX    16    2018-01-27 10:02:00 A333  VABB 
 2 LX066  SWR66R  LX    16    2015-04-11 09:59:00 A333  KMIA 
 3 LX1248 SWR121X LX    28    2019-03-12 07:05:00 A320  ESSA 
 4 LX1582 SWR158H LX    28    2018-01-11 17:27:00 B77W  LOWW 
 5 LX1250 SWR129Q LX    28    2015-08-21 13:32:00 A320  ESSA 
 6 LX2804 SWR51KJ LX    28    2017-03-09 09:46:00 A321  LSGG 
 7 LX160  SWR160  LX    16    2015-03-18 13:37:00 A343  RJAA 
 8 LX040  SWR40   LX    16    2016-06-15 14:04:00 B77W  KLAX 
 9 LX038  SWR38   LX    16    2015-12-07 13:58:00 A343  KSFO 
10 LX1200 SWR1200 LX    16    2015-08-07 17:04:00 A320  EFHK 
# ... with 32,575 more rows
```

- Probability for RWY16 or RWY28 departure is estimated as a function of airline, aircraft type, destination, weekday and month
- If no estimate can be found, a stepwise reduction of complexity is introduced, i.e. function of aircraft type and destination in the simplest case

Selection of training and test data

- Split data into a training and test dataset
- 80% as training data, 20% as test data
- Avoids systematic error when training the models

ID	ICT	RWY
1	A321	28
2	A333	16
3	A320	28
4	A320	16
5	A343	16
6	B77W	16
7	BCS1	28
8	A321	28
9	A319	28
10	A333	16

Random sampling

ID	ICT	RWY
8	A321	28
10	A333	16
3	A320	28
2	A333	16
6	B77W	16
1	A321	28
7	BCS1	28
5	A343	16
9	A319	28
4	A320	16

Split data

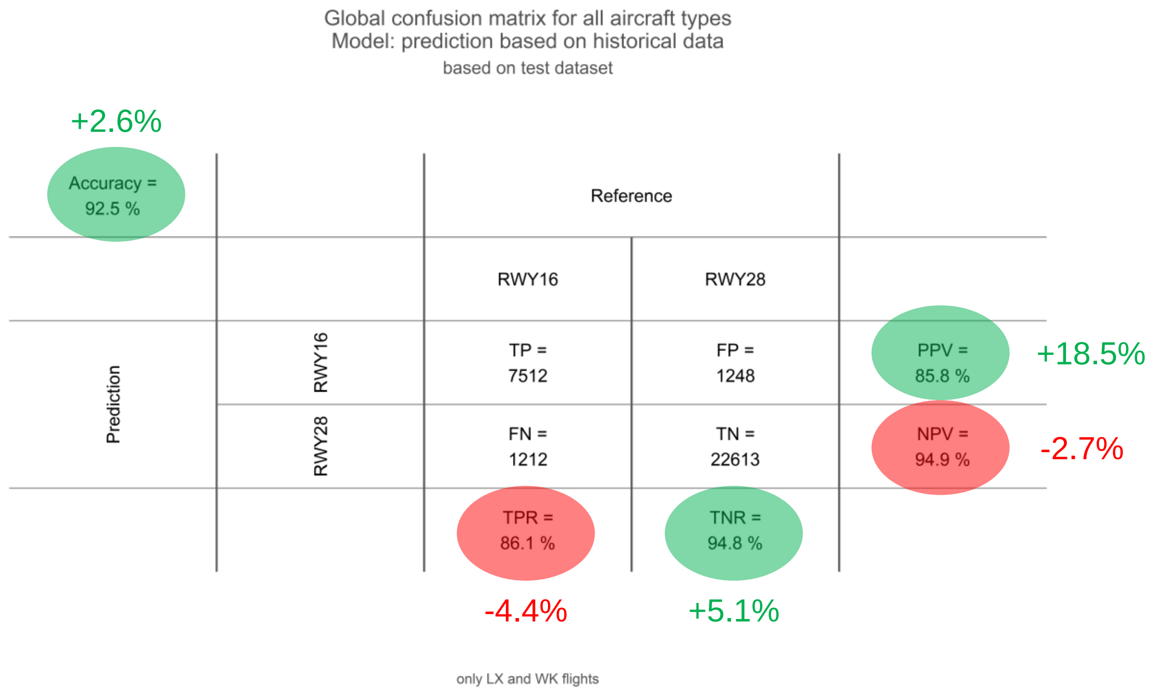
ID	ICT	RWY
8	A321	28
10	A333	16
3	A320	28
2	A333	16
6	B77W	16
1	A321	28
7	BCS1	28
5	A343	16
9	A319	28
4	A320	16

Training data

Test data

3rd model
Predicting the takeoff RWY based on historical data

- Confusion matrix



- The gains dominate and especially the PPV improved considerably (+18.5%)

3rd model
Predicting the takeoff RWY based on historical data

- Gives already a good accuracy
- Pretty modest with respect to required data
- Suitable to make good predictions weeks or even months ahead

- For day of operation prediction too static
- For example, depending on final takeoff weight an A321 takes off either on RWY28 or RWY16

4th model
Random forest base model

- Let’s see what a random forst model can do
- From an operational point of view, selection of RWY is influenced by airline, aircraft type and takeoff weight, temperature, pressure and wind
- Random forest base model with predictors: airline, aircraft type and takeoff weight
- Excerpt is provided below

```
# A tibble: 32,585 x 12
   RWY   FLC   ICT      TOW
   <chr> <fct> <chr>  <dbl>
 1 16    LX    A333  221707
 2 16    LX    A333  231819
 3 28    LX    A320   56161
 4 28    LX    B77W  214002
 5 28    LX    A320   67973
 6 28    LX    A321   74382
 7 16    LX    A343  252647
 8 16    LX    B77W  326850
 9 16    LX    A343  257017
10 16    LX    A320   68420
# ... with 32,575 more rows
```

4th model
Random forest base model

- Confusion matrix

Random forest base model: global confusion matrix for all aircraft types
based on test dataset

+3.1% Accuracy = 95.6 %		Reference		
		RWY16	RWY28	
Prediction	RWY16	TP = 7890	FP = 591	PPV = 93 % +7.2%
	RWY28	FN = 834	TN = 23270	NPV = 96.5 % +1.6%
		TPR = 90.4 % +4.3%	TNR = 97.5 % +2.7%	

only LX and WK flights

5th model
Extended random forest model

- Taking into account temperature, pressure and wind as well
- Required to join different datasets

```
# A tibble: 246,490 x 7
   FLT    CSG     FLC     RWY ABT                 ICT      TOW
   <chr>  <chr>   <chr> <dbl> <chr>               <chr>  <dbl>
 1 LX976  SWR976  LX       28 2015-01-01 09:04:00 A321   63174
 2 LX562  SWR50D  LX       28 2015-01-01 09:05:00 RJ1H   34363
 3 LX1188 SWR181X LX       28 2015-01-01 09:08:00 RJ1H   33259
 4 LX750  SWR750  LX       28 2015-01-01 09:11:00 RJ1H   35872
 5 LX1310 SWR1310 LX       28 2015-01-01 09:22:00 A320   57709
 6 LX2902 SWR27TG LX       28 2015-01-01 09:15:00 DH8D   21970
 7 LX2694 SWR2694 LX       16 2015-01-01 09:33:00 A343  232245
 8 LX1176 SWR117F LX       28 2015-01-01 09:38:00 RJ1H   33249
 9 LX1616 SWR121W LX       28 2015-01-01 09:42:00 RJ1H   32102
10 LX2804 SWR84KJ LX       28 2015-01-01 09:49:00 A320   59733
# ... with 246,480 more rows
```

Airport operational database

```
# A tibble: 2,433,198 x 3
   Datetime            Speed Direction
   <chr>               <dbl>     <dbl>
 1 2015-01-01 00:00:00     6        40
 2 2015-01-01 00:01:00     6        40
 3 2015-01-01 00:02:00     6        40
 4 2015-01-01 00:03:00     6        40
 5 2015-01-01 00:04:00     6        40
 6 2015-01-01 00:05:00     6        40
 7 2015-01-01 00:06:00     7        40
 8 2015-01-01 00:07:00     7        40
 9 2015-01-01 00:08:00     7        40
10 2015-01-01 00:09:00     7        40
# ... with 2,433,188 more rows
```

Wind data → averaged wind per minute

```
# A tibble: 237,213 x 4
   Datetime             Temp   QNH
   <chr>               <int> <int>
 1 2005-12-18 01:50:00    -3  1016
 2 2005-12-18 02:20:00    -5  1016
 3 2005-12-18 02:50:00    -4  1017
 4 2005-12-18 03:20:00    -3  1017
 5 2005-12-18 03:50:00    -3  1018
 6 2005-12-18 04:20:00    -4  1018
 7 2005-12-18 04:50:00    -4  1019
 8 2005-12-18 05:20:00    -3  1019
 9 2005-12-18 05:50:00    -5  1019
10 2005-12-18 06:20:00    -5  1019
# ... with 237,203 more rows
```

METAR data

```
# A tibble: 32,585 x 12
   RWY   FLC   ICT      TOW  Temp   QNH Speed Direction
   <chr> <fct> <chr>  <dbl> <dbl> <dbl> <dbl> <fct>
 1 16    LX    A333  221707     6  1029     4 60
 2 16    LX    A333  231819    13  1024    12 230
 3 28    LX    A320   56161    -3  1028     3 130
 4 28    LX    B77W  214002     7  1016     3 270
 5 28    LX    A320   67973    21  1025    12 50
 6 28    LX    A321   74382     8  1023    11 220
 7 16    LX    A343  252647    15  1023     2 280
 8 16    LX    B77W  326850    17  1004     3 10
 9 16    LX    A343  257017     3  1034     5 290
10 16    LX    A320   68420    36  1016     4 140
# ... with 32,575 more rows
```

5th model
Extended random forest model

- Confusion matrix

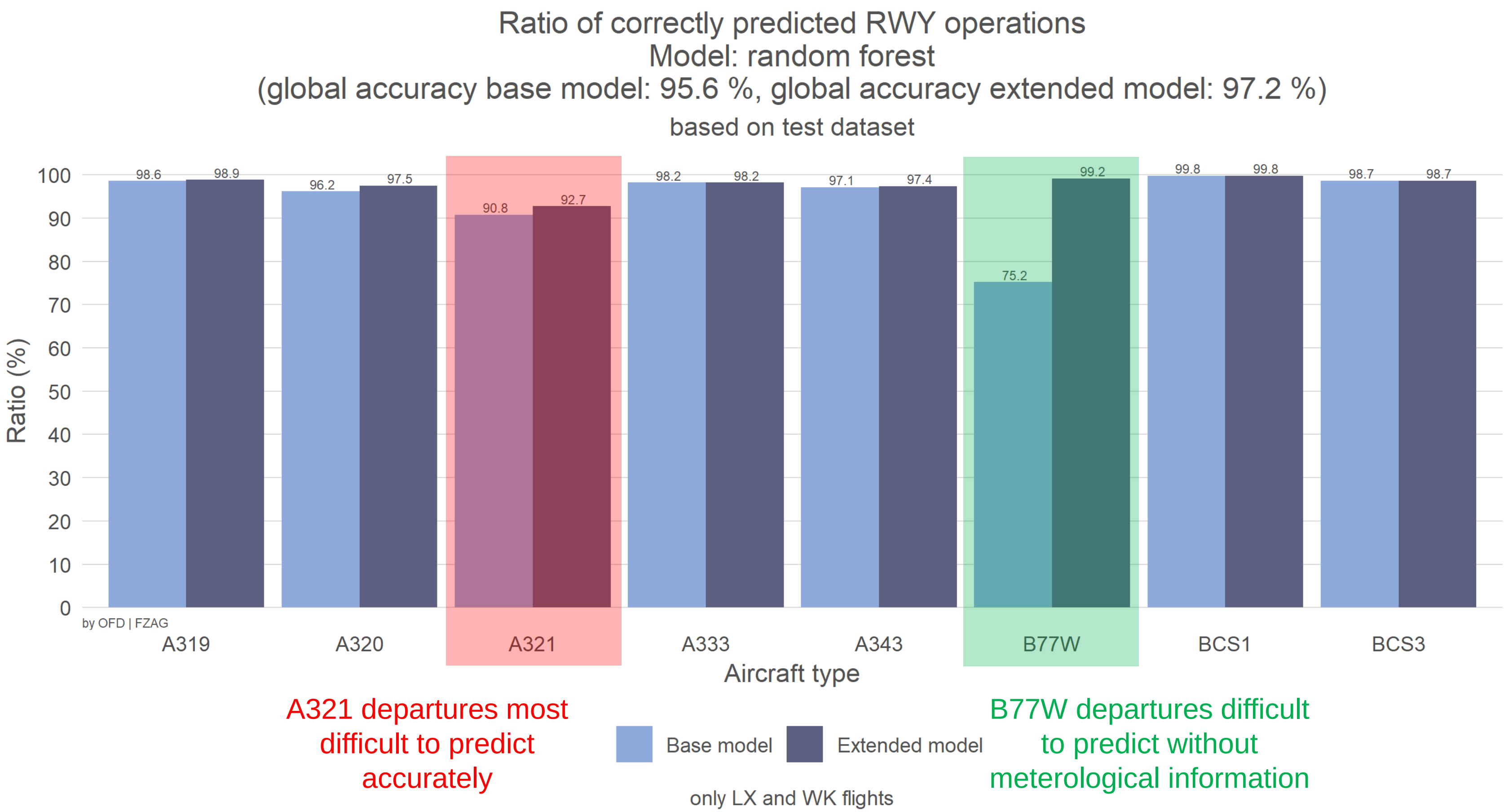
Random forest extended model: global confusion matrix for all aircraft types
based on test dataset

+1.6% Accuracy = 97.2 %		Reference		
		RWY16	RWY28	
Prediction	RWY16	TP = 8124	FP = 308	PPV = 96.3 % +3.3%
	RWY28	FN = 600	TN = 23553	NPV = 97.5 % +1.0%
		TPR = 93.1 % +2.7%	TNR = 98.7 % +1.2%	

only LX and WK flights

- Both random forest models lead to satisfying values in the confusion matrix, let’s see how they compare to each other in more detail

How accurate are the random forest models for the different aircraft types?

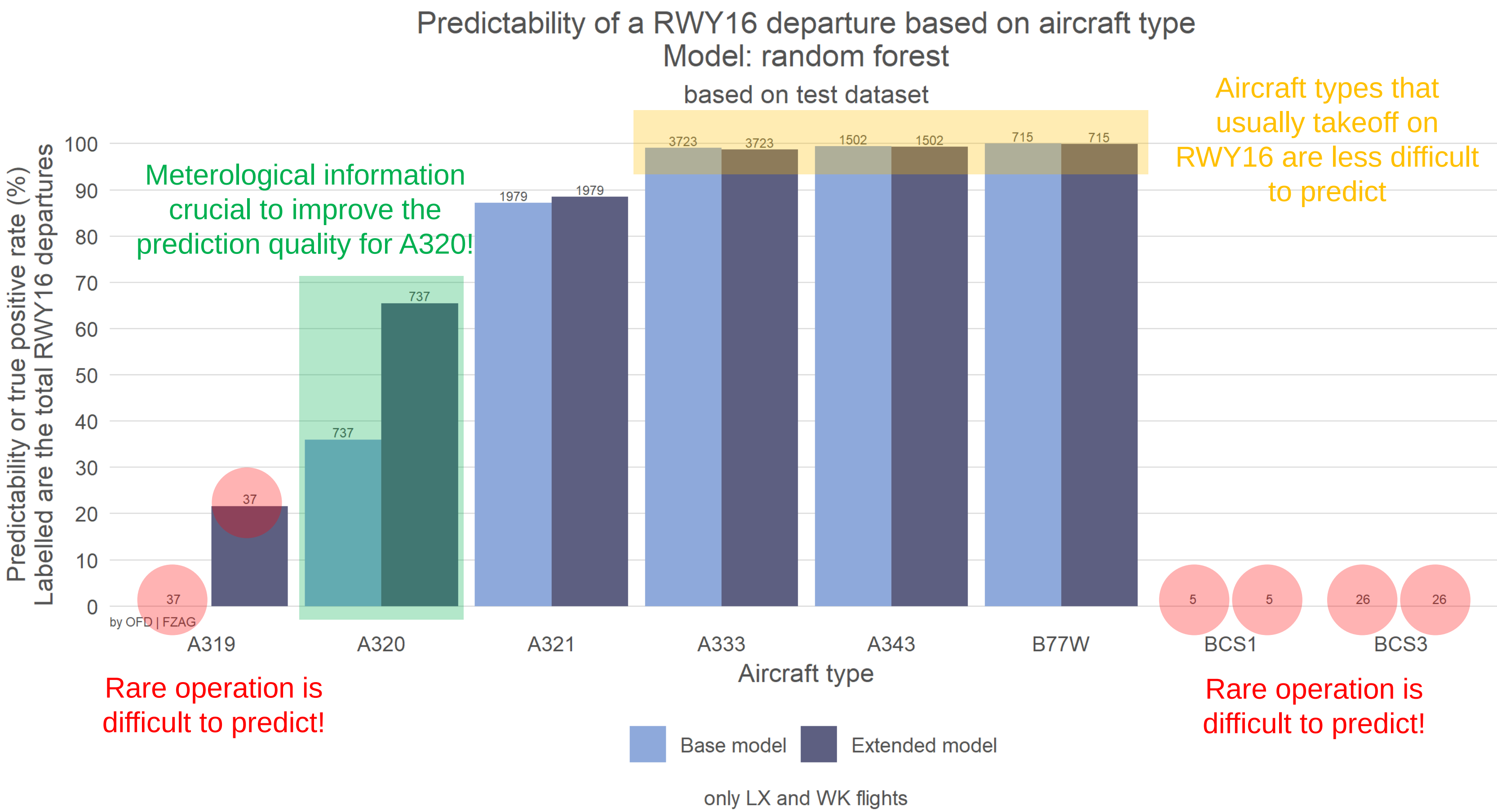


A321 departures most difficult to predict accurately

B77W departures difficult to predict without meterological information

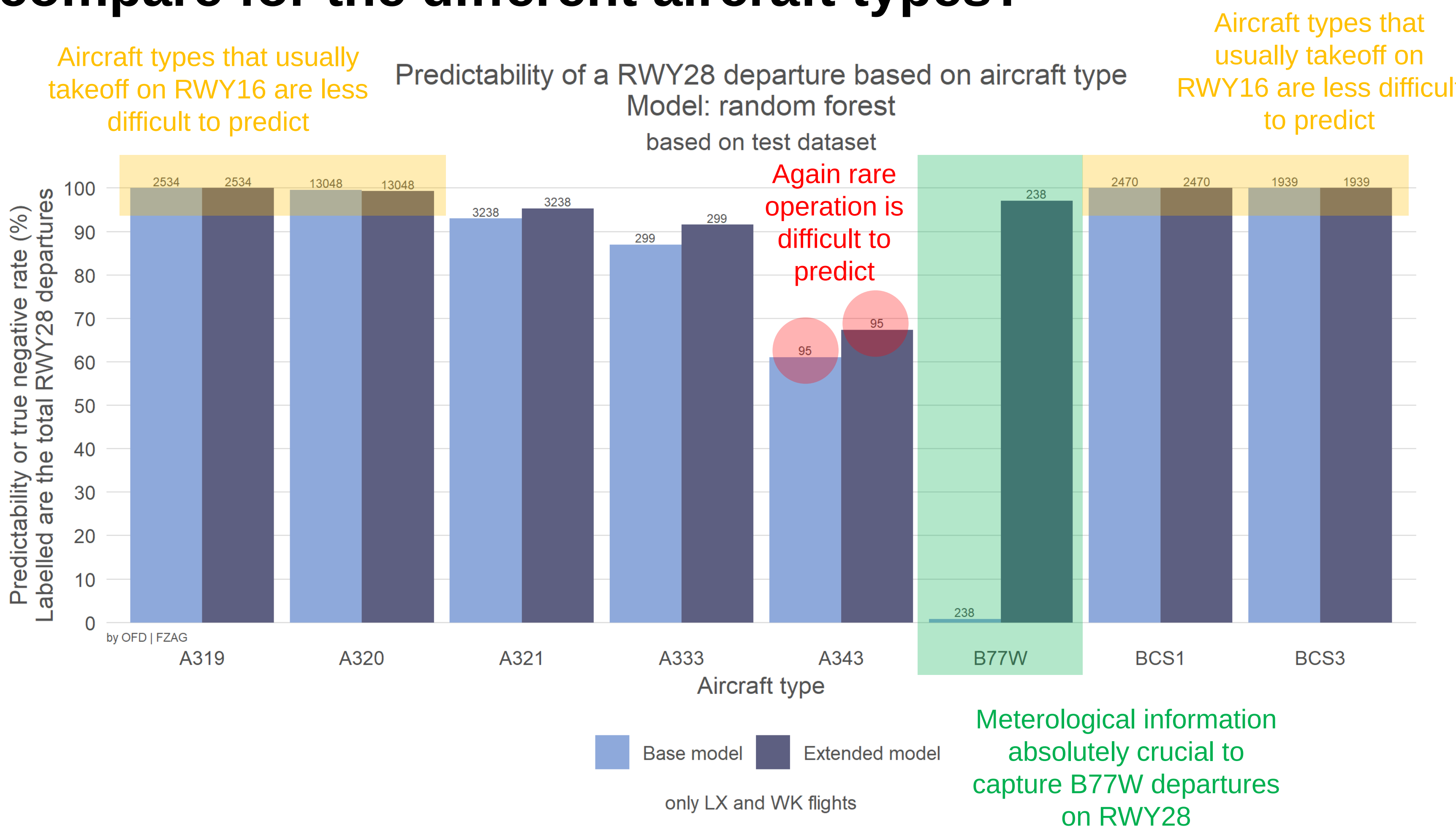
FLUGHAFEN ZÜRICH

How does the predictability for a RWY16 departure compare for the different aircraft types?



FLUGHAFEN ZÜRICH

How does the predictability for a RWY28 departure compare for the different aircraft types?



5th model
Extended random forest model

- Provides a very high accuracy
- Suitable to make very good day of operation prediction

- Takeoff weight and meterological information themself are predictions
- Can limit the accuracy of the model

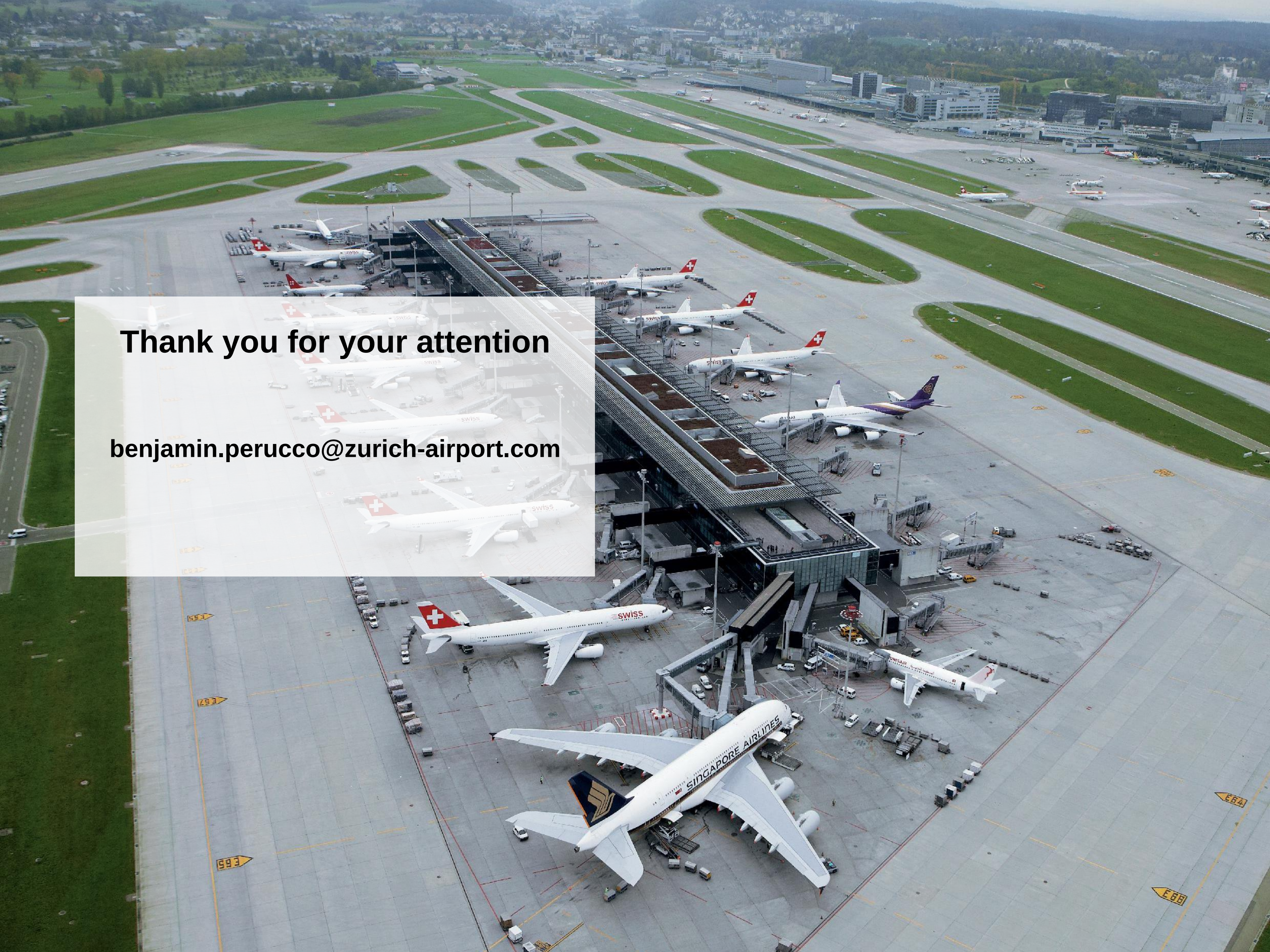
What's next?

- Testing other machine learning algorithms
- Study how the prediction of the takeoff weight and meterological information impacts the extended random forest model

- Demand on RWY in the next weeks or months?
  → Model based on historical data
- Demand on RWY in the next couple of hours?
  → Extended random forest model

- An exchange of information will become more crucial to apply the models in an operational environment

Questions?



Thank you for your attention

benjamin.perucco@zurich-airport.com