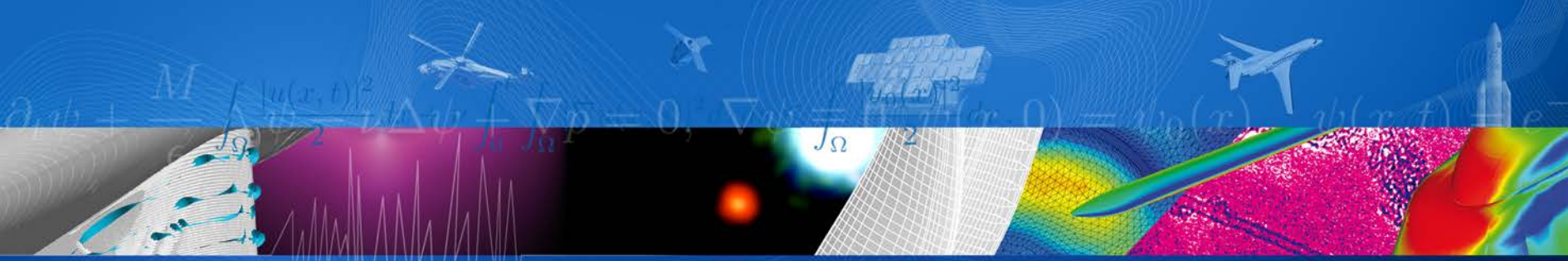


Occupancy Peak Estimation from Sector Geometry and Traffic Flow data

Luis Basora, Valentin Courchelle, Judicaël Bedouet, Thomas Dubot

+ Operational team in Bordeaux ACC

SIDs18

December 6th, Salzburg

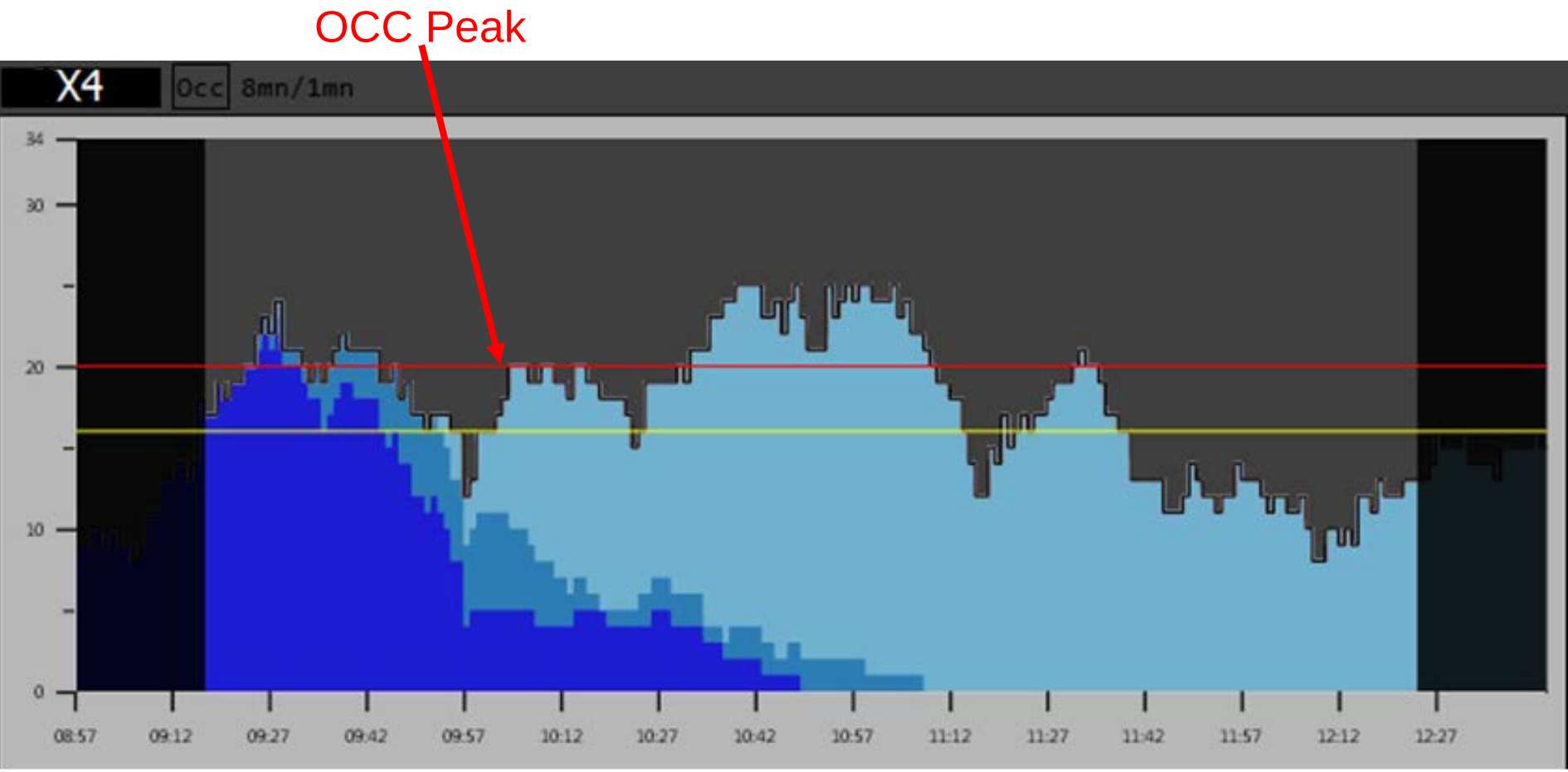
ONERA

THE FRENCH AEROSPACE LAB

Outline

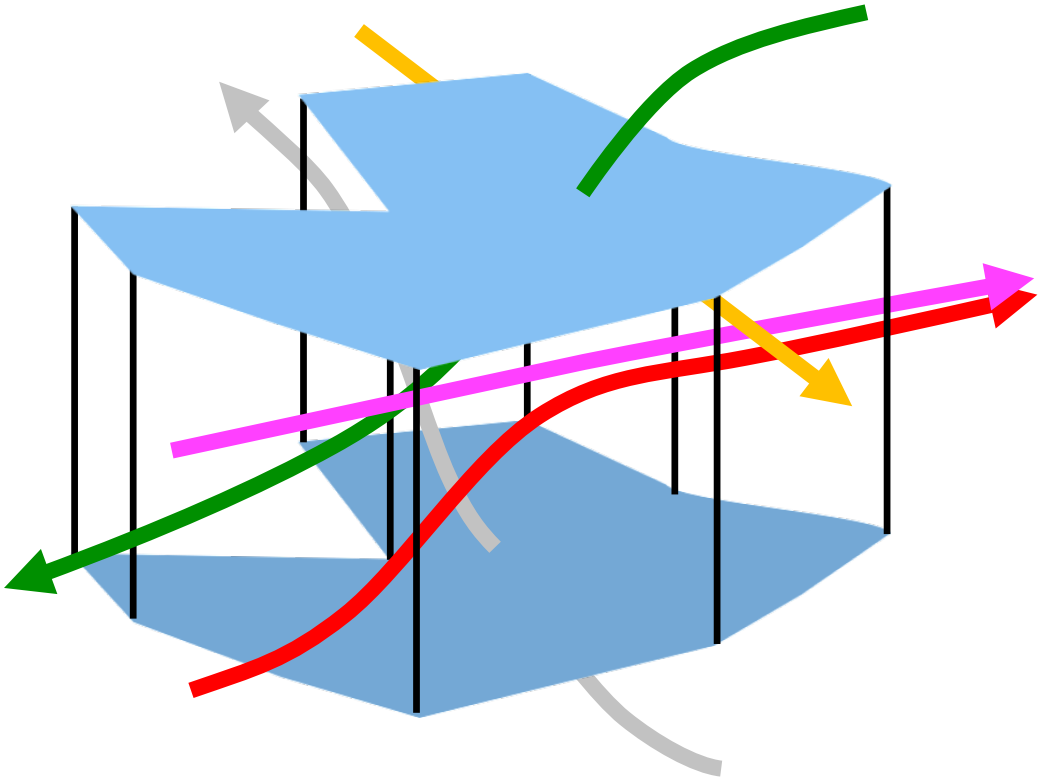
- **Introduction**
  - Occupancy count and Peak value
- **Methodology**
  - Scenario and datasets
  - Flow identification
  - Metrics generation
  - ML model for peak estimation
- **Preliminary results**
- **Conclusions**
  - Limitations, issues, future work

Occupancy count (OCC) metric example



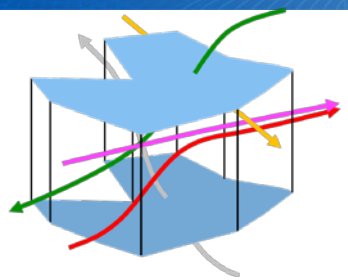
Occupancy = Number of flights inside a sector during a period of time

The « peak »: an empirical threshold



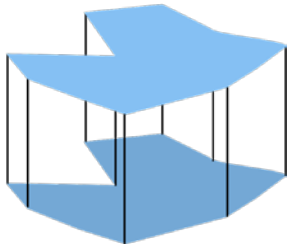
- Peak for a new designed sector
- Fine-tuning existing peak values

Methodology overview

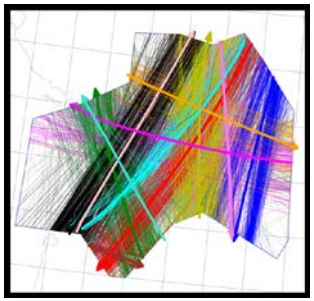


Geometry

Traffic complexity



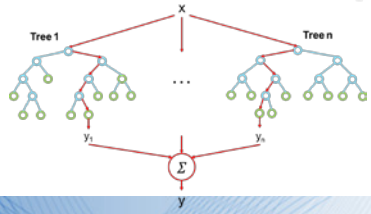
Metrics generation



Flow identification

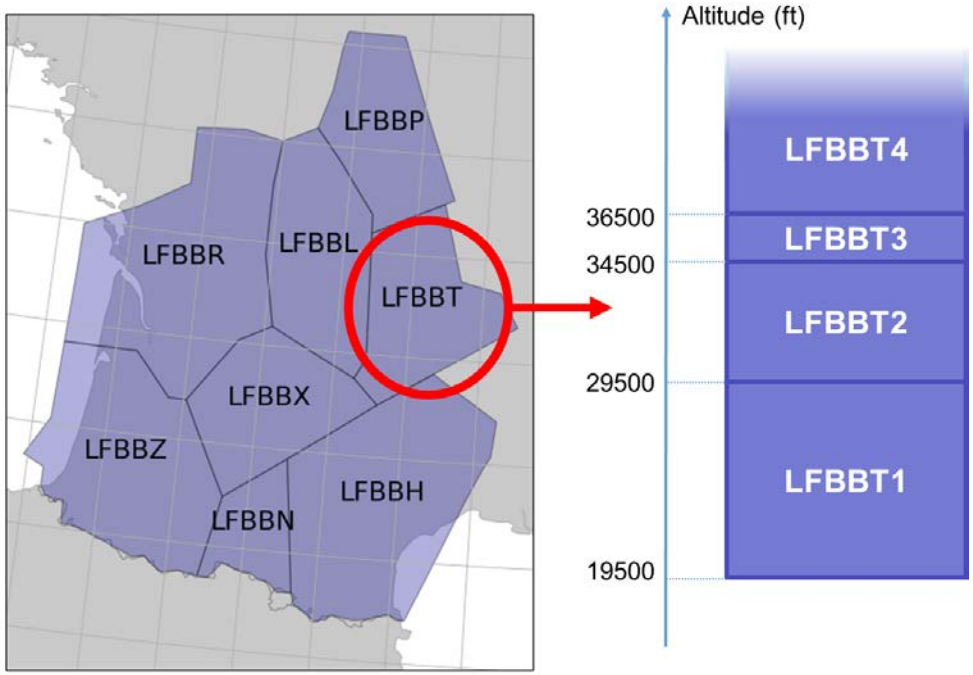
Sector	Metrics	OCC peak

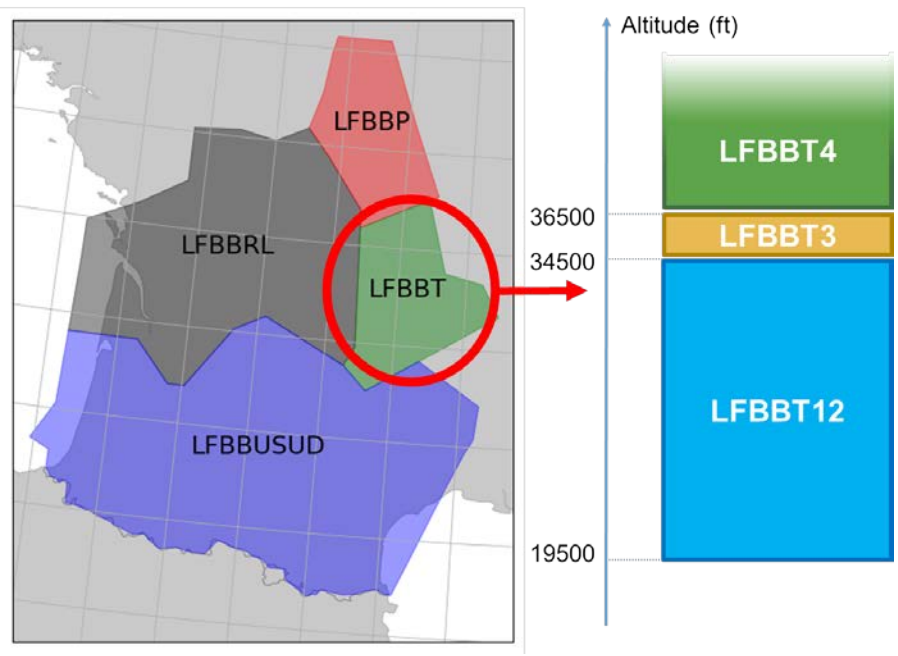
Model learning



Scenario: Bordeaux ACC (LFBB)

Sectorisation in 2017



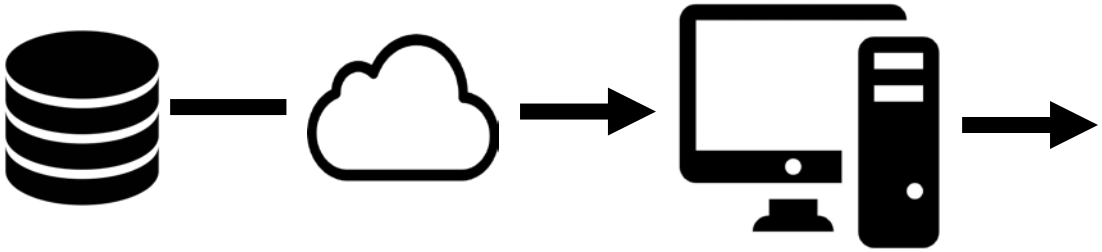


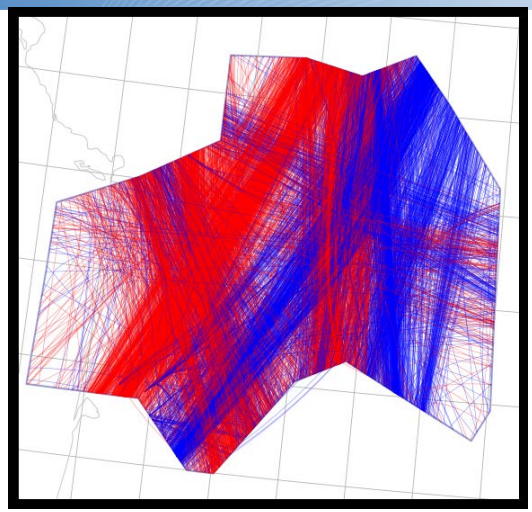
Dataset of 75 sectors in total

Traffic

OpenSky Network

Trajectory Preprocessing





2	00:00:00	04:19:00	259'	UBDX	FBDX									
2	04:19:00	04:22:00	3'	UBDX	FBDX									
3	04:22:00	04:38:00	16'	UNOR	FBDX	USUD								
4	04:38:00	04:41:00	3'	PT	RL	FBDX	USUD							
5	04:41:00	04:46:00	5'	PT	RL	FBDX	US12	US34						
6	04:46:00	04:47:00	1'	PT	RL12	RL34	FBDX	US12	US34					
7	04:47:00	05:00:00	13'	P14	T14	RL12	RL34	FBDX	US12	US34				
8	05:00:00	05:38:00	38'	P14	T14	RL12	RL4	RL3	FBDX	US12	US34			
9	05:38:00	09:24:00	226'	P14	T14	RL12	RL4	RL3	FBDX	US12	US3	US4		
10	09:24:00	10:20:00	56'	P14	T14	RL12	RL4	RL3	FBDX	US12	US3	ZX4	NH4	
9	10:20:00	11:01:00	41'	P14	T14	RL12	RL4	RL3	FBDX	US12	US3	US4		
10	11:01:00	11:27:00	26'	P34	T14	RL12	RL4	RL3	P12	FBDX	US12	US3	US4	
9	11:27:00	12:57:00	90'	P14	T14	RL12	RL4	RL3	FBDX	US12	US3	US4		
8	12:57:00	13:41:00	44'	P14	T14	RL12	RL34	FBDX	US12	US3	US4			
9	13:41:00	14:04:00	23'	P14	T14	RL2	RL1	RL34	FBDX	US12	US3	US4		
10	14:04:00	14:43:00	39'	P14	T14	RL2	RL1	RL4	RL3	FBDX	US12	US3	US4	
11	14:43:00	14:59:00	16'	P14	T14	RL2	RL1	RL4	RL3	FBDX	US1	US2	US3	US4
10	14:59:00	15:27:00	28'	P14	T14	RL12	RL4	RL3	FBDX	US1	US2	US3	US4	
11	15:27:00	16:32:00	65'	P12	T14	RL12	RL4	RL3	P34	FBDX	US1	US2	US3	US4

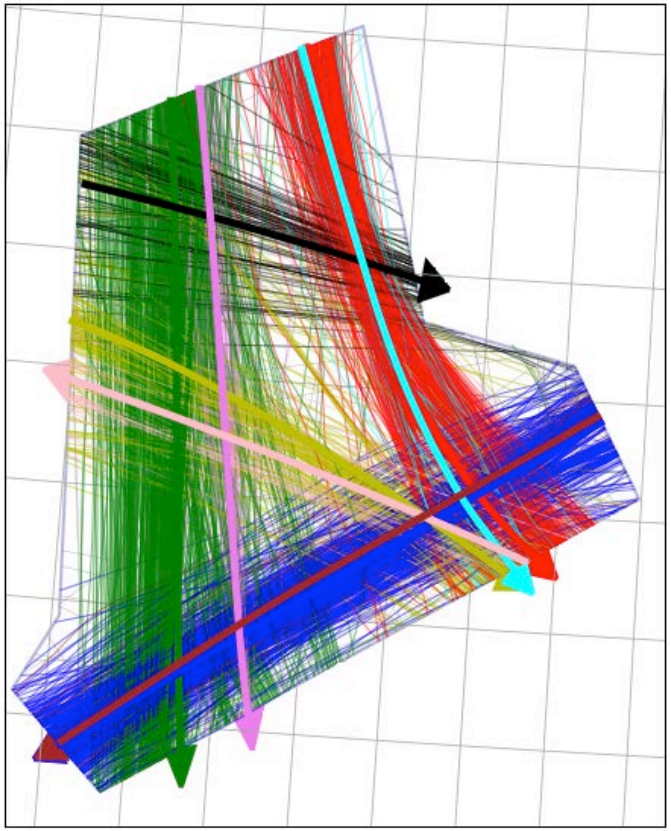
Sector Configuration Plan (SCP)

Traffic: July 13th - 19th 2017
(+ traffic from July and 1st semester for some sectors)

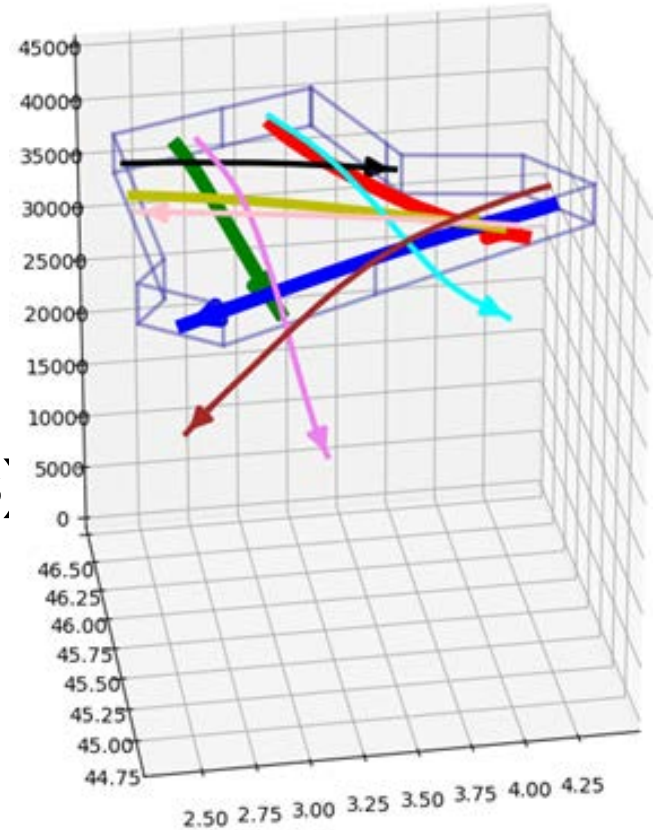
mean	2523
std	1422
min	481
25%	1354
50%	2201
75%	3678
max	4965

Flow identification

- Flow: trajectories with « similar » Entry/Exit points in the sector
- Trajectory clustering (DBSCAN, progressive clustering)
- Usability: only 6 parameters (no specific parameters per sector)
- Verification by visual inspection

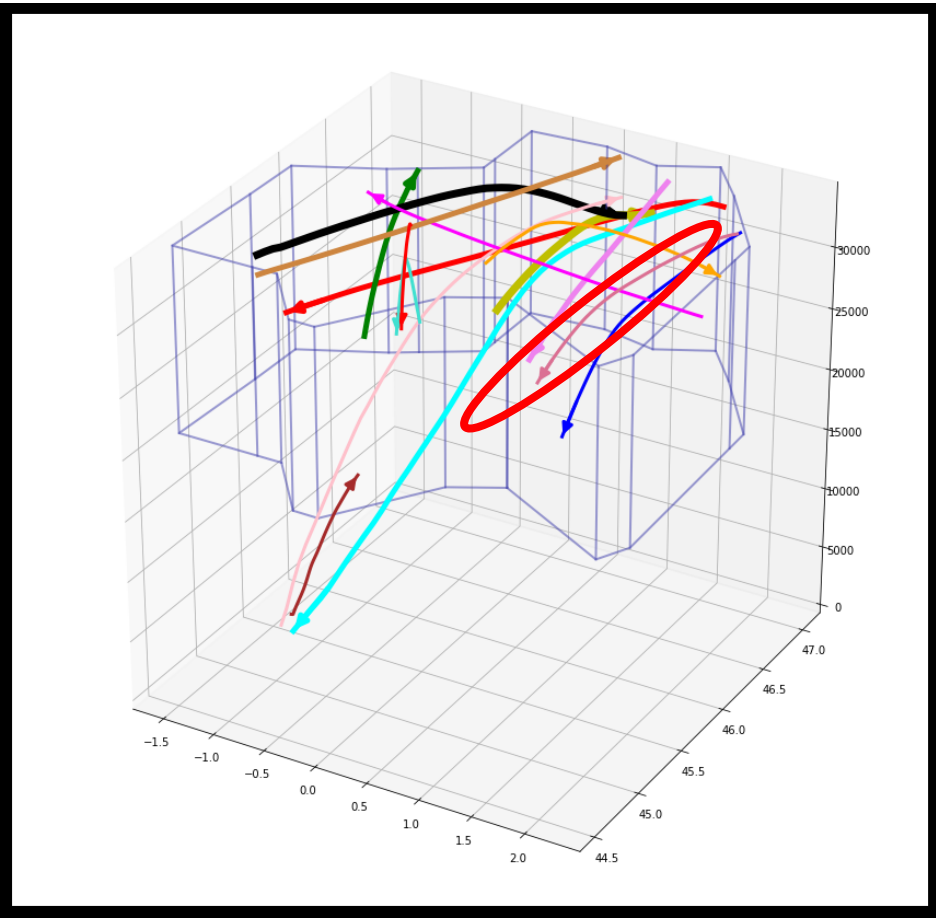


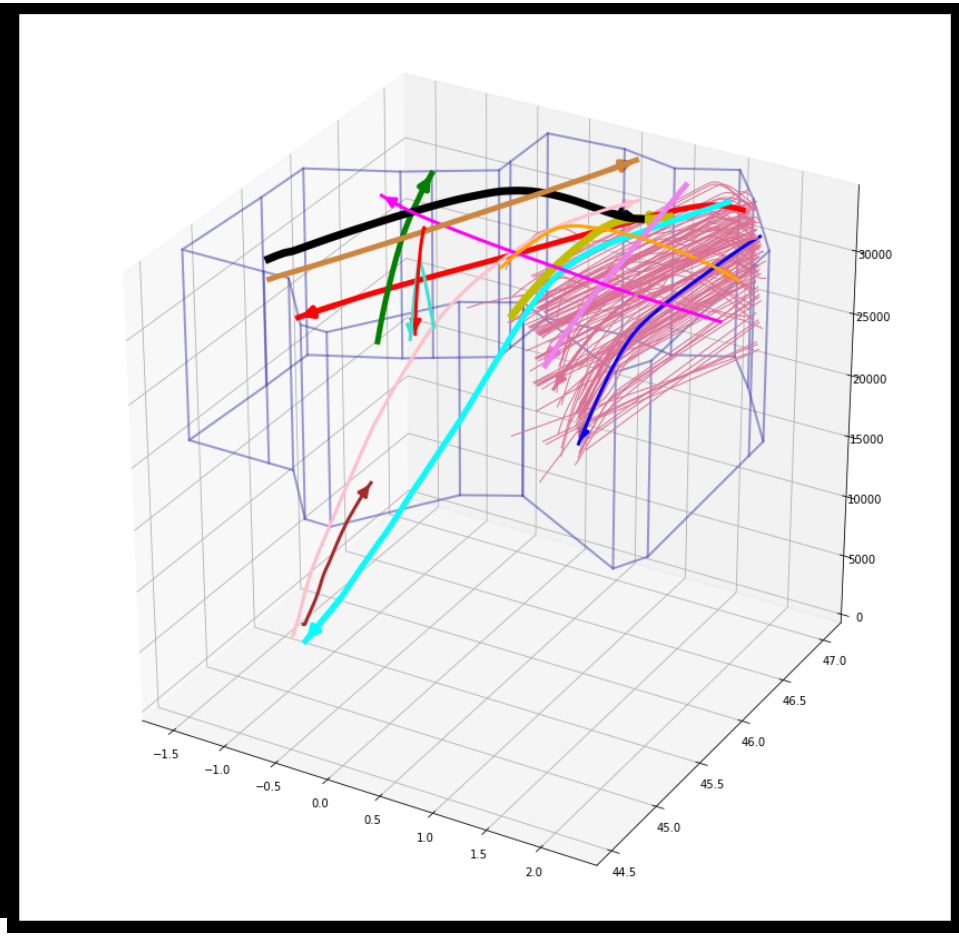
Sector T4
(9 flows,
6,4% outliers)



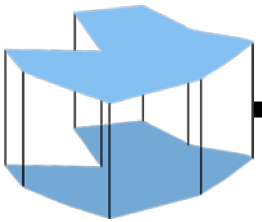
Operational validation of flows

Sector RL12 (15 flows, 13% outliers)





Metrics generation

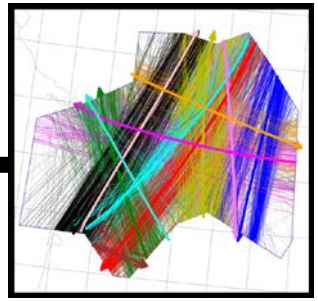


Geometry

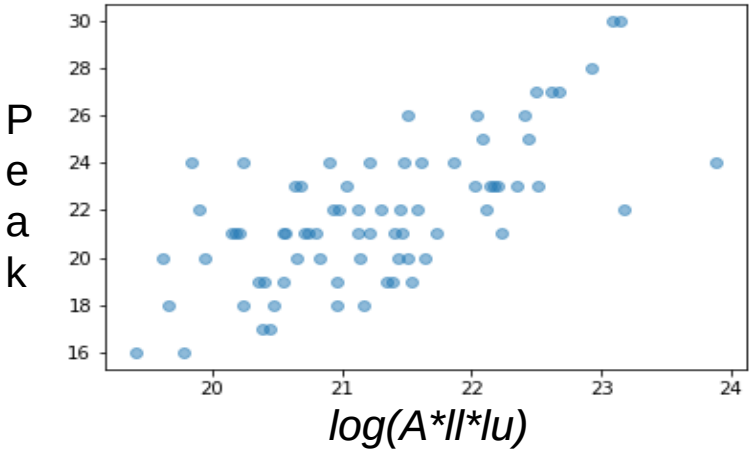
V	Sector volume (NM^3)
A	Sector area (NM^2)
H	Sector floor-to-ceiling heigh (ft)
l_l	Sector lower limit (sector floor altitude) (ft)
l_u	Sector upper limit (sector ceiling altitude) (ft)

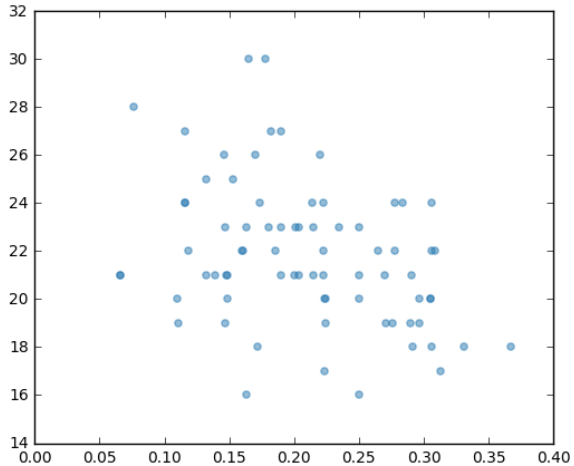
N_{flo}	Total number of flows
N_{stb}	Number of stable flows
N_{evol}	Number of evolving flows
N_{conv}	Number of converging flows
N_{cros}	Number of crossing flows

Traffic complexity



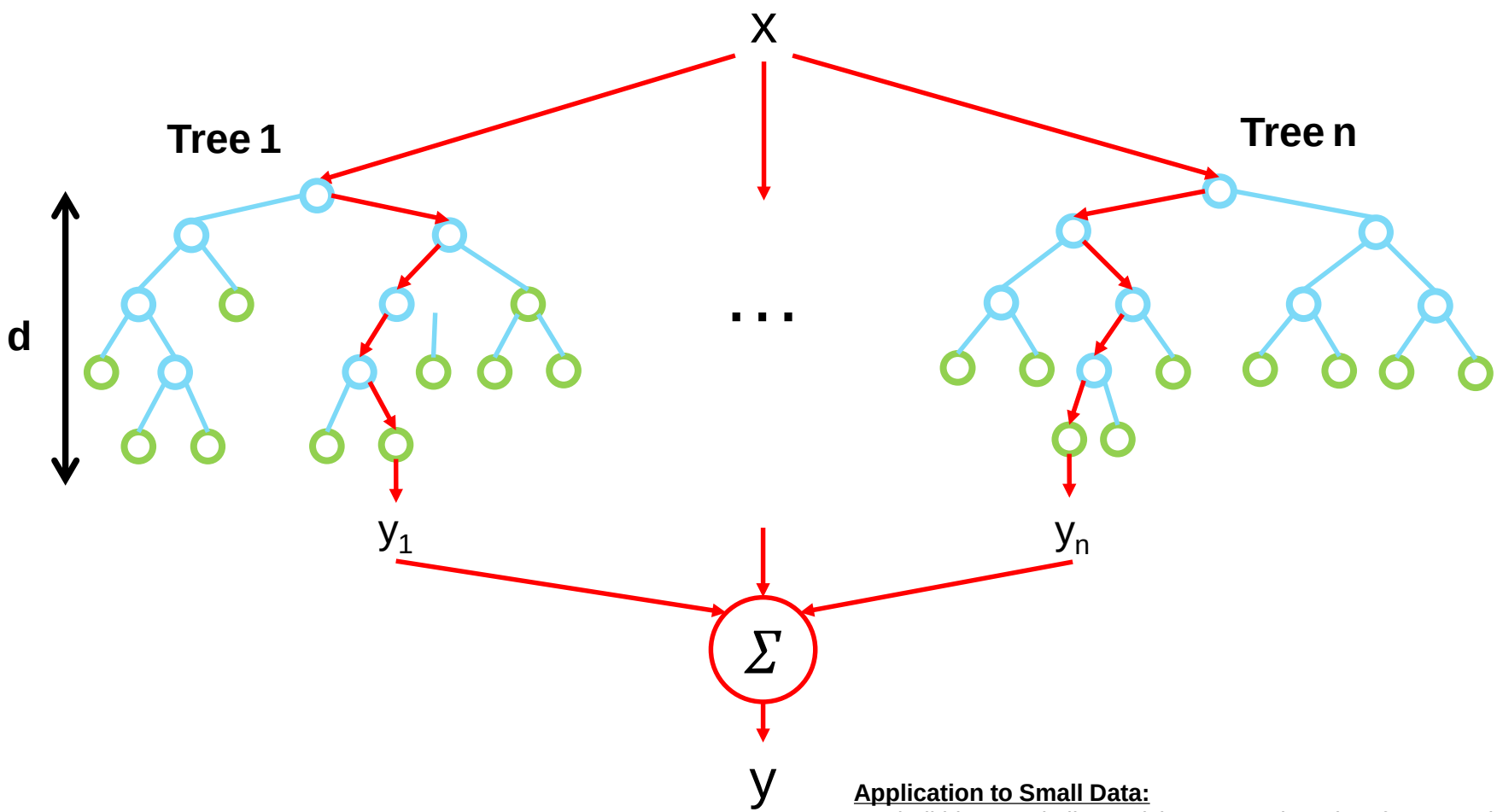
r: 0.28 (A) → 0.67 (log(A*ll*lu))





r: -0.16 (Ncros)→ -0.37 (Rcros)

Model selection: Random forest (RF)

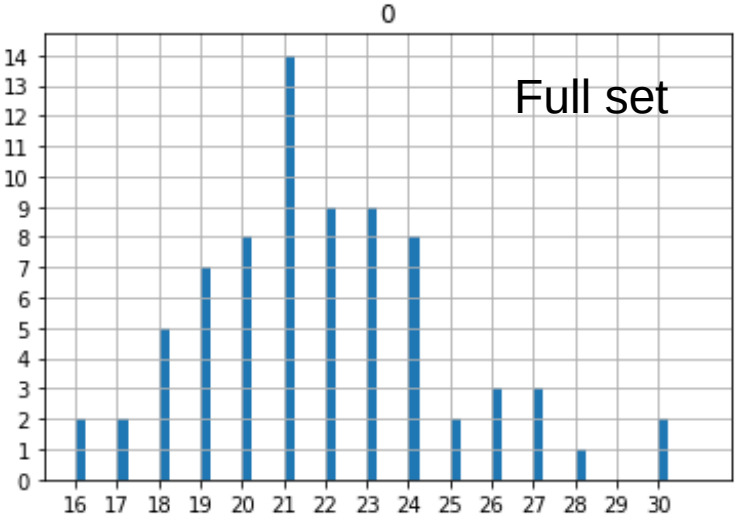


Application to Small Data:
T. Shaikhinaa and all, “Decision tree and random forest models for outcome prediction in antibody incompatible kidney transplantation,” *Biomedical Signal Processing and Control*, 2017.

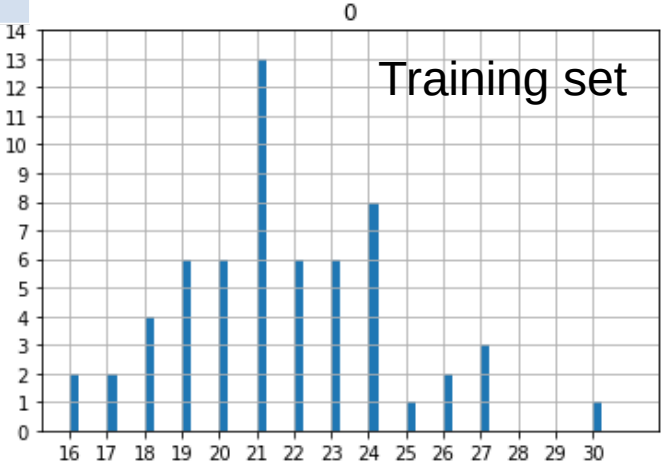
Training and test sets

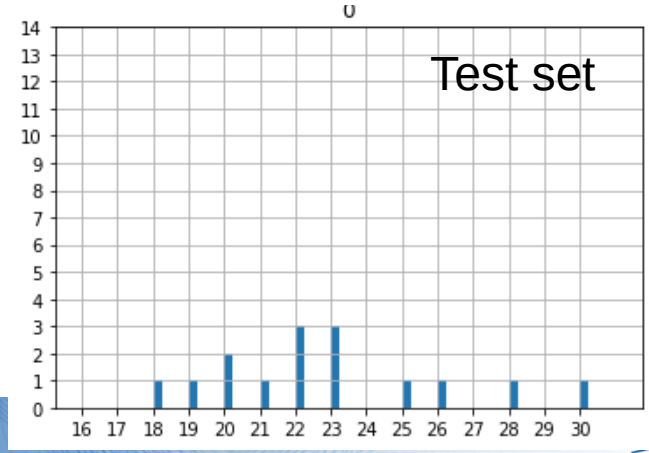
	Sector	Metrics	OCC peak (16-30)
1			
…			
75			

- Splitting: 80% training (60 sectors)
  20% test (15 sectors)
- Maximise diversity in training set (stratified sampling)



Peak distribution





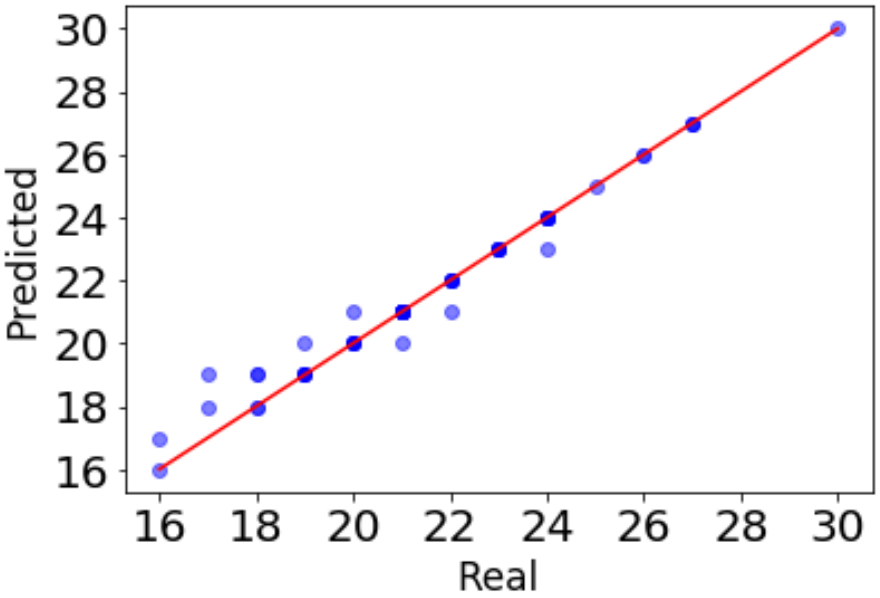
Model training and testing

- **RF hyperparameters (n, d) selection**
  - Grid search using 5-fold CV on the training set
  - Best model train with full training set
  - Test set only for final evaluation of best model
- **Explicit feature selection**
  - Strong correlation with peak
  - Keep as few features as possible

=> Final settings: n=100, d=7

Feature	Importance
$log(A * l_l * l_u)$	32.6%
l_u	23.6%
R_{conv}	12.7%
l_l	11.0%
R_{cros}	8.7%
N_{st}	6.6%
R_{evol}	4.8%

Preliminary Results: Training set

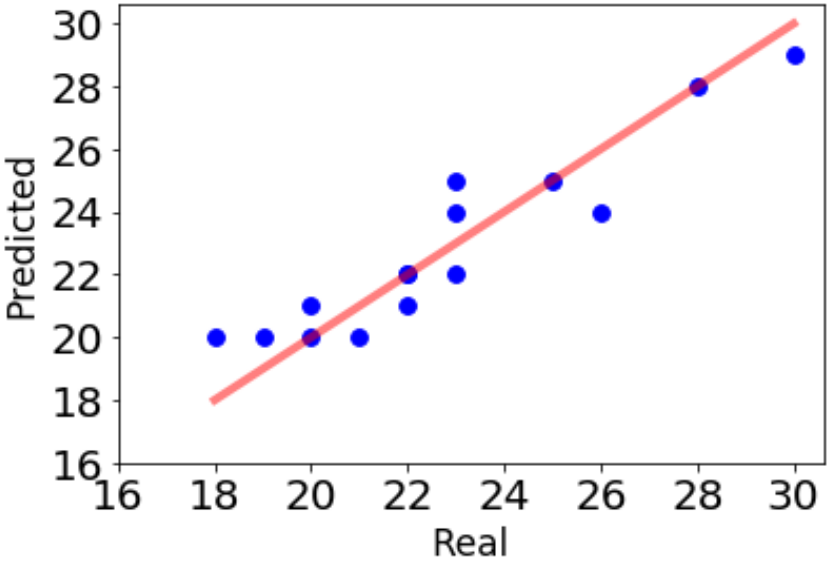


$R^2 = 0{,}979$

Sectors with errors:

Sector	Real peak	Prediction	Error
X2	18	19	1
X3	17	18	1
H2	22	21	-1
T12	24	23	-1
R3	18	19	1
L1	16	17	1
L12	20	21	1
L3	21	20	-1
L2	19	20	1
R1	17	19	2

Preliminary Results: Test set



$R^2 = 0,838$

Sector	Real peak	Prediction	Error
H3	20	20	0
ZX34	25	25	0
ZNH4	28	28	0
T123	22	22	0
FNOR	22	22	0
Z2	21	20	-1
NS34	23	22	-1
US4	30	29	-1
USUD	24	23	-1
R12	19	20	1
L4	20	21	1
RL12	22	21	-1
X1	18	20	2
NH34	23	25	2
RL	26	24	-2

Conclusions and future work

- Risk of overfitting: generalisation to new data?
- Better operational validation of traffic flows
- Optimisation of clustering parameters
- Traffic complexity metrics to be improved
- Applicability of the method to other ACCs?
- Learning from a multi-ACC dataset?
- Applicability to estimate other capacity metrics' thresholds (e.g. Hourly Rate for the Entry Count)?

Thank you!
Questions?