7th SESAR Innovation Days

Probabilistic Traffic Models for Occupancy Counting

J. Boucquey[1], F. Gonze[2], A. Hately[1], E. Huens[2], R. Irvine[1], S. Steurs[1], R.M. Jungers[2]

[1] EUROCONTROL ATM/RDS/ATS
[2] UCLouvain ICTEAM Institute

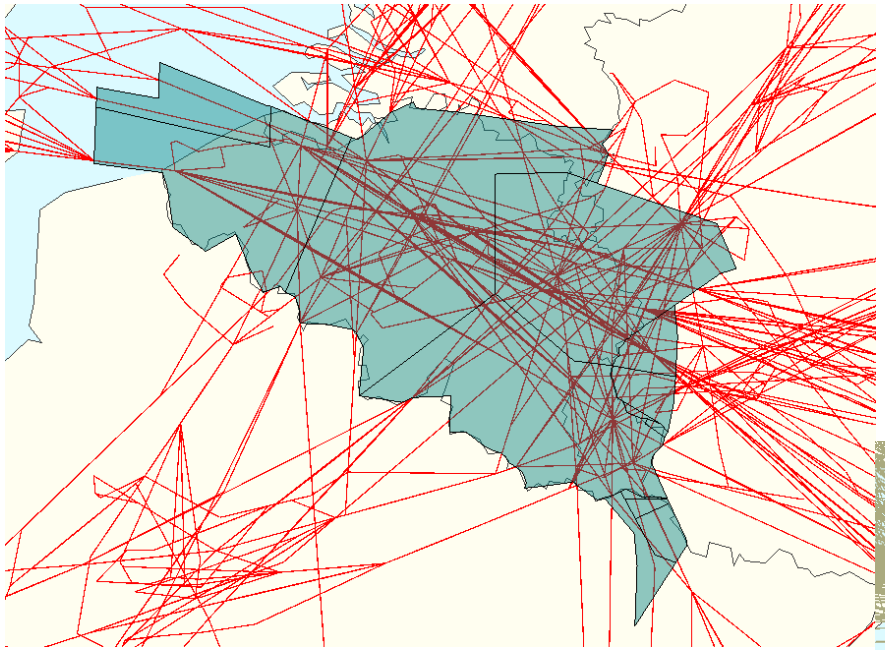
COPTRA

SESAR JOINT UNDERTAKING

Founding Members

EUROPEAN UNION EUROCONTROL

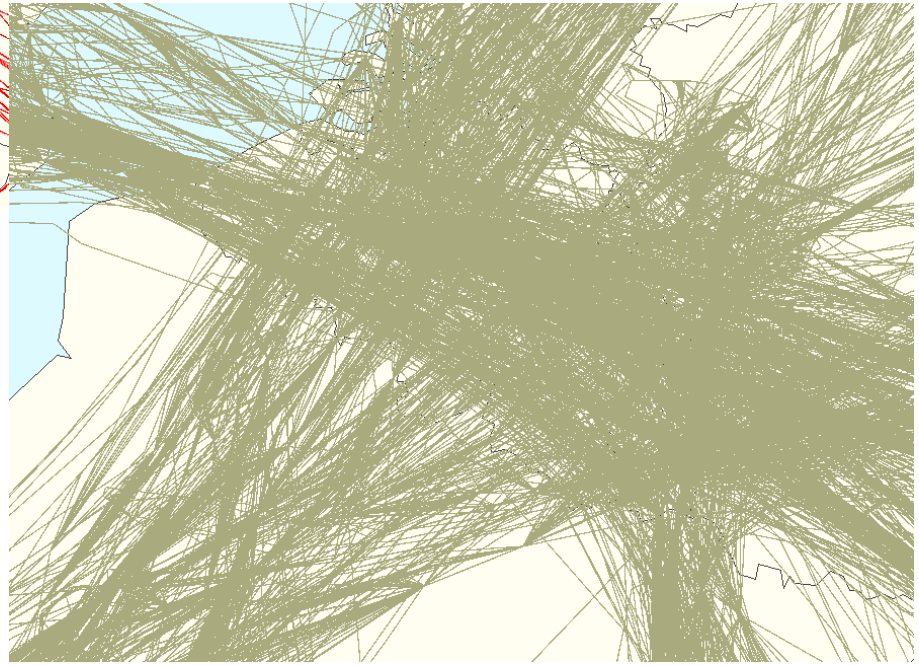
Traffic Uncertainty



Planned traffic

- T/O time?
- Directs?
- Conflicts
- Weather
- …

Actual

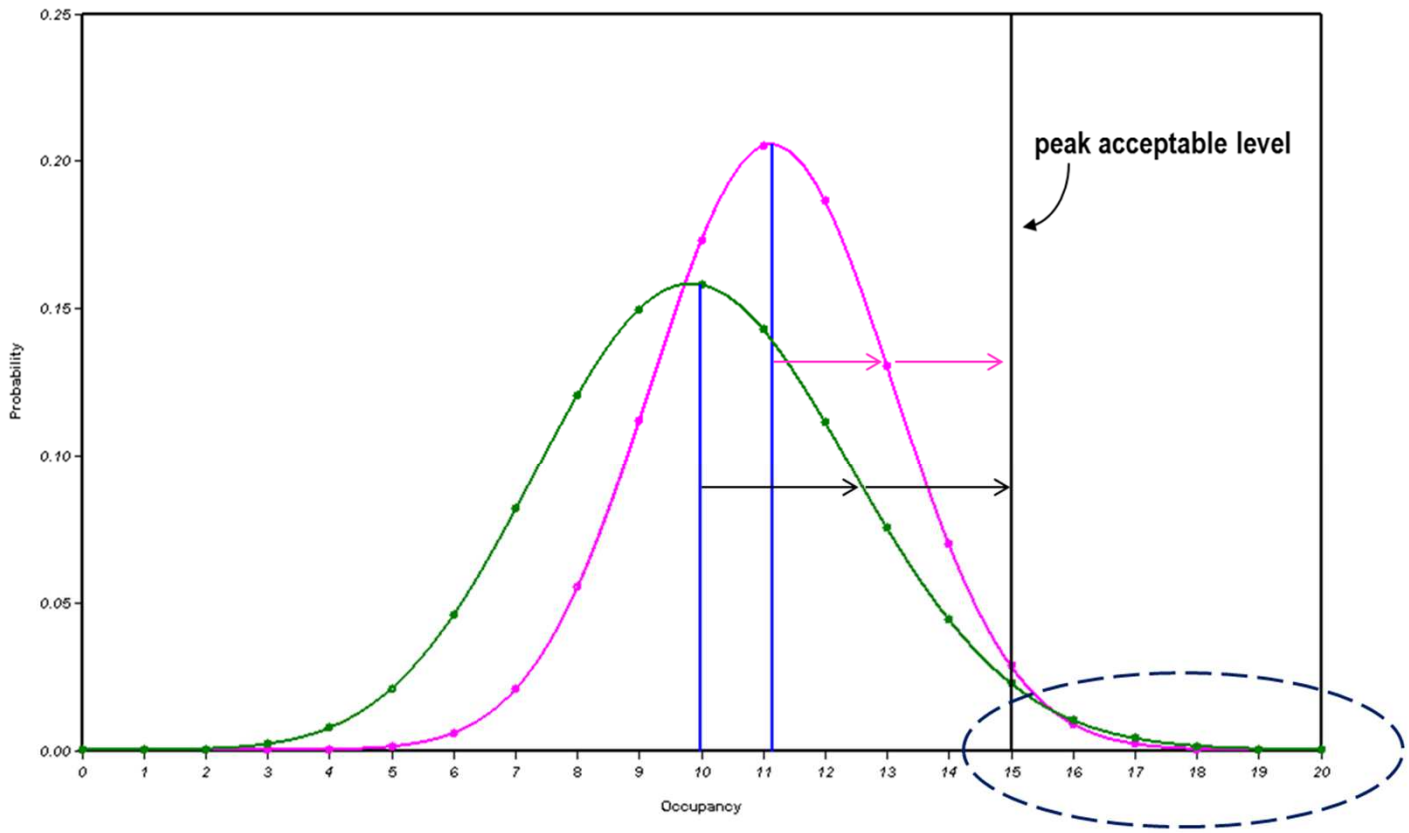


- Predictions based on planning info, route structure
- Do not materialize

Traffic Uncertainty

Impact on Capacity

"Sector capacity is set to control the probability of occupancy counts exceeding the peak acceptable level."



[2] R. Irvine, *Enhanced DCB Step 1 R4 Validation Report (VALR) - Part II EXE-13.02.03-VP723*, SESAR Joint Undertaking, 2011.

COPTRA

- COPTRA: **CO**mbining **P**robable **TRA**jectories

"COPTRA proposes an operational concept where the uncertainty of the predicted trajectories is made explicit at trajectory prediction level and combined using state of the art applied mathematics methods to build a probabilistic traffic situation."

"These probabilistic traffic situations will be used to improve the prediction of occupancy counts used in ATC Planning and convey better information to the human operator."

COPTRA

- COPTRA: **CO**mbining **P**robable **TRA**jectories

"COPTRA p ... *rtainty of the predicted* ... *ion level and combined* ... *ods to build a probabilisti*

Probabilistic Trajectory Model

"These prob ... *ve the prediction* ... *nvey better informatio*

Occupancy Count Distributions

$\Theta_{(s,t)}$

Probabilistic Trajectory Model

- Principle:
  - To each planned flight attach several probable trajectories (i)
  - Probable trajectory = sequence of probabilistic states (j)

- In practice:
  - Elementary sector sequences
  - State = Entry and exit times
  - Gaussians

Sector sequence probability

Sectors in sequence

$$T_f = \left(p^*_{(f,i)}, \left(\mathrm{es}_{(f,i,j)}, \mu^e_{(f,i,j)}, \sigma^e_{(f,i,j)}, \mu^l_{(f,i,j)}, \sigma^l_{(f,i,j)} \right)_j \right)_i$$

Entry time mean & standard deviation

Exit time mean & standard deviation

Occupancy Count Distributions

$$\Theta_{(s,t)} : \mathbf{N} \rightarrow [0,1] : k \rightarrow \Theta_{(s,t)}(k)$$

Probability having k flights
In sector s at time t.

- Convolution of the binomial distributions giving the probability of having each flight in s at t.
- By standard methods requires exponential computing cost
- [1] describes a polynomial time algorithm

[1] F. Gonze, A. Simonetto, E. Huens, J. Boucquey and R. Jungers, *Probabilistic Occupancy Counts and Flight Criticality Measures for ATM*, Twelfth USA/Europe Air Traffic Management Research and Development Seminar (ATM2017).

Problem at hand

To get the occupancy count distributions for time t at a look ahead time of l.

For each possible flight, we need

- A set of probabilistic sector sequences with their respective probabilities $p^*_{(f,i)}$

- For each sequences:
  - Entry time distribution (mean & standard deviation)
  - Exit time distribution (mean & standard deviation)

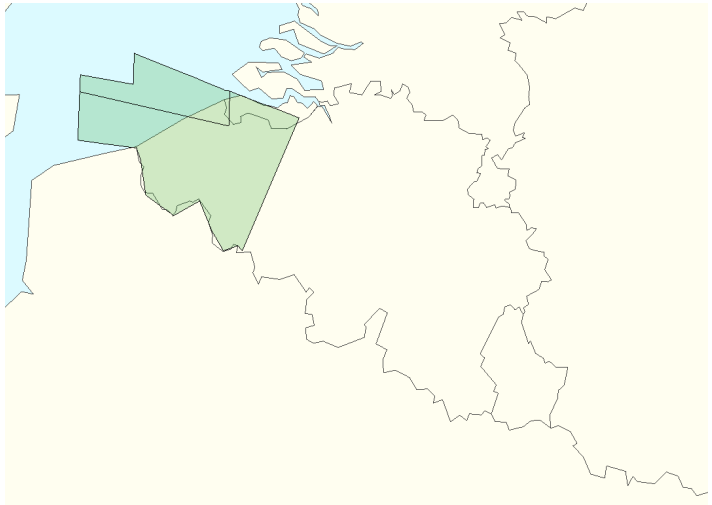
$$\mu^e_{(f,i,j)}, \sigma^e_{(f,i,j)}, \mu^l_{(f,i,j)}, \sigma^l_{(f,i,j)}$$

- How to determine this?

  Use of historical data

Dataset

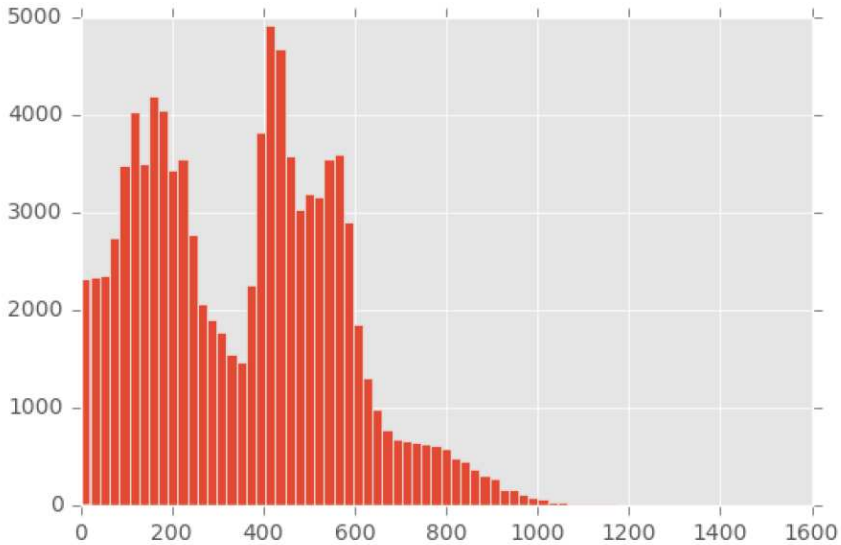
- AllFt+ data (from DDR)
- AIRAC 1607, 1608, 1609
- 1 323 866 crossings for 22 elementary sectors
- 91 389 crossings for EDYYB5KL



Extracted Features:

- Delta off-block time
- Delta entry time
- Sector crossing time

Data modelling



(EDYYB37EH crossing durations for AIRAC 1607, 1608, 1609)

- Multi-modal
- Non normal

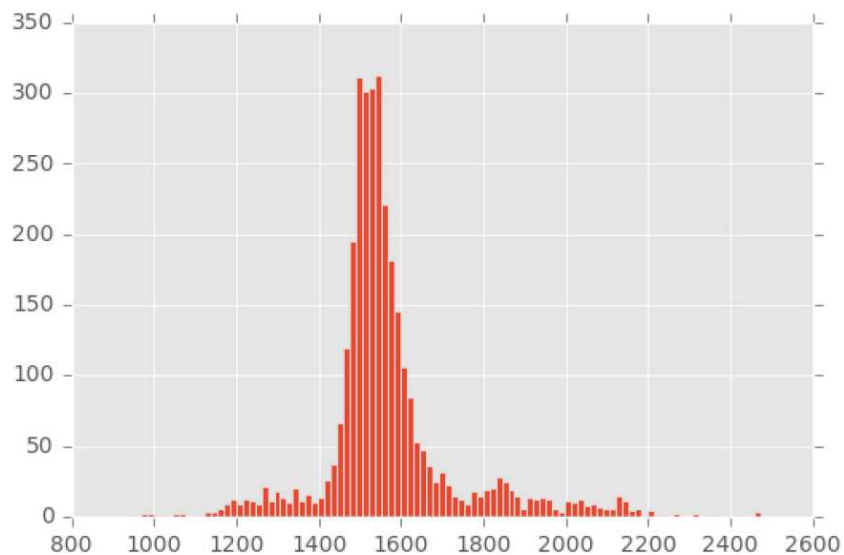
(Unconditioned distributions)

Gaussian Mixture Model:

$$\text{GMM} \sim \sum_{i=1}^{n} w_i N(\mu_i, \sigma_i)$$

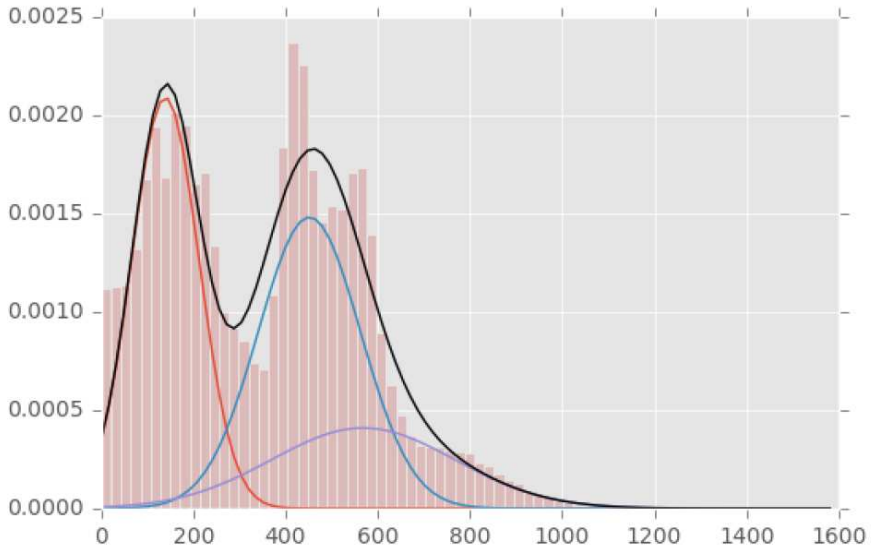
Fitting = unsupervised machine learning problem

- n as parameter
- Maximum Likelihood Estimation (MLE)



(eti from EHAM to EDYYB37EH for AIRAC 1607, 1608, 1609)

Data modelling



(MLE with 3 components)

- Multi-modal
- Non normal

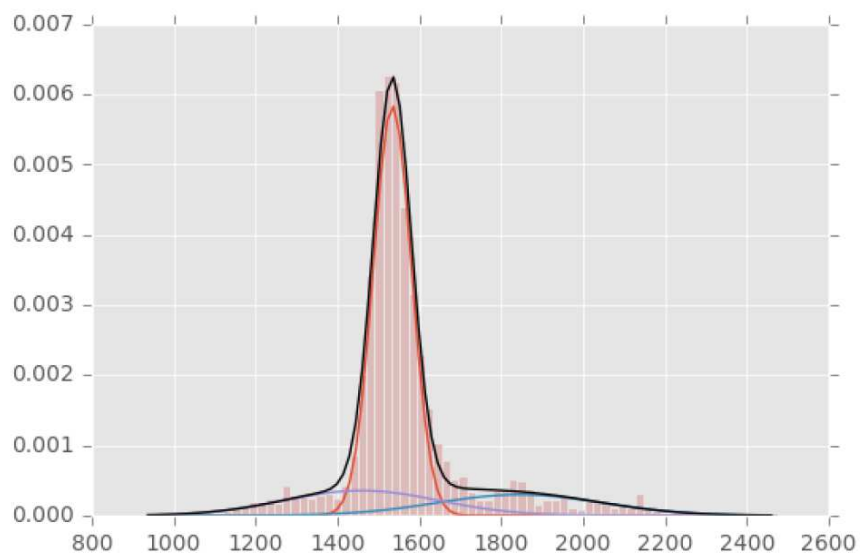
(Unconditioned distributions)

Gaussian Mixture Model:

$$\text{GMM} \sim \sum_{i=1}^{n} w_i N(\mu_i, \sigma_i)$$

Fitting = unsupervised machine learning problem

- n as parameter
- Maximum Likelihood Estimation (MLE)



(MLE with 3 components)

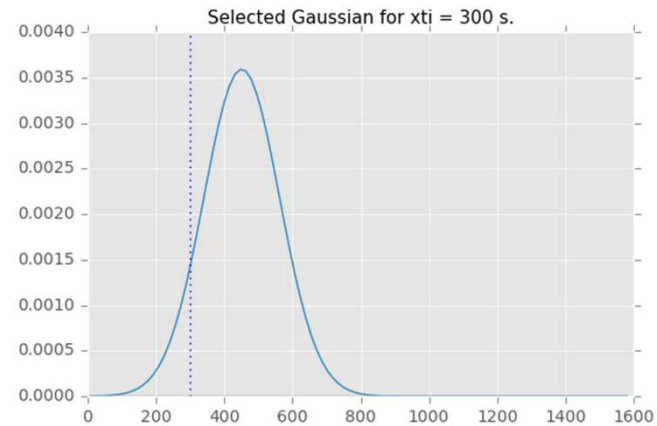
GMM Usage

Classifier or Predictor

- Classifier

$$\mathrm{GMM}_{\mathrm{class}}(t) \rightarrow i_{\mathrm{max}}$$

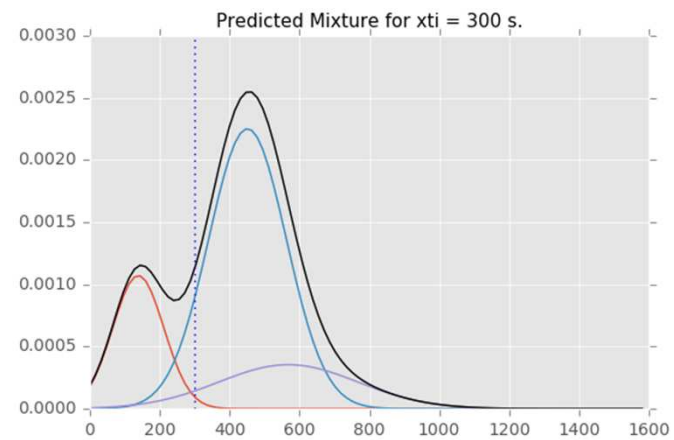
  - Gives the most probable Gaussian



- Predictor

$$\mathrm{GMM}_{\mathrm{pred}}(t) \rightarrow \left(p_k^t\right)_{k=1}^{n_{\mathrm{GMM}}}$$

  - Gives the probability of the respective Gaussians



Model Fitting

MUAC EDYYB5KL

- ADEP dependent models
  - **Off-Block delay model**
    - 11 Predictor GMM (based on ADEP) for the 11 most frequent ADEPs
    - 1 Classifier GMM (based on ICAO region) for the remaining ADEP
  - **Delta entry-time model**
    - 11 Predictor GMM (based on ADEP) for the 11 most frequent ADEPs
    - 1 Classifier GMM (based on ICAO region) for the remaining ones

Airport	%	n
EGGL	17.00	1
EGKK	12.19	3
EGSS	8.71	2
EGGW	4.50	2
LFPG	4.48	2
EDDF	2.71	2
EBBR	2.34	1
EGLC	2.30	2
EDDM	1.65	2
LEBL	1.58	1
EDDL	1.54	2

- **Crossing time model**
  - 1 Predictor GMM

i	w_i	μ_i (s)	σ_i (s)
1	.37	115.04	77.38
2	.31	330.67	159.85
3	.32	474.25	90.54

- In total, 25 GMM

Model Use

- An EZY flight from EGKK to EDMM,
  - Actual off-block time 05:48
  - Predicted to cross EDYYB5KL at 06:13:32 (DETI = 932 s) during 9 min and 15 sec (EGTI = 555 s)

DETI GMM for EGKK

i	w_i	μ_i (s)	σ_i (s)
1	0.35	2160.56	308.12
2	0.25	1453.21	319.70
3	0.40	2313.88	181.32

932 s →

Joint Probability Table

i	w_i			
1	0.01	~ 0.0	~ 0.0	~ 0.0
2	0.99	$\boxed{0.77}$	~ 0.0	0.23
3	~ 0	~ 0.0	~ 0.0	~ 0.0

XGTI GMM for EDYYB5KL

i	w_i	μ_i (s)	σ_i (s)
1	0.32	474.25	90.54
2	0.37	115.04	77.38
3	0.31	330.67	159.85

555 s →

i	1	2	3
w_i	0.77	~ 0	0.23

$p^*_{(f,i)}$	$\mu^e_{(f,i,j)}$	$\sigma^e_{(f,i,j)}$	$\mu^l_{(f,i,j)}$	$\sigma^l_{(f,i,j)}$
0.77	06:12:13	319.70	06:20:07	332.27

Compatible with Probabilistic Trajectory Model!

Model Use

Input/modelling dataset

- AllFt+ data
- AIRAC 1607, 1608, 1609
- 1 323 866 crossings for 22 elementary sectors
- 91 389 crossings for EDYYB5KL

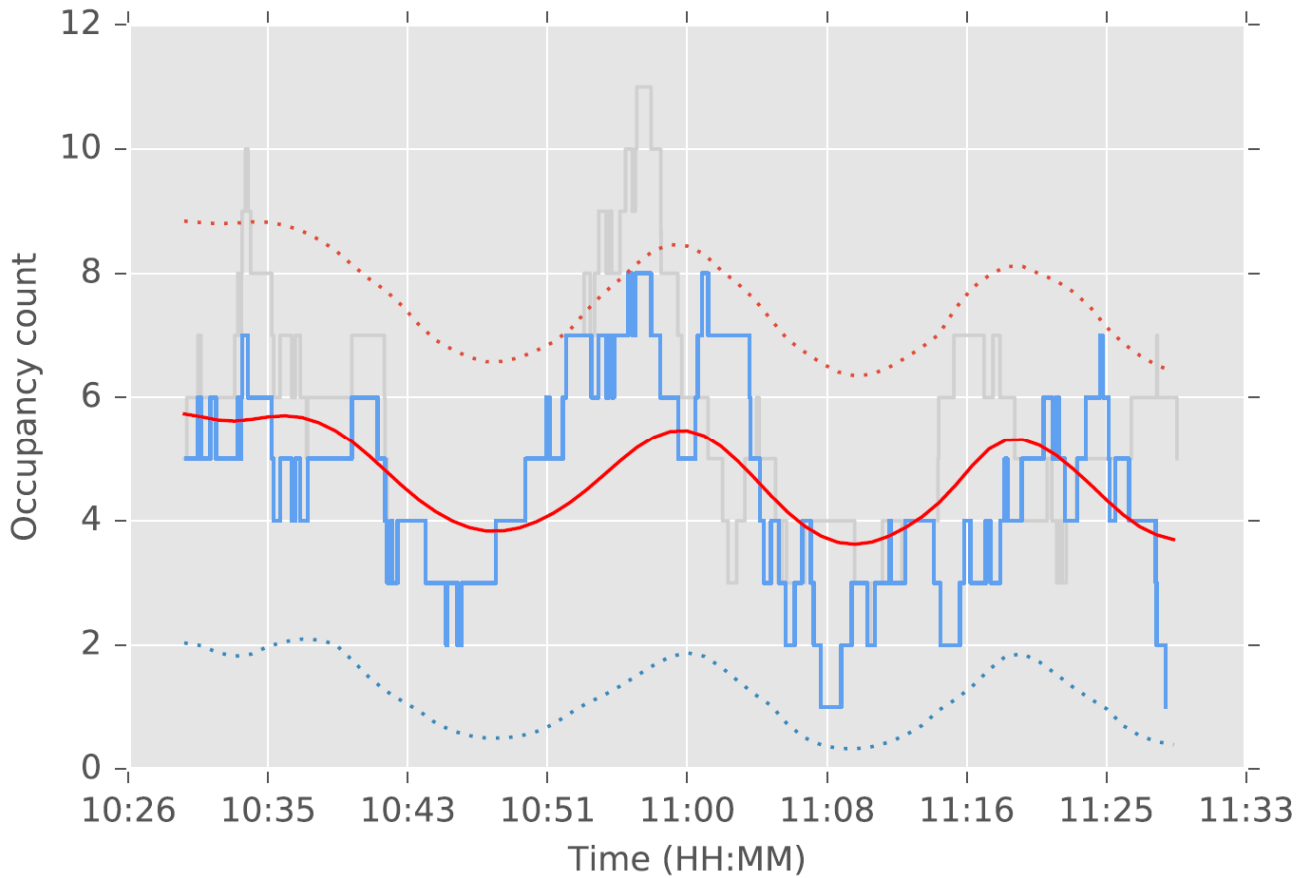
Target dataset

- 5th of May 2017
- ETFMS OPLOG for baseline
- 113 880 EFDs for 3413 unique flights

- AllFt+ for actuals
- 1131 flights

- Occupancy count distributions computed
  - at t every 30' from 05:00 to 23:00
  - For look-ahead time for $t - 5h$ to t (every 30')

Results



Occupancy count distributions (red and dashed) along actual (blue) and predicted (grey) occupancies.
EDYYB5KL – 5th of May 2017 – 11:00

Validation

- Validation approach
  - Baseline and probabilistic counts compared to actual counts (AllFt+)
  - Every 30’ from 05:00 to 23:00 predicted every 30’ from t -5h to t:
    - 37 target times
    - 11 look-ahead times
    - -> 407 (37 x 11) comparisons aggregated by look-ahead time

- Probabilistic and deterministic forecasts
  - Count distributions -> Probabilistic
  - Baseline counts -> Deterministic

- Ranked Probability Score:

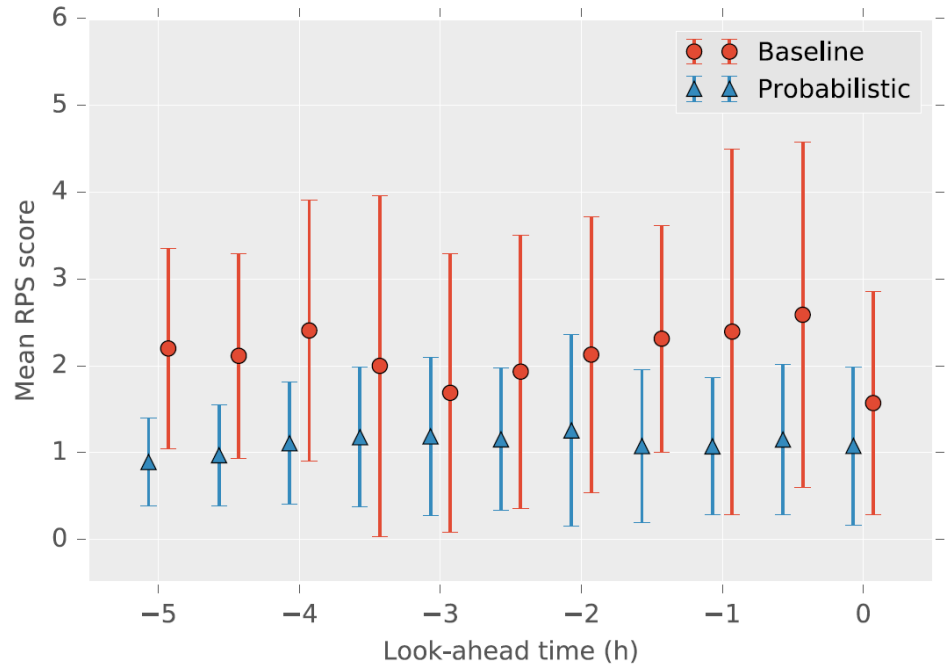
$$\mathrm{RPS}\left(F^{\Theta}_{(s,t,l)}, o_{(s,t)}\right) = \sum_{n=0}^{\infty}\left(F^{\Theta}_{(s,t,l)}(n) - H[n - o_{(s,t)}]\right)^2$$

  - Deterministic = Distribution with 1 value of probability 1
  - In deterministic case, RPS = Absolute Error

Validation

MEANS AND STANDARD DEVIATIONS OF THE BASELINE AND PROBABILISTIC SCORES

l	Baseline		Probabilistic	
(h)	mean	stdev	mean	stdev
−5.0	2.20	1.155	0.89	0.507
−4.5	2.12	1.177	0.97	0.579
−4.0	2.41	1.500	1.11	0.701
−3.5	2.00	1.963	1.18	0.804
−3.0	1.63	1.606	1.19	0.914
−2.5	1.93	1.574	1.15	0.821
−2.0	2.13	1.586	1.26	1.100
−1.5	2.31	1.306	1.08	0.878
−1.0	2.39	2.106	1.07	0.789
−0.5	2.59	1.987	1.15	0.861
0.0	1.57	1.290	1.08	0.910



- Standard deviations are significantly different:
  - Uncertainty reduction
- Means are significantly different (except @ t and $t - 3h$):
  - Better accuracy

Statistical significance level: 5%

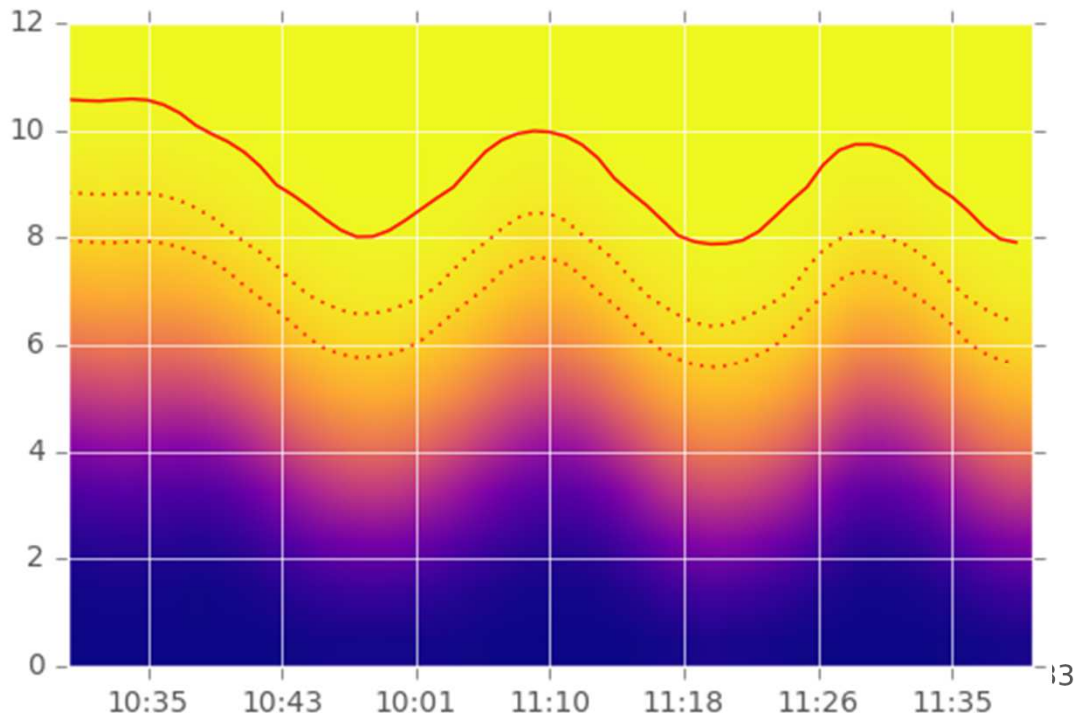
Conclusions

- Flexible and extensible approach based on historical data to attach uncertainty to traffic demand
- Based on Gaussian Mixture Models (GMM)
- Compatible with the “COPTRA probabilistic trajectory model”
    - Occupancy count distributions can be computed in polynomial time
- Brings:
    - Reduced uncertainty
    - Improved accuracy

On going work

- "Hotspot" prediction
  - Probability to exceed a given capacity

- Visualization
  - How to convey uncertainty to the human operator?



Questions

?

Thank you very much for your attention!

This project has received funding from the SESAR Joint Undertaking under the European Union's Horizon 2020 research and innovation programme under grant agreement No 699274

SESAR JOINT UNDERTAKING

Founding Members

EUROPEAN UNION EUROCONTROL