Probabilistic Traffic Models for Occupancy Counting

J. Boucquey\textsuperscript{1}, F. Gonze\textsuperscript{2}, A. Hately\textsuperscript{1}, E. Huens\textsuperscript{2}, R. Irvine\textsuperscript{1}, S. Steurs\textsuperscript{1}, R.M. Jungers\textsuperscript{2}

\textsuperscript{1} EUROCONTROL ATM/RDS/ATS
\textsuperscript{2} UCLouvain ICTEAM Institute
Traffic Uncertainty

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

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.”

COPTRA

• COPTRA: COmbining Probable TRAjectories

“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 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.
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

\[
T_f = \left( p^*_i(f, i), \left( es(f, i, j), \mu^e_{(f, i, j)}, \sigma^e_{(f, i, j)}, \mu^l_{(f, i, j)}, \sigma^l_{(f, i, j)} \right)_j \right)_i
\]
Occupancy Count Distributions

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

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

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) $\mu_e^{(f,i,j)}, \sigma_e^{(f,i,j)}$
  - Exit time distribution (mean & standard deviation) $\mu_l^{(f,i,j)}, \sigma_l^{(f,i,j)}$
- 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

- 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

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)

- Multi-modal
- Non normal

(Unconditioned distributions)
GMM Usage

Classifier or Predictor

- **Classifier**
  
  \[ \text{GMM}_{\text{class}}(t) \rightarrow i_{\text{max}} \]

  - Gives the most probable Gaussian

- **Predictor**

  \[ \text{GMM}_{\text{pred}}(t) \rightarrow (p_k^t)_{k=1}^{n_{\text{GMM}}} \]

  - 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

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

- In total, 25 GMM

<table>
<thead>
<tr>
<th>Airport</th>
<th>%</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGGL</td>
<td>17.00</td>
<td>1</td>
</tr>
<tr>
<td>EGKK</td>
<td>12.19</td>
<td>3</td>
</tr>
<tr>
<td>EGSS</td>
<td>8.71</td>
<td>2</td>
</tr>
<tr>
<td>EGGW</td>
<td>4.50</td>
<td>2</td>
</tr>
<tr>
<td>LFGP</td>
<td>4.48</td>
<td>2</td>
</tr>
<tr>
<td>EDDF</td>
<td>2.71</td>
<td>2</td>
</tr>
<tr>
<td>EBBR</td>
<td>2.34</td>
<td>1</td>
</tr>
<tr>
<td>EGLC</td>
<td>2.30</td>
<td>2</td>
</tr>
<tr>
<td>EDDM</td>
<td>1.65</td>
<td>2</td>
</tr>
<tr>
<td>LEBL</td>
<td>1.58</td>
<td>1</td>
</tr>
<tr>
<td>EDDL</td>
<td>1.54</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>i</th>
<th>( w_i )</th>
<th>( \mu_i ) (s)</th>
<th>( \sigma_i ) (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.37</td>
<td>115.04</td>
<td>77.38</td>
</tr>
<tr>
<td>2</td>
<td>.31</td>
<td>330.67</td>
<td>159.85</td>
</tr>
<tr>
<td>3</td>
<td>.32</td>
<td>474.25</td>
<td>90.54</td>
</tr>
</tbody>
</table>
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)

<table>
<thead>
<tr>
<th>DETI GMM for EGKK</th>
<th>Joint Probability Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i$</td>
<td>$w_i$</td>
</tr>
<tr>
<td>1</td>
<td>0.35</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
</tr>
<tr>
<td>3</td>
<td>0.40</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>XGTI GMM for EDYYB5KL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$i$</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>

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 3,413 unique flights

- AllFt+ for actuals
- 1,131 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’

7th SESAR Innovation Days - Belgrade
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:
  \[ \text{RPS} \left( F_{(s,t,l)}, o_{(s,t)} \right) = \sum_{n=0}^{\infty} \left( F_{(s,t,l)}(n) - H[n - o_{(s,t)}] \right)^2 \]
  • 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**

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

- 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!