Data-Driven Occupancy Prediction in Adverse Weather Conditions using Thunderstorm and Traffic Observations

Aniel Jardines
PhD Advisor: Manuel Soler
ATFM Delays due to Weather

• Weather delays have increased over the last 5 years
• Weather is the 2nd cause of ATFM delay after ATC Capacity.
• Cost due to weather delay in 2018 estimated at 0.48 Billion Euros
2018 Weather Regulations

- Majority ATFM weather regulations occur in the summer months (May, Jun, Jul, Aug).
- ATFM weather regulations are concentrated in a few severe days. (14 days responsible for over 1/5 of total weather regulations in 2018.)
- Convective weather (Thunderstorms) responsible for majority of weather regulations.
Research Questions

• Can we reduce delay/number of regulations due to weather?
• Can we improve ATFM operations during convective weather events?
• Can we improve weather predictability?
• Can we make better use of weather information during the pre-tactical phase?
Methodology

• Data Science
• Develop models to predict and quantify capacity and demand imbalances due to weather in the airspace network.
• Analysis of historical data:
  • Traffic
  • Weather Observations
  • Weather Forecast
Approach

Step 1:
If we had access to a “perfect forecast” how would be implement ATFM solutions?
• Could we predict values of sector capacity and demand values?
• Could we improve ATFM operations (manage traffic flows, ATFM Regulations, etc)?

Analysis of historical traffic data and weather observations. Find correlations between storms and traffic patterns.

Step 2:
How can we still implement ATFM solutions with a probabilistic forecast?
• At what time horizon can we expect a reliable forecast, how reliable does my forecast need to be?
• How much lead time is necessary to make a significant impact on ATFM plan?

Analysis of historical weather observations and weather forecasts. Leverage Numerical Weather Prediction to anticipate storms.
Thunderstorms

• Three ingredients:
  • Lifting force (heat)
  • Moisture
  • Unstable atmosphere (temp. lapse rate)
• Shelf cloud is formed at the condensation layer
• Top cloud spreads as updrafts reach the tropopause
• Overshoots can occur due to strong updrafts
Weather Data

• Weather observations data provided by EUMETSAT’s Rapidly Developing Thunderstorm (RDT) product.
• Satellite images of convective cells available every 15 minutes
• Parameters provided for each cell: Top/Shelf Cloud Contour, Overshoots Location, Severity, Cloud Top, Phase Life, Velocity, Direction, etc.
Traffic Data

- Historical flight trajectories from DDR.
- 4D Trajectory with non-uniform points
- Current Tactical Flight Model trajectory provides closest representation of actual flight.
- Unique IFPS ID of each trajectory allows us to identify and cross-reference each trajectory.
Spatio-temporal Grid

- Necessary to integrate traffic and weather data by mapping onto a common domain.
- Defined grid using latitude, longitude, altitude and time

<table>
<thead>
<tr>
<th>Altitude Level</th>
<th>Pressure Level</th>
<th>Flight Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>&lt;= 200</td>
<td>&gt;= 390</td>
</tr>
<tr>
<td>4</td>
<td>200-250</td>
<td>390-340</td>
</tr>
<tr>
<td>3</td>
<td>250-300</td>
<td>340-300</td>
</tr>
<tr>
<td>2</td>
<td>300-400</td>
<td>300-240</td>
</tr>
<tr>
<td>1</td>
<td>400-700</td>
<td>240-100</td>
</tr>
<tr>
<td>0</td>
<td>700-Surface</td>
<td>100-Surface</td>
</tr>
</tbody>
</table>
### Processing Data - Weather

<table>
<thead>
<tr>
<th>Time</th>
<th>Lon</th>
<th>Lat</th>
<th>Alt Level</th>
<th>Overshoot</th>
<th>Storm Cell</th>
<th>Shelf Cloud</th>
<th>Not Defined</th>
<th>Low</th>
<th>Moderate</th>
<th>High</th>
<th>Very High</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018-05-29 16:45:00</td>
<td>10.6</td>
<td>49.9</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2018-05-29 13:15:00</td>
<td>1.9</td>
<td>41.7</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2018-05-29 11:00:00</td>
<td>13.1</td>
<td>46.5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2018-05-29 22:15:00</td>
<td>17.2</td>
<td>45.7</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2018-05-29 13:00:00</td>
<td>4.2</td>
<td>51.6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Processing Data - Traffic

<table>
<thead>
<tr>
<th>Time</th>
<th>Lon</th>
<th>Lat</th>
<th>AltLevel</th>
<th>FlightIDs</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018-06-27 08:00:00</td>
<td>10.0</td>
<td>44.0</td>
<td>5</td>
<td>[AA02918295]</td>
<td>1.0</td>
</tr>
<tr>
<td>2018-05-18 09:00:00</td>
<td>2.0</td>
<td>46.0</td>
<td>1</td>
<td>[AA01469354, AA01470024, AA01474501, AA01470387]</td>
<td>4.0</td>
</tr>
<tr>
<td>2018-05-25 16:00:00</td>
<td>-6.0</td>
<td>51.0</td>
<td>3</td>
<td>[AA01704145, AA01736414]</td>
<td>2.0</td>
</tr>
<tr>
<td>2018-06-19 17:00:00</td>
<td>8.0</td>
<td>42.0</td>
<td>1</td>
<td>[AA02634936, AA02626809, AA02623041, AA02623042]</td>
<td>4.0</td>
</tr>
<tr>
<td>2018-06-22 10:00:00</td>
<td>9.0</td>
<td>46.0</td>
<td>1</td>
<td>[AA02742016, AA02741969, AA02739767, AA02753810]</td>
<td>4.0</td>
</tr>
</tbody>
</table>
MUAC Case Study

• Maastricht Upper Area Control Center manages the upper airspace over Belgium, the Netherlands, Luxembourg and north-west Germany from FL245 to FL660 - one of Europe’s busiest and complex airspace areas.
• By projecting MUAC onto our grid we are able to analyse the traffic and weather data.

3,694 air-blocks per level, 14,596 air-blocks total
• Decrease in MUAC Occupancy on Tuesday, May 29th compared to other Tuesdays.
• Decrease can be attributed to afternoon thunderstorms.
Data Science Algorithms

- Use of time series data to create algorithms to predict MUAC occupancy.
- Multiple linear regression, $y = a + b_1x_1 + b_2X_2 + \ldots$
- Decision Tree Regression, determine data splits in a tree fashion using independent variables.
- Neural Network, hidden node "network" between input and output layers, each node includes weights and has sigmoidal activation.
- LSTM, similar to NN also includes values at previous time steps as input.

<table>
<thead>
<tr>
<th>Model Type</th>
<th>RMSE</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple Linear Regression</td>
<td>25.6</td>
<td>0.978</td>
</tr>
<tr>
<td>Decision Tree Regression</td>
<td>24.4</td>
<td>0.980</td>
</tr>
<tr>
<td>Neural Network</td>
<td>19.6</td>
<td>0.987</td>
</tr>
<tr>
<td>LSTM</td>
<td>17.7</td>
<td>0.990</td>
</tr>
</tbody>
</table>
More Features...

• Arrivals
• Regulations
Conclusions

• Quantitative impact of weather is noticeable in data.
• Traffic data is highly cyclical and allows for fairly accurate modelling of time series.
• More data is needed to validate if models accurately capture impact of weather.
• Exploring correlation with other weather and traffic features.
• Exploring multi-dimensional input vs 1-D time series.
Step 2: Predicting Storms

• Integrating EPS forecast and storm observation data.

• Using ML to predict storms based on NWP forecast parameters.
Thank you

Questions?
Comments?
Suggestions?