Toward a machine learning based ceiling forecast diagnosis for TAF initialization (IniTAF project)

P. Crispel, P. Jaunet, A. Drouin, T. Kanitz, A. Audevant, C. Delin
Météo France DSM/AERO

Context – IniTAF project

The IniTAF project is intended to provide TAF initialization by using time to forecast the evolution of the most critical parameters (i.e., horizontal visibility in case of fog formation, etc.).

Several inputs are required to initialize TAFs from model data: wind, cloud cover, ceiling, etc.

Some of them are direct calculations from WNP models: wind speed and direction.

Some require further development: visibility, ceiling.

Ceiling – Aeronautical definition and thresholds

Aeronautical ceiling is the lowest cloud layer base height with:
- cloudiness greater than 50% r a radius of 8 km from the airport
- cloud base height ≤ 5000 ft

Note that:
- A ceiling upper than 5000 ft does not have to be mentioned in TAF messages.
- Only ceiling lower than 1500 ft does not have to be mentioned in the evolution groups of the TAF.

Table: Ceiling observations year 2016.

<table>
<thead>
<tr>
<th>Ceiling height</th>
<th>High Score</th>
<th>False Alarm</th>
</tr>
</thead>
<tbody>
<tr>
<td>200 ft</td>
<td>47%</td>
<td>63%</td>
</tr>
<tr>
<td>&lt; 1500 ft</td>
<td>61%</td>
<td>34%</td>
</tr>
<tr>
<td>&lt; 5000 ft</td>
<td>50%</td>
<td>17%</td>
</tr>
</tbody>
</table>

Conclusion: Ceiling forecast needs to be improved for all thresholds.

State of the art - Ceiling forecast by using cloud fraction vertical profiles from AROME - Comparison to METAR observations on French airports

Machine learning approach:

Linear regression, Random Forest, Extra Trees

Learning dataset:
Observations: Ceiling information from METARs (year 2016 – 66 French airports – hourly data) + 10’ observations.

Predictors:
- AROME forecast parameters: humidity, horizontal wind, turbulence, etc.
- Hourly extracted in a 3D grid around the airport (20 km x 20 km x 4000 ft).

Data engineering:
- Pre-treatment step: rescaling, correlation study, predictor number optimization.
- Unbalanced data: learning mechanism to sample data.
- Too much predictors due to 3D extraction - Spatial information has to be critically compressed.

Parameter on 3D grids (humidity, visibility)

Summarized in a vertical profile

Predictors

Figure: Example of 3D data reduction for the humidity predictor.

Scores

<table>
<thead>
<tr>
<th>Ceiling height</th>
<th>High Score</th>
<th>False Alarm</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 200 ft</td>
<td>46%</td>
<td>65%</td>
</tr>
<tr>
<td>&lt; 1500 ft</td>
<td>73%</td>
<td>40%</td>
</tr>
<tr>
<td>&lt; 5000 ft</td>
<td>79%</td>
<td>26%</td>
</tr>
</tbody>
</table>

Conclusion: Ceiling forecast needs to be improved for all thresholds.

Unique model vs Multi-airport models

<table>
<thead>
<tr>
<th>Ceiling height</th>
<th>High Score</th>
<th>False Alarm</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 200 ft</td>
<td>46%</td>
<td>65%</td>
</tr>
<tr>
<td>&lt; 1500 ft</td>
<td>73%</td>
<td>40%</td>
</tr>
<tr>
<td>&lt; 5000 ft</td>
<td>79%</td>
<td>26%</td>
</tr>
</tbody>
</table>

Conclusion:
- Performance: Single model for each airport show in general better results.
- Some airports in Mediterranean region have few cases of low ceiling to perform final learning procedure.
- Still very difficult to forecast efficiently low ceiling events.

10 Convolutional Neural Network (CNN) approach

Advantages: Introduce a convolutional step inside CNN to combine local information and identify features in vertical profiles.

Neural network and 3D images

Handle vertical and horizontal information.

Methodology: add 3D convolutional filters in the CNN to locally combine and compress horizontal information.

Neural network and times series

Handle time information.

Methodology: use Recurrent Neural Network to learn from predictor sequences.

Figure: Example of a humidity vertical profile in a 20 x 20 km airport neighborhood.

Figure: Example of a humidity time series compared to ceiling observation in METAR (blue line).