Some aspects of the verification of deterministic ECMWF forecasts at Météo-France

Frédéric Atger
November 2003
Topics

• 1) Objective comparison ECMWF T511 vs Arpège
  • Which one is better in average?
  • Are forecasts worse when not in agreement?

• 2) Subjective comparison ECMWF T511 vs Arpège
  • Is there a better model?
  • Are forecasts better when in agreement?
  • What about using a 3rd model?

• 3) Local wind forecasts
  • ECMWF T511 model vs Arpège/Aladin: the impact of model resolution
T511 vs Arpège: which one is better?

500-hPa geopotential height (day+1)
1) Objective verification

- 500 hPa geopotential height RMSE (error)
- T511 - Arpège RMSD (difference)
- Europe-Atlantic domain (synoptic scale)
- Verified wrt Arpège analysis
RMSE and RMSD
778 days (2001-2003) +72/+84 forecasts

Both poor (28%)

Arpège good, T511 poor (9%)

Both good (45%)

T511 good, Arpège poor (18%)
Graphical representation

T511 RMSE

ref

Arpège RMSE

RMSD

Arpège

Both good

Arpège poor, T511 good

Arpège good, T511 poor

Both poor

Classification according to RMSD and RMSEs
Are forecasts better when in agreement?

- What is a "good" forecast?
- What are "different" forecasts?
- 35m ~ day+2/day+3 forecast RMSE

RMSE < 35m
RMSD < 35m

"Good" cases
Arpège RMSE < 35m 54%
T511 RMSE < 35m 62% (+8)

General case
54%
62% (+8)

RMSD > 35m (48%)
29%
45% (+16)
2) Subjective assessment by the forecasters (synoptic pattern wrt the weather in France)

<table>
<thead>
<tr>
<th></th>
<th>Day+1</th>
<th>Day+2</th>
<th>Day+3</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Very good&quot; T511</td>
<td>83%</td>
<td>46%</td>
<td>30%</td>
</tr>
<tr>
<td>&quot;Very good&quot; Arpège</td>
<td>83%</td>
<td>51%</td>
<td>27%</td>
</tr>
<tr>
<td>Synoptic difference</td>
<td>27%</td>
<td>62%</td>
<td>82%</td>
</tr>
<tr>
<td>&quot;Very good&quot; Arpège when no difference</td>
<td>90%</td>
<td>62%</td>
<td>52%</td>
</tr>
</tbody>
</table>

A 2-member poorman ensemble!
Using UK when T511 and Arpège differ?
When T511 and Arpège differ, where is the UK model?

<table>
<thead>
<tr>
<th></th>
<th>Day+1</th>
<th>Day+2</th>
<th>Day+3</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK supports T511 or Arpège</td>
<td>66%</td>
<td>67%</td>
<td>71%</td>
</tr>
<tr>
<td>UK between T511 and Arpège</td>
<td>30%</td>
<td>27%</td>
<td>18%</td>
</tr>
<tr>
<td>UK gives a 3rd alternative</td>
<td>4%</td>
<td>6%</td>
<td>11%</td>
</tr>
</tbody>
</table>

Most of the time a 2-member poorman ensemble is large enough to sample the synoptic uncertainty.
A 3-member poorman ensemble?

When Arpège and T511 are not in agreement, and one of them is supported by UK, does it help to choose?

<table>
<thead>
<tr>
<th></th>
<th>Arpège</th>
<th>T511</th>
<th>The one supported by UK</th>
</tr>
</thead>
<tbody>
<tr>
<td>67%</td>
<td>67%</td>
<td>67%</td>
<td>78%</td>
</tr>
<tr>
<td>47%</td>
<td>36%</td>
<td></td>
<td>47%</td>
</tr>
<tr>
<td>22%</td>
<td>26%</td>
<td></td>
<td>31%</td>
</tr>
</tbody>
</table>
3) Local wind forecasts
The impact of model resolution

Black : T511 (0.5°)
Red : Aladin (0.1°)
VENT 10m valable le 08/10/2003
Aladin base 07/10/2003 base 00UTC P24 rouge
Aladin base 07/10/2003 base 06UTC P18 violet
Aladin base 07/10/2003 base 12UTC P12 bleu
10-meter wind direction

12 months (2000-2001)
587 locations in France
16 sectors percent correct

12 months (2000-2001)
587 locations in France
16 sectors percent correct
• Red/orange = Aladin better
• Green/blue = T511 better
• Level of significance of the difference between T511 and Aladin (non parametric statistical test based on resampling)
• Wind direction percent correct
• Day+1, 12 UTC
10-meter wind speed

12 months (2000-2001)
587 locations in France
+/- 2 kts percent correct (5kts, 10kts, etc)
• Red/orange = Aladin better
• Green/blue = T511 better
• Level of significance of the difference between T511 and Aladin (non parametric statistical test based on resampling)
• Wind direction percent correct
• Day+1, 12 UTC
Summary

• Objective verification says T511 gives better guidance
• Subjective evaluation says T511 and Arpege have a similar level of performance
• Both subjective and objective verification show the efficiency of a poorman ensemble approach
• Model resolution does matter when forecasting local surface wind, but:
  • Impact is clear for direction, not really for speed
  • Local effects dominate the performance
Acknowledgements

• Objective verification: Marc Tardy
• Subjective evaluation: Bruno Gillet-Chaulet
• Wind forecasts: Isabelle Souyri