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

Frédéric Atger, Météo-France

1 ECMWF T511 model vs. Arpège model – objective verification

1.1 *Which one is better in average?*

In this section 12-UTC ECMWF T511 +84h forecasts have been compared to 00-UTC Météo-France Arpège +72h forecasts, with respect to the 500-hPa geopotential height over a 'synoptic' Europe-Atlantic domain (1.5° grid). Figure 1 shows an over-performance of T511 forecasts, despite the lead-time advantage of Arpège forecasts (justified in this comparison by the fact that both are available on the bench at the same time in the early morning).

Figure 1 also shows a decrease of the forecast performance in case of a large difference between Arpège and T511 forecasts. This aspect is investigated further in the next sub-section.

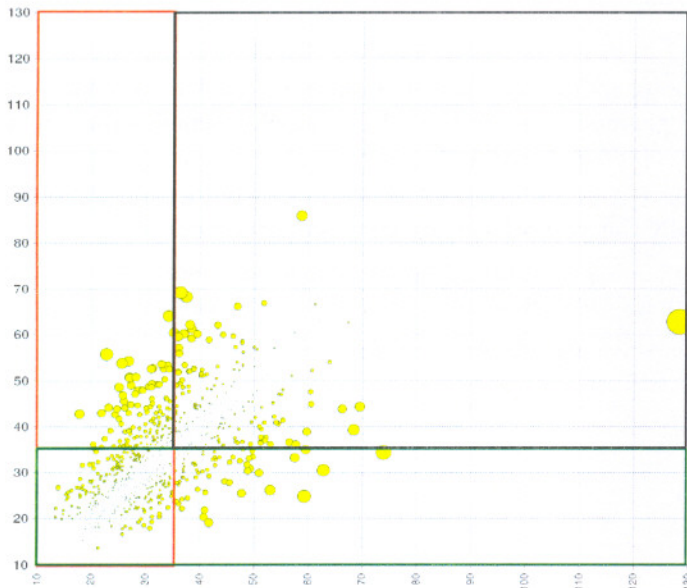### 1.2 *Are forecasts worse when not in agreement?*

Fig. 1 T511 RMS error (x-axis) vs. Arpège RMS error (y-axis). The size of the dots (778 days in 2001-2003) indicates the RMS difference between the T511 and the Arpège forecasts. The red frame contains 'good' T511 forecasts, while the green frame contains 'good' Arpège forecasts ('good' means here with RMSE<35m).

Table 1 confirms a strong reduction of the forecast performance when the 2 models are not in agreement. This reduction is stronger for Arpège forecasts than it is the case for T511 forecasts, so that the over-performance of T511 forecasts is emphasized when the 2 models are not in agreement.

2 ECMWF T511 model vs. Arpège model - Subjective assessment

2.1 *Is there a better model?*

Table 1	General case	RMS (Arpège-T511) > 35m
Proportion of cases when Arpège RMS error < 35m	54%	29%
Proportion of cases when T511 RMS error < 35m	62%	45%

A subjective comparison has been done routinely by national forecasters between synoptic forecasts based on the interpretation of 12-UTC ECMWF T511 forecasts and 00-UTC Météo-France Arpège forecasts. Table 2 shows a similar performance for the 2 models. This result (based on a 6-month evaluation in 2003, but supported by previous results based on longer evaluations) contrasts with the objective over-performance of T511 forecasts shown in the previous section.

2.2 *Are forecasts better when in agreement?*

Table 3 (first row) shows the proportion of cases when T511 forecasts and Arpège forecasts exhibit synoptic differ-

Table 2	Day + 1	Day + 2	Day + 3
'Very good' T511 forecast	83%	46%	30%
'Very good' Arpège forecast	83%	51%	27%

ences that are considered significant by the forecasters. The numbers increase with the lead-time. Table 3 (second row) shows that forecasts are definitely better, from the forecaster's point of view, when they are in agreement than in the general case, especially on day +3: the proportion of 'good' Arpège forecasts increases from 27% in the general case (Table 2) to 52% when the models are in agreement. This result suggests the potential usefulness of a 2-member poor man ensemble forecasting strategy.

2.3 What about using a 3rd model?

Table 3	Day +1	Day + 2	Day + 3
Significant synoptic difference	27%	62%	82%
'Very good' Arpège forecast when no significant difference	90%	62%	52%

In this section the relevance of a 3-member poor man ensemble strategy is investigated by considering forecasts based on the global UK Met Office model (available in real time for interpretation by French forecasters). Table 4 shows the proportion of cases when forecasters judge that UK forecasts support or not T511 and/or Arpège forecasts.

Table 4 suggests that most of the time a 2-member poor man ensemble samples the main part of the synoptic uncertainty apprehended by an operational forecaster. Nevertheless, when considering the special (and crucial) case when Arpège and T511 forecasts are not in agreement, Table 5 shows that the 3rd member may still help to arbitrate between them, at least on day +1.

3 Conclusions

Table 4	Day + 1	Day + 2	Day + 3
UK forecast supports T511 or Arpège	66%	67%	71%
UK forecast between T511 and Arpège	30%	27%	18%
UK forecast gives a 3rd alternative	4%	6%	11%

Table 5	Day +1	Day + 2	Day + 3
Arpège and T511 forecasts are not in agreement; one of them is supported by the UK forecast			
'Very good' Arpège forecast	67%	47%	22%
'Very good' T511 forecast	67%	36%	26%
'Very good' UK-supported forecast	78%	47%	31%

The main conclusions are the following:

- Objective verification shows an over-performance of T511 forecasts compared to Arpège forecasts, despite the lead-time advantage of the latter.

- By contrast, subjective evaluation shows no significant difference in performance between synoptic forecasts based on an interpretation of T511 and Arpège models.

- Both objective and subjective evaluation shows the potential usefulness of a 2-member poor man ensemble forecasting strategy, based on a daily comparison of T511 and Arpège forecasts.

- Although most of the time a 2-member poor man ensemble seems sufficient to sample the synoptic uncertainty, a 3rd member may still help to arbitrate between 2 conflicting options, at least on day +1.

Acknowledgements