Forecast Inconsistencies - how often do forecast jumps occur in the models?

Ervin Zsótér[1], Roberto Buizza[2] and David Richardson[2]

[1]Hungarian Meteorological Service, [2]European Centre for Medium-Range Weather Forecasts

1. Introduction

One of the key advantages that should be expected from using an ensemble prediction system (EPS) instead of a single forecast, i.e. the EPS control or the high resolution deterministic forecast, is a better consistency between consecutive forecasts (Buizza 2008a). However, forecasters occasionally experience large inconsistencies in the EPS, some of which show a zigzagging fashion when not only the EPS (represented by the ensemble mean) but also the single forecasts jump in phase for a few consecutive runs. In this work we investigated this important property of the forecasts, i.e. the inconsistency or *'jumpiness'*, putting special emphasis on the relationship between the control's and the ensemble-mean's inconsistency behaviour and on the connection between forecast inconsistency and forecast skill.

2. Inconsistency indices

We used a set of indices to measure forecast inconsistency and 'jumpiness'.

- The inconsistency index $INC_{\Sigma}[f_j^C(d,t),\delta]$ between two forecasts issued δ-hours apart (in our case it was 12 hours) over the area Σ is defined as the ratio between their distance and their average standard deviation:

$$INC_{\Sigma}[f_j^C(d,t),\delta] = \frac{d_{\Sigma}[f_j^C(d,t), f_j^C(d-\delta,t+\delta)]}{0.5\{std_{\Sigma}[f_j^C(d,t)] + std_{\Sigma}[f_j^C(d-\delta,t+\delta)]\}}.$$

In $f_j^C(d,t)$ d is the run date and t is the step, j is either 0 or m corresponding to the control or the ensemble-mean, and C identifies the origin (EC- or UK-EPS). d_{Σ} is the average difference of the two forecast fields, and std_{Σ} is the standard deviation of the field values over the area Σ. The denominator has been introduced to take into account the fact that different forecasts have different characteristics (e.g. resolution, variability). This scaling factor normalizes the differences between consecutive forecasts to the variability of the forecasts themselves.

- The inconsistency correlation $CORR_{\Sigma}[F_i, G_i]$ between two inconsistency index time series, F_i and G_i (where $i=1,2,\ldots,N$ refers to the runs) over the area Σ, is defined by the commonly used product-moment correlation coefficient (Neter et al. 1988). The inconsistency correlation measures how similar the inconsistency behaviour of two different forecasts is, either in an individual EPS system (with f and g representing the control and the ensemble-mean), or between two EPS systems (with f and g representing the two controls or the two ensemble-means).
- A forecast inconsistency is characterized as a forecast jump when the absolute value of the corresponding inconsistency index is larger than half the period-average inconsistency over N forecast runs (defined as the root-mean-square average of the daily inconsistency index values). The single 'flip', single 'flip-flop' and single 'flip-flop-flip' indices are defined as the frequency of events when a forecast makes a single jump ('flip'), or two ('flip-flop') or three ('flip-flop-flip') consecutive jumps with alternating signs (zigzagging).
- To help characterize the strength of the relationship between two forecast systems – such as e.g. the control and the ensemble-mean forecasts of an EPS, or the control forecasts of two different EPSs – we extend the concept of the single forecast jump indices. A two-forecast system makes a parallel 'flip', parallel 'flip-flop', or parallel 'flip-flop-flip', when both forecasts make a single 'flip', single 'flip-flop', or single 'flip-flop-flip' respectively, and the jumps are in phase.

3. Framework of the forecast inconsistency experiments

The forecast inconsistency behaviour was investigated by considering the control and the ensemble-mean forecasts, both 00 UTC and 12 UTC runs, of the ECMWF (EC-EPS) and the UK Met Office (UK-EPS) operational ensemble systems for a period from 16 February 2007 to 31 August 2009 (~2.5 years). The forecasts were extracted from the TIGGE archive hosted at ECMWF (see e.g. http://tigge.ecmwf.int/ for more information) on a regular 1°x1° latitude-longitude grid. To account for the pole-ward distortion of the grid, the grid point values were weighted by the cosine of the latitude in area-average computations.

The inconsistency behaviour was investigated on 4 parameters: 500 and 1000 hPa geopotential height, and 500 and 850 hPa temperature. Four different areas were considered to assess the sensitivity of the inconsistency indices to the area: three over North-Western Europe with different size but centered on the same point, and a separate area over south-east Europe with the same size as the middle NW Europe area. Furthermore, to assess the seasonal variability of the ensemble 'jumpiness', ensemble data from June 2003 to August 2009, i.e. for about 6 years, have also been considered for the EC-EPS.

4. A prominent example of jumpy forecasts

During this event of February 2008, the event which actually triggered our investigation, the EPS control and ensemble-mean forecasts were very jumpy for a few days in a row over northern Europe. The jumpy behavior of the 12-hour apart consecutive forecasts is highlighted on Fig. 1 by the arrows indicating the dominant flow direction. The forecasts started on the 16th (t+180h) and the 18th (t+156h) show a ridge stretching over the UK, while the intermediate forecasts started on the 17th (t+168h) indicate a significantly more zonal flow. Note that the ensemble-mean followed the control forecast very closely, although the jumps in the ensemble-mean seem to show smaller amplitude.

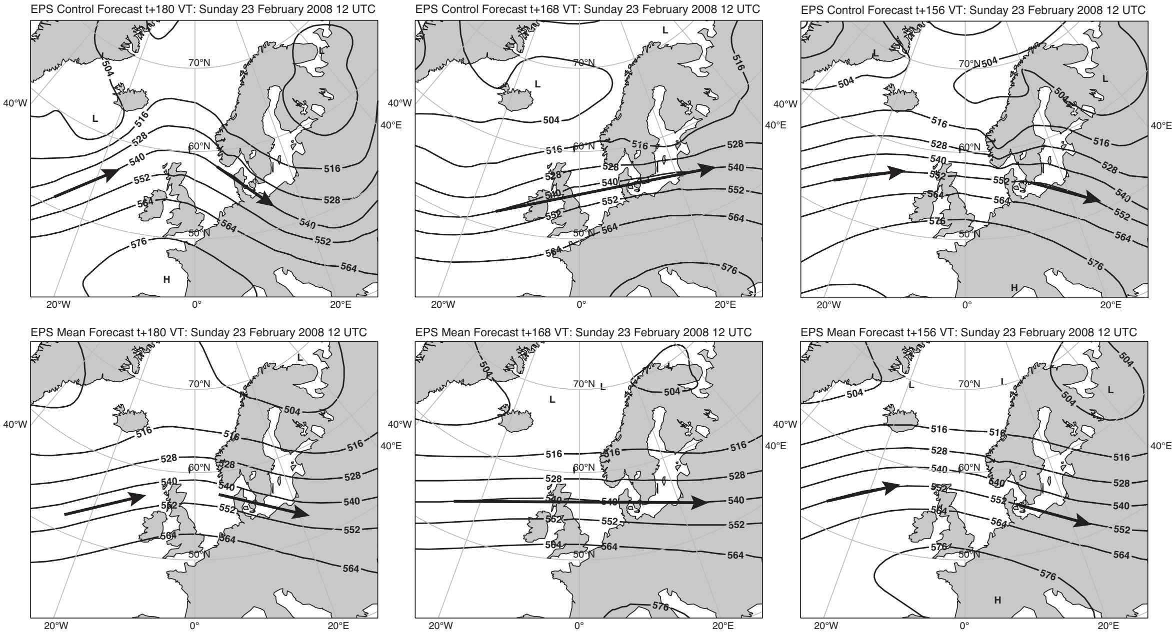

Fig. 1 EC-EPS control (upper row) and ensemble-mean (bottom row) forecasts of 500 hPa geopotential height (contour interval 120m) started at 00 UTC on 16 (left column), at 12 UTC on 16 (middle column) and at 00 UTC on 17 February (right column) and valid for 12 UTC on 23 February 2008. The black arrows highlight the dominant flow. Source: Zsoter et al 2009.

Fig. 2 shows the inconsistency index of the EC-EPS control and ensemble-mean forecasts as a function of forecast range for all forecasts verifying at 12 UTC on 23 February. The ensemble-mean forecast follows the control forecast rather closely between forecast steps t+240h and t+144h, and it jumps less for longer forecast ranges. At short forecast range both forecasts change less and follow each other very closely.

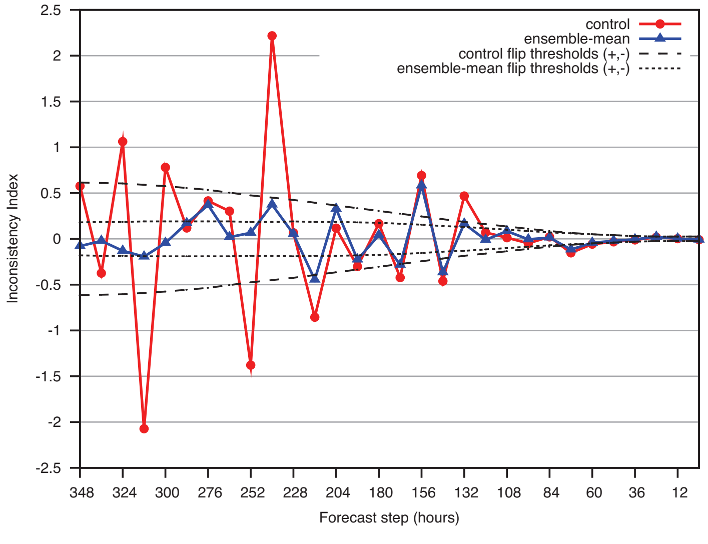

Fig. 2 EC-EPS control (red line) and ensemble-mean (blue line) inconsistency indices for 500 hPa geopotential height forecasts valid for 12 UTC on 23 February 2008. To help interpreting the concept of forecast jumps, half the period-average inconsistency (with positive and negative signs) is also indicated by dashed (control) and dotted (ensemble-mean) lines. The forecast steps are from t+348h to t+0h, every 12 hours. The indices have been computed over the medium North-western European area.

The dashed and dotted lines in Fig. 2 show the threshold values used in the definition of single and parallel forecast jumps. The control and / or the ensemble-mean makes a 'flip' if the corresponding inconsistency indices (circles for the control, triangles for the ensemble-mean) are outside of the area bounded by the two dashed (control) or dotted (ensemble-mean) lines. The inconsistency values at t+168h and t+156h correspond to the forecasts displayed in Fig. 1. These inconsistency-index pairs actually rank as a single 'flip-flop' of the control and also of the ensemble-mean, because they zigzag outside of the jump-threshold lines (first both geopotential height fields get lower, then on the next run both get higher again). Moreover, since the control and the ensemble-mean jump in the same direction, they also make a parallel 'flip-flop'. A closer inspection actually indicates a parallel 'flip-flop-flip' of the four consecutive 180-, 168-, 156-, and 144-hour forecasts. Note also that from t+228h to t+144h the control and the ensemble-mean always jump in phase.

5 Statistics of the 'jumpiness'

For reason of space we summarize the results in the following two sections only for the 500 hPa geopotential height over the medium North-western European area for the whole 2.5-year sampling period. The forecast steps are displayed consistently from t+348h to t+0h, every 12 hours.

Figure 3a shows that all average inconsistency indices increase with the forecast range in agreement with the practical experience that the forecasts are usually more consistent at short forecast range. Figure 3a also shows that for both ensemble systems, the ensemble-mean period-average inconsistency index is smaller than the control one, especially at long forecast ranges.

In both the EC- and UK-EPS, the initial perturbations are centered on the control forecast, therefore the correlation between control and ensemble-mean inconsistency will be high at short range, while initial perturbations evolve linearly, and then decrease as perturbation growth becomes more non-linear (Fig. 3b). The correlation is slightly stronger in the EC-EPS than in the UK-EPS during the first few days, while it is weaker in the medium-range.

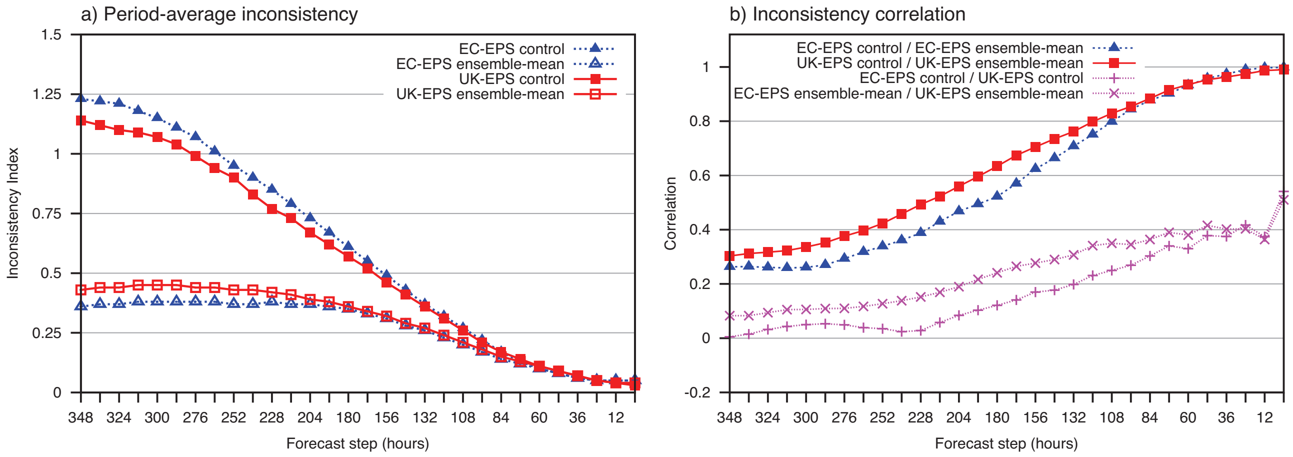

Fig. 3 a) Period-average inconsistency for the control and the ensemble-mean forecasts of the EC- and the UK-EPS; b) Inconsistency correlation between the control and the ensemble-mean forecasts of the EC- and the UK-EPS separately, plus between the EC- and the UK-EPS control forecasts, and between the EC- and the UK-EPS ensemble-mean forecasts.

It is interesting to contrast these correlations with the correlation between the inconsistencies of the two control forecasts or the two ensemble-mean forecasts. Figure 3b also shows that the inconsistency correlation between the two control forecasts (pluses) and the two ensemble-mean forecasts (crosses) are much lower than the correlations obtained for each individual system (squares and triangles). The large difference between the same-system and the different-system correlations throughout the forecast range could be seen as an indication that each single system is still not capable to properly simulate the effect of model or analysis uncertainty.

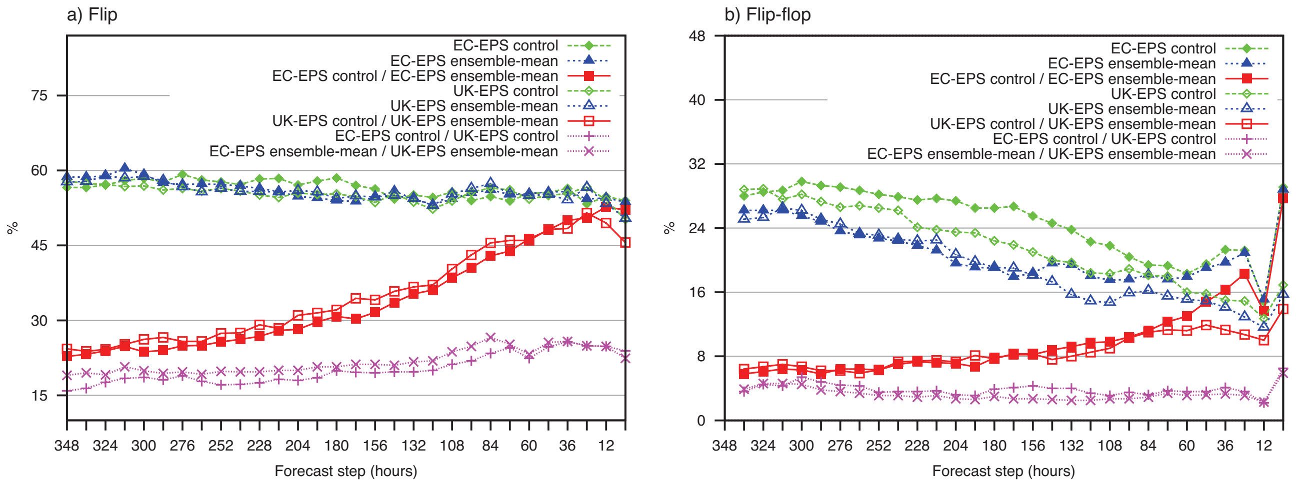

Fig. 4 Frequency statistics of the 'flip' (a) and the 'flip-flop' (b) occurrences for the EC- and the UK-EPS control and ensemble-mean forecasts. Beside the single versions also the statistics for parallel 'flips' and 'flip-flops' are displayed for the EC- and the UK-EPS respectively and also by combining the EC- and the UK-EPS's control forecasts and ensemble-mean forecasts.

Occasionally, for example the case shown earlier, the ensemble-mean follows the control very closely in a zigzagging manner for a number of runs. If such situations happened too often, the capability of the EPS to improve the prediction of potentially damaging weather scenarios could be reduced. To measure the frequency of occurrence of such events we counted the number of cases when the control and the ensemble-mean forecasts - either individually or in parallel - made 'flips', 'flip-flops' and 'flip-flop-flips'.

The frequency of individual 'flips' is similar throughout the forecast range for the control and ensemble-mean forecasts for both systems (Fig. 4a). The parallel 'flip' frequency between the control and the ensemble-mean forecasts of each system reflects the weakening relationship between the control and the ensemble-mean with the forecast range: at short range almost every jump occurs simultaneously, but this decreases steadily through the medium-range. The parallel 'flips' between EC- and UK-EPS also reflect the correlations discussed earlier: the two systems jump together substantially less often than the same-system control and ensemble-mean.

The frequencies of 'flip-flops' (Fig. 4b) and 'flip-flop-flips' (not shown) are lower than for 'flips'. For the ensemble-mean, the 'flip-flops' and 'flip-flop-flips' indices are noticeably smaller than the control indices for both EPS systems, especially in the medium range. Thus, although the control and ensemble-mean forecasts have a similar numbers of 'flips', the ensemble-mean has a definite smaller frequency of 'flip-flops' and 'flip-flop-flips'. This confirms the earlier results that the ensemble-mean jumps less than the control forecast.

6 Relationship between 'jumpiness' and forecast error

Routine assessment of the ECMWF ensemble performance has indicated that there is a certain degree of correlation between ensemble spread and forecast error, with cases with lower-than-average ensemble standard deviation characterized by lower-than-average ensemble-mean errors (see e.g. Buizza and Palmer 1998). It is useful to know whether there is a similar relationship between 'jumpiness' and forecast error and the cases characterized by less inconsistent forecasts are easier to predict?

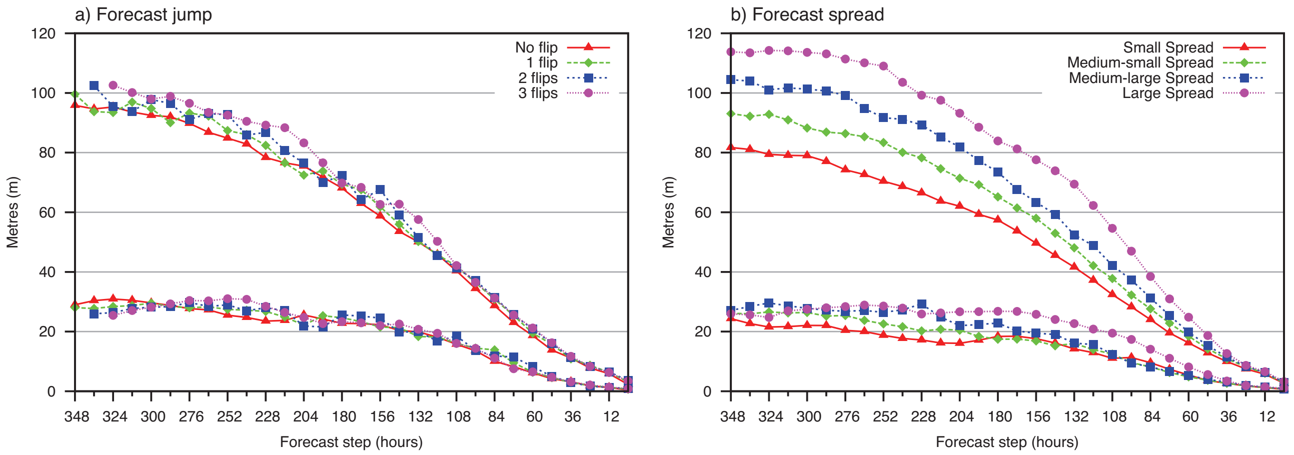

Fig. 5 a) Mean (lower cluster of lines) and standard deviation (upper cluster of lines) of forecast errors in subsamples with 0 (red), 1 (green), 2 (blue) and 3 (magenta) consecutive 'flips'. b) Mean (lower cluster of lines) and standard deviation (upper cluster of lines) of forecast errors in subsamples with small (red) medium-small (green), medium-large (blue) and large (magenta) ensemble spread. The statistics are based on a 9-month sample consisting only of the summer months (June, July, August).

To assess this, forecast error statistics have been computed for subsets of cases characterized by either 0, 1, 2 or 3 consecutive 'flips'. Similarly, forecast error statistics have been computed for cases with small, medium-small, medium-large and large ensemble spread (the spread was classified as small when it was below the average spread minus one spread standard deviation, medium-small when it was between the average and the average minus one standard deviation, etc). This assessment was made for the EC-EPS over the three summer months June, July and August, so that 9 months worth of data were available within the 2.5-year total period. Figure 5 shows the average and standard deviation of rms errors of the control and ensemble-mean forecasts for the cases with flip numbers equal to 0, 1, 2 and 3 (Fig 5a), and the corresponding average and standard deviation of errors for the cases with small, medium-small, medium-large and large spread (Fig 5b). Results show clearly that while the ensemble spread is a good predictor of the forecast error, there is only a weak relationship between forecast 'jumpiness' and forecast error.

For either the control or the ensemble-mean forecast, the 'flip' number is a measure of the spread of consecutive, lagged forecasts. This result indicates that there is only a moderate correlation between the spread of a lagged ensemble and the error of the most recent forecasts. This result is in line with the conclusions of Buizza (2008b) who showed that the ECMWF operational EPS performs better than a lagged ensemble based on the 6 most recent high-resolution forecasts.

For further details on the inconsistency experiments and related results please consult Zsoter et al 2009.

References

Buizza, R., and **T. N. Palmer**, 1998: Impact of ensemble size on ensemble prediction system. *Mon. Wea. Rev.*, **126**, 2503-2518.

Buizza, R., 2008a: The value of Probabilistic Prediction. *Atmos. Sci. Lett.*, **9**, 36-42 DOI: 10.1002/asl.170.

Buizza, R., 2008b: Comparison of a 51-member low-resolution (TL399L62) ensemble with a 6-member high-resolution (TL799L91) lagged-forecast ensemble. *Mon. Wea. Rev.*, **136**, 3343-3362.

Neter, J., W. Wasserman, and **G. A.Whitmore**, 1988: Applied Statistics. Allyn and Bacon Inc., pp. 1006.

Zsoter, E., R. Buizza, D. Richardson, 2009. 'Jumpiness' of the ECMWF and the UK Met Office EPS control and ensemble-mean forecasts. *Mon. Wea. Rev.*, **137**, 3823-3836.