Recent Developments in the ECMWF EPS

R. Buizza - J. Barkmeijer - D.S. Richardson - T.N. Palmer
European Centre for Medium-Range Weather Forecasts

Abstract

Three changes in the ECMWF operational forecast productions are discussed: the resolution increase in the operational European Centre for Medium-range Weather Forecasts (ECMWF) Ensemble Prediction System to 80-km, the introduction in the EPS of tropical perturbations defined by singular vectors targeted to maximize total-energy inside an area centred on tropical storms and the implementation of the multi-analysis stream based on 4 forecasts run with the ECMWF $T_L255L40$ version of the ECMWF model but starting from analyses produced at different centres.

The increase in EPS resolution and the inclusion of tropical perturbations has led to an improvement in the EPS accuracy and capability to forecast tropical storms' tracks. Since implementation, the multi-analysis stream has been used to identify failures of the ECMWF forecasting system due to analysis problems. Results from pre-operational experimentation that lead to these three latest operational implementations are summarized. On-going research projects aiming to develop a severe-weather warning system are also discussed, and preliminary results are shown.

1. Introduction

A complete description of the weather prediction problem can be stated in terms of the time evolution of an appropriate probability density function in the atmosphere's state space. Ensemble prediction based on a sampling of this probability density function by a finite number of deterministic integrations designed to represent both initial and model uncertainties, appears to be the only feasible method to predict the PDF beyond the range of linear error growth (*Leith* 1974). The operational Ensemble Prediction System (EPS) of the European Centre for Medium-Range Weather Forecasts (ECMWF) has been developed to provide an estimate of the probability density function of forecast states.

The EPS became part of the ECMWF operational suite in December 1992 (*Molteni et al.* 1996) with 33 members run at T63L19 resolution (spectral triangular truncation T63 with 19 vertical levels. The system was upgraded to 51 members run at $T_L159L31$ resolution (spectral truncation T159 with linear grid) in 1996 (*Buizza et al.* 1998). Since March 1998, the EPS initial conditions have been defined using initial and evolved singular vectors so that the direction of fastest error growth during both the data-assimilation period and the first 2 days of the forecast period are sampled (*Barkmeijer et al.* 1999). Since October 1998, random model errors due to parameterised physical processes have been simulated (*Buizza et al.* 1999). The EPS resolution has been further increased to $T_L255L40$ on 21st November 2000 (*Buizza et al.* 2002). On 22nd January 2002 tropical singular vectors designed to sample fastest growing perturbations in the tropical region were added to the EPS initial perturbations (*Barkmeijer et al.* 2001).

This work discusses the impact of the two latest EPS changes and reports first results from an EPS parallel stream constituted of an ensemble of 5 TL255L40 ECMWF forecasts started from initial conditions provided by 4 different centres (UK Meteorological Office, Deutsche Wetterdienst, Météo-France and National Centers for Environmental Prediction). Then, it discusses preliminary results from ongoing research projects designed to further improve the ECMWF probabilistic forecasting system.

2. The high-resolution (80-km) EPS

Buizza et al. (2002) compared the performance of the high-resolution EPS (HEPS) and the EPS for 87 cases covering two periods, summer 1999 (30 cases, from 2 to 30 August) and winter 1999-2000 (57 cases, from 26 November to 27 December 1999 and from 22 January to 15 February 2000). HEPS scores have been contrasted with EPS scores and with EPS scores shifted by 1-day (EPS(d-1)), i.e. with the scores of an EPS system characterised by a 1-day gain in skill. For any score measure SC, skill gains/losses have been quantified by the Relative Improvement index RI(SC)

$$RI(SC) = \frac{SC[HEPS] - SC[EPS]}{SC[EPS(d-1)] - SC[EPS]} \quad \text{Eq. 1}$$

The RI index is a normalised measure of the gain in skill obtained by configuration HEPS. RI=100% indicates an improvement equivalent to a 1-day gain in skill when measured using the score SC. Since the EPS has been principally designed for forecast ranges of 2 days or longer, the RI has been computed only for forecast days 2 to 10.

The accuracy of deterministic 850 hPa temperature forecasts by the EPS control and the ensemble-mean has been measured using anomaly correlation coefficient (ACC). The following two events have been considered: '850hPa temperature positive anomalies larger than one standard deviation' and '850hPa temperature negative anomalies larger than one standard deviation'. The accuracy of the control, the ensemble-mean and the EPS probabilistic prediction of these two events have been assessed using the Brier skill score (BSS).

Figure 1 shows the relative improvement index, RI, computed over the Northern Hemisphere (NH) for 5 skill measures: control ACC and BSS, ensemble-mean ACC and BSS, and EPS BSS. Results indicate that for summer (Fig. 1a) RIs are positive for all but forecast days 2 and 10, while for winter (Fig. 1b) all RIs are positive. The day-2 negative RI shown for the control and the ensemble-mean are due to the fact that the $T_L319L60$ analysis is used as verification. Note that only the control forecasts (but not the ensemble-mean or EPS forecasts) show a negative RI at day-10. Overall, RI results show that the summer HEPS probabilistic predictions are 55-70% better than the EPS (Fig. 1a) and that the winter HEPS are 45-66% better than the EPS (Fig. 1b) between forecast day 5 and 7. Figure 2 shows the equivalent RIs computed for Europe.

Comparing the RIs computed for the BSS of the control, the ensemble-mean and the EPS it can be seen that for all forecast steps, the largest RIs are those for the EPS. In particular, the EPS RIs are always larger than the control RIs, especially at the end of the forecast period. This indicates that the system upgrade from EPS to HEPS has induced a larger relative impact on the ensemble probability forecasts than on the deterministic forecasts given by the control or the ensemble-mean forecast.

3. Tropical singular vectors

With the development of linearised versions of important components of the ECMWF forecast model (Mahfouf 1999), such as, vertical diffusion, sub grid-scale orographic effects, large-scale condensation, long-wave radiation and deep cumulus convection, it is possible to determine singular vectors for situations where physical processes may contribute significantly to perturbation growth. The development of this package of linearised physics made it possible to extent the area of initial EPS perturbations to the tropical region.

Barkmeijer et al. (2001) and *Puri et al.* (2001) have described several properties of tropical singular vectors and their impact on the EPS performance. To benefit from tropical singular vectors in tropical cyclone ensemble forecasting, it was necessary to define target areas in the vicinity of tropical cyclone locations. As a consequence, multiple computations are required when several tropical cyclones coexist. In order to reduce the overall numerical costs for determining perturbations in the tropics, a multi-Gaussian sampling technique was adopted as described in Ehrendorfer (1999). For each tropical cyclone, 50 initial EPS perturbations are determined by sampling from the leading five singular vectors and added to the already existing perturbations.

The tropical singular vector computation does not differ from the extra-tropical computation with respect to the defining norms. Both at initial and final time the total energy norm is used. The main difference is the use of the newly developed package of linear physics. The number of singular vector computations in the tropical region may vary from day to day depending how many target areas are defined. However, to limit the numerical costs, the maximum number of target areas is set to four:

i. the Caribbean area (0°-25°N, 100°W-60°W) is always selected, since weather systems originating from the Caribbean area may influence medium-range European forecasts;

ii. for each tropical system with WMO classification index larger than 1 (hurricanes, typhoons, cyclones, severe tropical storms or tropical storm) in the tropical strip 25°S-25°N a 30°x40° target area centred at the system location and intersected by the strip 25°S-25°N is defined;

iii. if more than four areas have been defined, then merge the closest target areas to a maximum of four.

The tropical singular vectors have a clear impact on the EPS cyclone tracks. Figure 3 shows the tracks for cyclone ANDO of the 50 ensemble members for a forecast period of 4 days and starting from 4 January 2001. In the operational EPS, as shown in Figure 3a, all tracks follow the same route up to day 3 and there is a high probability that the cyclone will make landfall over La Reunion between days 2-3. The experimental EPS, however, makes this scenario less likely by indicating other possible cyclone tracks.

To investigate whether tropical perturbations are beneficial for the skill of the tropical cyclones tracks in the ensemble, distances between the ensemble tracks and the analysis for the tropical cyclone location have been computed for certain forecast periods and averaged over 14 cases. Figure 4 gives the number of members, averaged over the 14 cases, closer to the analysed tropical cyclone location for selected distances and forecast periods. For a lead-time of 1 day, the EPS unperturbed control forecast is quite capable to position the tropical cyclone accurately. As seen from Fig. 4, initial perturbations may slightly deteriorate the skill of the ensembles members for this short forecast range. However, for longer lead times the number of skilful members is larger in the experimental ensemble.

4. Multi-analysis forecast stream

ECMWF is now routinely running forecasts using the ECMWF model initialised with the 12 UTC analyses from Deutsche Wetterdienst (DWD), Météo-France, the National Centers for Environmental Predictions (NCEP) and the UK Meteorological Office (UK). A 10-day forecast is run from each analysis and from a 'consensus' analysis generated as a simple average over all the available analyses, including the ECMWF analysis. The forecasts are run at T_L255L40 resolution using the same configuration of the ECMWF model as is used for the EPS. This set of forecasts will be referred to as the multi-analysis (MA) system.

Because the analyses generated by different centres are not directly compatible (for example they use different horizontal grid spacing, different vertical levels and different orography), interpolation is needed to produce an analysis on the ECMWF model grid. This is done using the following procedure. Data is received from each centre as fields of wind, temperature and humidity on standard pressure levels. The difference between this pressure level analysis and the ECMWF pressure level analysis is calculated and then interpolated onto the ECMWF model levels. The resulting perturbation is added to the ECMWF analysis to produce an approximation to the analysis of the other centre. This method has been chosen to minimise the disturbance to the analysis fields – any inconsistencies are in the small perturbations rather than in the full fields. Only the upper-air fields are perturbed; there is no perturbation to the ECMWF surface fields. The only exception to the above procedure is for the Météo-France analysis, which is provided directly on the ECMWF grid.

A key aim of this project is to investigate the sensitivity of the forecast to analysis differences. During the next year, ECMWF will conduct an investigation of the performance of the system, including a comparison with the operational EPS. The following figures show first results from a preliminary evaluation of the performance of the multi-analysis system. Scores are calculated for 500 hPa height fields for 57 cases from July and August 2001 (some analysis data was missing for the remaining 5 dates). Figure 5 shows the average anomaly correlation over the Northern Hemisphere for each member of the MA system and for some of the corresponding operational forecast from the centre providing the analysis. There are clear differences in overall performance between the MA forecasts started from different analyses (Fig 5d). Over this set of cases the predictions from the DWD and Météo-France analyses have the lowest scores while the NCEP analyses provides the most skilful forecast. The performance of the MA-forecast from the DWD analysis is similar to that of the DWD operational forecast (Fig 5a; the DWD forecast extends to 7 days ahead) and the same correspondence can be seen for the forecasts from the NCEP (Fig 5b) and UK (Fig 5c) analyses (the Météo-France operational forecast is made only for 3 days ahead and is not shown).

Figure 6 compares the day-to-day variability in performance of each MA member with the corresponding operational forecast over Europe at day 3. Here again the similarity between forecasts from the same analysis (but using different models) is clear. By day 6 (Fig 7) there are cases where there are differences in skill between the MA-member and the operational forecast from the same analysis, but such examples are relatively infrequent.

These results suggest that differences in initial conditions may account for much of the difference in performance between medium-range forecast systems.

5. Severe weather prediction

The performance of the EPS depends on both resolution and membership. High resolution is needed to be able to describe in the most accurate way small-scale features and large membership is required to be able to sample the tail of the initial probability distribution function and to describe more correctly its time evolution. Previous works have shown that ensemble performance improves if resolution (*Buizza et al.* 1998, 2002) and membership (*Buizza and Palmer* 1998, *Mullen & Buizza* 2002) are increased. Computer power and the necessity to issue forecasts in a reasonable amount of time limit the operational resolution of the non-linear model and the ensemble membership. It should also be stressed that the optimal combination of membership and resolution is user dependent, with a small-size high-resolution ensemble being the preferred option for certain users and a large-size small-resolution ensemble the preferred option for others. At ECMWF, the configuration of the operational EPS run daily has been defined to try to satisfy a broad range of users by considering a large and comprehensive set of accuracy measures (*Talagrand et al* 1999).

One of the weaknesses of the current EPS is that it does not use the ECMWF analysis at its full accuracy. The EPS initial conditions are generated by interpolating the $T_L511L60$ analysis to the EPS $T_L255L40$ resolution, thus losing details that are important during the first few days of time integration.

A Variable Resolution EPS (VAREPS), whereby a $T_L511L40$ resolution (same as the as the analysis) is used until a forecast time TTRUNC followed by a $T_L255L40$ resolution, is under test. Such an approach of truncating the forecast resolution during the time integration has always been used at the National Centers for Environmental Prediction (NCEP, *Toth & Kalnay* 1997). This would move the interpolation from time 0 to time the TTRUNC when predictability is lost in the small-scales. In terms of CPU cost, a VAR2 forecast, with truncation applied at forecast day 2 (i.e. $T_L511L40$ up to day 2 and $T_L255L40$ from day 2 to day 10) needs approximately 2.6 time the CPU time required to run a $T_L511L40$ 10-d forecast.

Single deterministic experiments should be performed to assess the impact of applying a truncation at time TTRUNC at forecast day 2, 3 and 5 (VAR2, VAR3 and VAR5 experiments). These forecasts have been run with 40 vertical levels (L40), starting from the same $T_L511L60$ analysis interpolated to 40 vertical levels, and compared with the performance of constant-resolution $T_L255L40$ and $T_L511L40$ forecasts for 30 summer (August 1999) and 30 winter (December 1999) cases. The accuracy of the forecasts have been measured by computing the anomaly correlation and the root-mean-square error over the Northern Hemisphere and Europe for the prediction of the 500 hPa geopotential height field, and the root-mean-square error of the prediction of the 850hPa vector-wind over the Tropics.

Denote by SC(CONF,d) the seasonal average score 'sc' of the day-d forecast run in configuration CONF, and define the Relative Improvement index RI(CONF,d) as following:

$$RI(CONF,d) = \frac{SC(CONF,d) - SC(T_L255,d)}{SC(T_L255,d-1) - SC(T_L255,d)} \qquad \text{Eq. 2}$$

For each forecast-day d the index RI(CONF,d) gives the difference in skill of the day-d forecast given by configurations CONF and T_L255, normalized by the difference in skill of the day-d and the day-(d-1) T_L255 forecasts. In other words, RI(CONF,d) gives the gain in predictability compared to a 1-d gain in predictability at truncation T_L255.

Figure 8 shows the winter-average Relative Improvement index RI for configurations T_L511, VAR5, VAR3 and VAR2 at forecast day 3, 5 and 7. Results indicate that all forecast configurations have a higher skill than T_L255. Considering, for example, the anomaly correlation of the day-5 forecast for the 500hPa geopotential height (Fig. 8, middle panel, fourth group of bars) results indicate a 0.30 RI for all configurations. Note that for all forecast steps and all scores, the RI values are similar for all configurations. Figure 9 is the equivalent of Fig. 2 but for the 30-winter cases. Again, results are always positive for all configurations and depend only weakly on truncation. These results suggest that applying a truncation at forecast day 2 or 3 has a weak effect on the skill of a forecast. In other words, the benefit of running from a high-resolution analysis (T_L511) is maintained even if a forecast started at T_L511 resolution is truncated during the time integration. These results discussed above suggest that VAREPS could be a viable configuration.

6. Conclusions

Three changes in the ECMWF operational forecast productions are discussed. On 21st November 2000 the resolution of the forecast model in the operational European Centre for Medium-range Weather Forecasts (ECMWF) Ensemble Prediction System was increased from a 120-km truncation scale (EPS) to an 80-km truncation scale (High-resolution EPS, HEPS). On 1st September 2001 the multi-analysis stream based on 4 forecasts run with the ECMWF $T_L255L40$ version of the ECMWF model but starting from analyses produced at Deutsche Wetterdienst, the UK Meteorological Office, the National Centers for Environmental Predictions and from an average analysis was added to the ECMWF operational suite. On 21st January 2002 tropical perturbations defined by singular vectors targeted to maximize total-energy inside an area centred on tropical storms have been added to the EPS.

The increase on the EPS resolution and the inclusion of tropical perturbations has lead to an improvement in the EPS accuracy and capability to forecast tropical storms' tracks. Since implementation, the multi-analysis stream has been used to identify failures of the ECMWF forecasting system due to analysis problems. Results from pre-operational experimentation that lead to these three latest operational implementations have been summarized. On-going research projects aiming to develop a severe-weather warning system have also been discussed.

References

Barkmeijer, J, Buizza, R., & Palmer, T.N., 1999: 3D-Var Hessian singular vectors and their potential use in the ECMWF Ensemble Prediction System. *Q.J.R. Meteor. Soc.*, **125**, 2333-2351.

Barkmeijer, J., Buizza, R., Palmer, T.N., Puri, K., & Mahfouf, J.-F., 2001: Tropical singular vectors computed with linearised diabatic physics. *Q.J.R. Meteorol. Soc.*, **127**, 685-708.

Buizza, R., & Palmer, T.N., 1998: Impact of ensemble size on the skill and the potential skill of an ensemble prediction system. *Mon. Wea. Rev.*, **126**, 9, 2503-2518.

Buizza, R., Petroliagis, T., Palmer, T.N., Barkmeijer, J., Hamrud, M., Hollingsworth, A., Simmons, A., & Wedi, N., 1998: Impact of model resolution and ensemble size on the performance of an ensemble prediction system. *Q.J.R. Meteorol. Soc.*, **124**, 1935-1960.

Buizza, R., Miller, M., & Palmer, T.N., 1999: Stochastic simulation of model uncertainties. *Q.J.R. Meteorol. Soc.*, **125**, 2887-2908.

Buizza, R., Richardson, D.S., & Palmer, T.N., 2002: Global high-resolution (80-km) ensemble prediction. *Q.J.R. Meteorol. Soc.*, submitted.

Ehrendorfer, M., 1999: Prediction of the uncertainty of numerical weather forecasts: Problems and approaches. *Proceedings of the ECMWF Workshop on Predictability*, 20-22 October 1997, pp. 27-99.

Leith, C.E., 1974: Theoretical skill of Monte-Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409-418.

Mahfouf, J.-F., 1999: Influence of physical processes on the tangent-linear approximation. *Tellus*, **51A**, 147-166.

Molteni, F., Buizza, R., Palmer, T.N., & Petroliagis, T., 1996: The new ECMWF ensemble prediction system: methodology and validation. *Q.J.R. Meteorol. Soc.*, **122**, 73-119.

Mullen, S., & Buizza, R., 2002: The Impact of horizontal resolution and ensemble size on probabilistic forecasts of precipitation by the ECMWF Ensemble Prediction System. *Weather & Forecasting*, **17**, 173-191.

Puri, K., Barkmeijer, J. & Palmer, T.N., 1999: Ensemble prediction of tropical cyclones using targeted diabatic singular vectors. *Q.J.R. Meteor. Soc.*, **127**, 709-731.

Talagrand, O., Vautard, R., & Strauss, B., 1999: Evaluation of probabilistic prediction systems. *Proceedings of the ECMWF Workshop on Predictability*, ECMWF, Reading, 20-22 October 1997, 1-26.

Toth, Z., & Kalnay, E., 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297-3319.

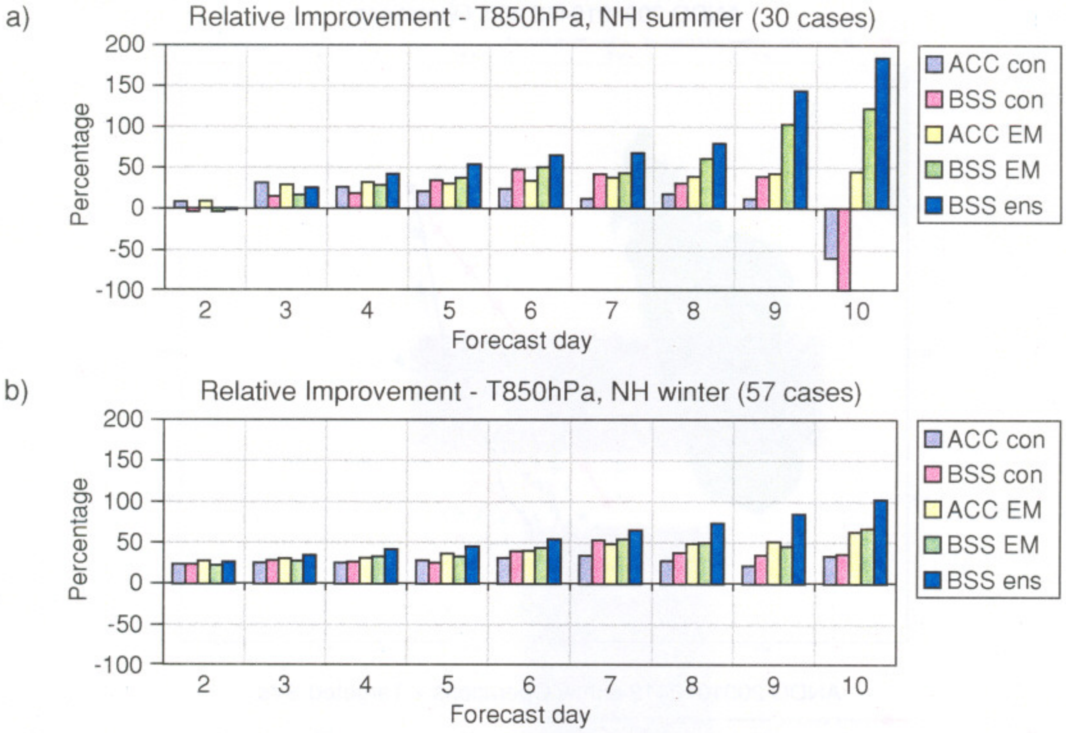

Fig 1. (a) Relative improvement index for summer computed over the Northern Hemisphere: control ACC skill score (mauve) the control Brier skill score (magenta), the ensemble-mean ACC skill score (yellow), the ensemble-mean Brier skill score (green) and the EPS Brier skill score (blue). A Relative Improvement of 100% indicate a gain in predictability of 1-day (see text for details). (b) as (a) but for winter.

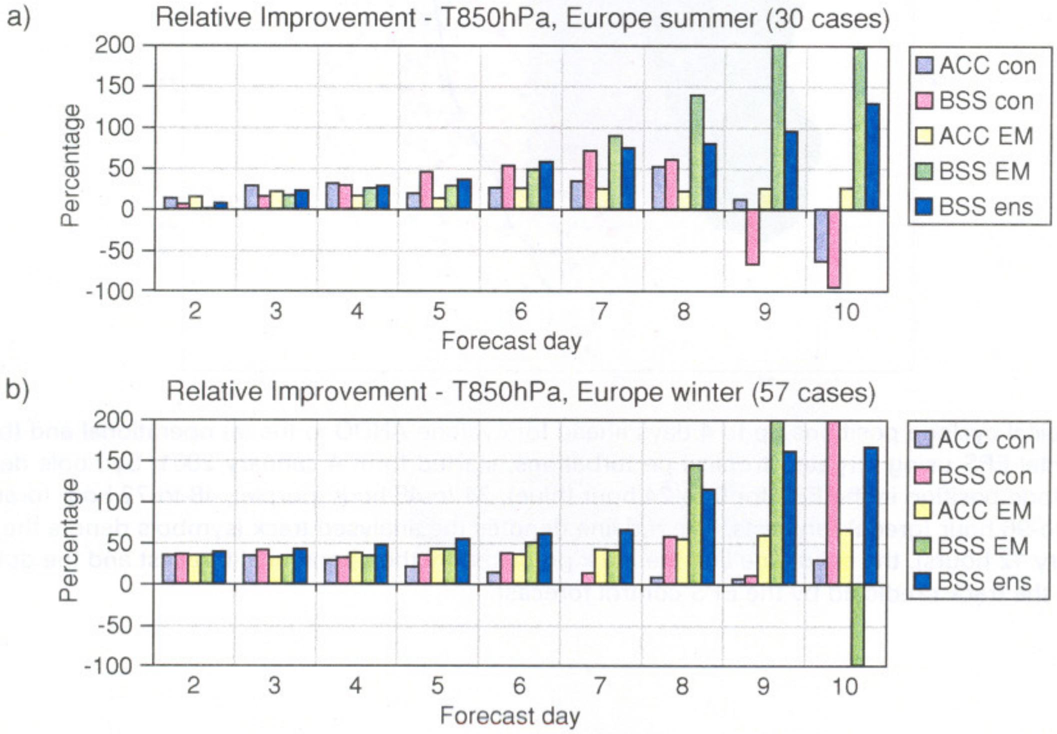

Fig 2. (a) Relative improvement index for summer computed over the Europe: control ACC skill score (mauve) the control Brier skill score (magenta), the ensemble-mean ACC skill score (yellow), the ensemble-mean Brier skill score (green) and the EPS Brier skill score (blue). A Relative Improvement of 100% indicate a gain in predictability of 1-day (see text for details). (b) as (a) but for winter.

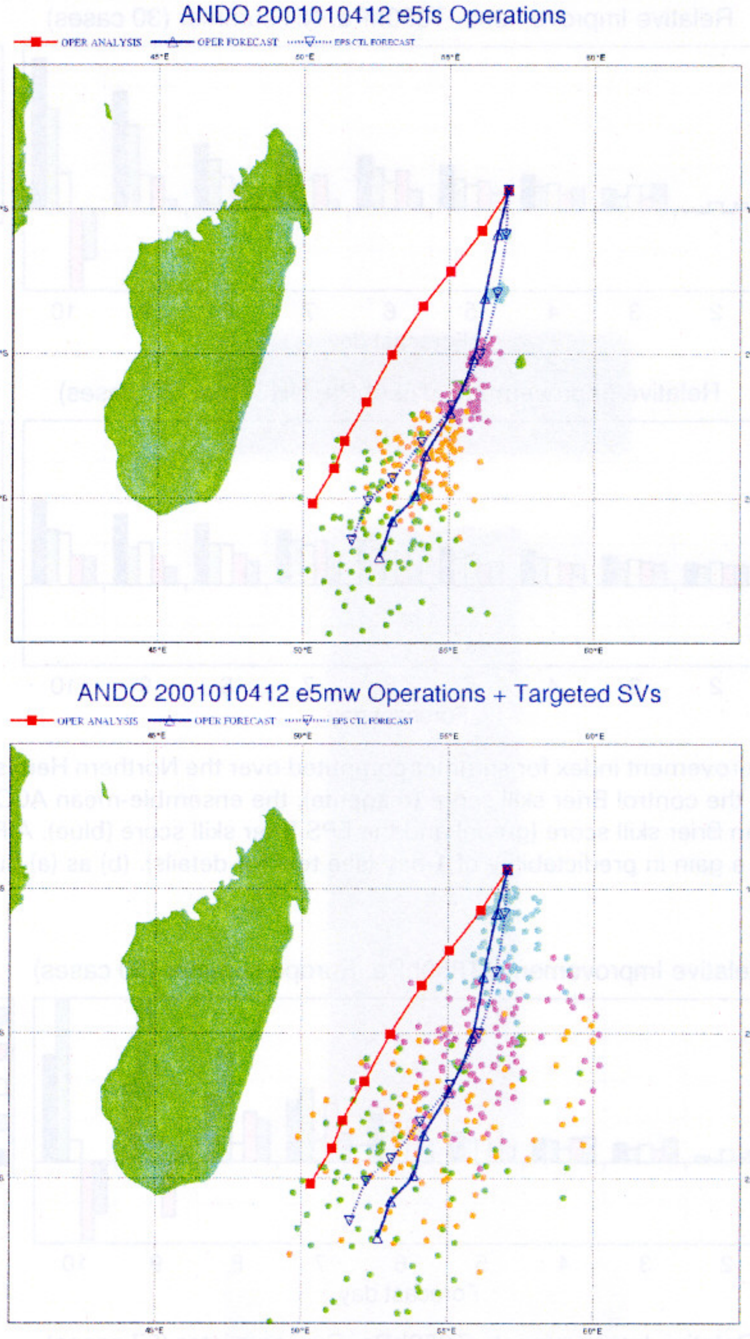

Fig 3. Tropical cyclone positions up to 4 days ahead for cyclone ANDO in the (a) operational and (b) experimental EPS using targeted tropical perturbations, started form 4 January 2001. Symbols denote the cyclone position in the EPS for 0-to-24 hour (blue), 24-to-48 hour (purple), 48-to-72 hour (orange) and 72-to-96 hour (green) forecasts. The red line denotes the analysed track (symbols denote the position every 12 hours), the solid blue line the track predicted by the operational forecast and the dotted blue line the track predicted by the EPS control forecast.

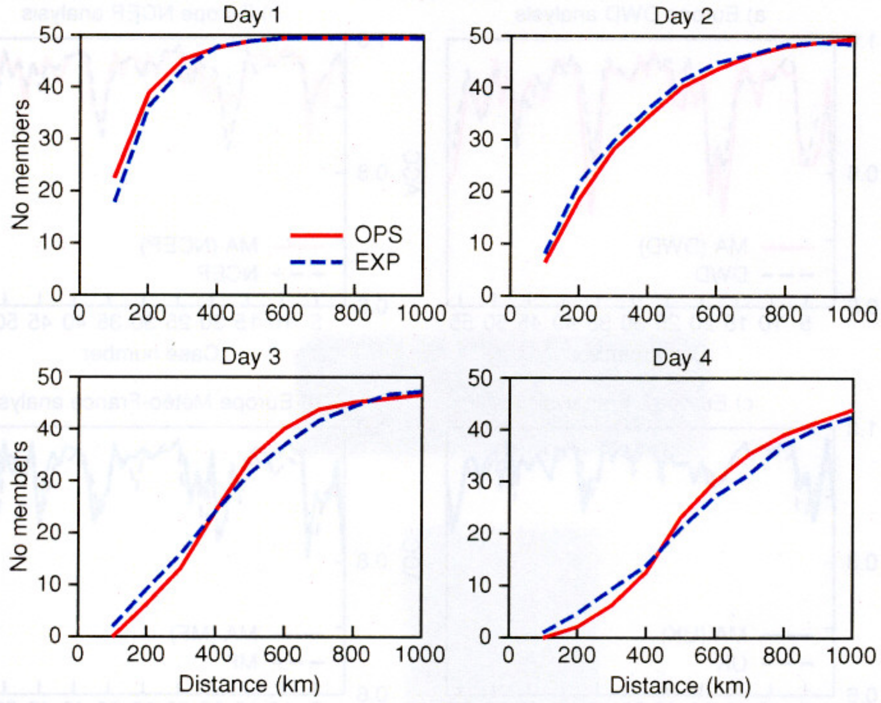

Fig 4. Number of ensemble members (averaged over 14 cases) closer to the analysed position of the tropical cyclone as a function of distance, for the operational EPS (solid red line) and the EPS with tropical perturbations (dashed blue line). The forecast period is given above each panel.

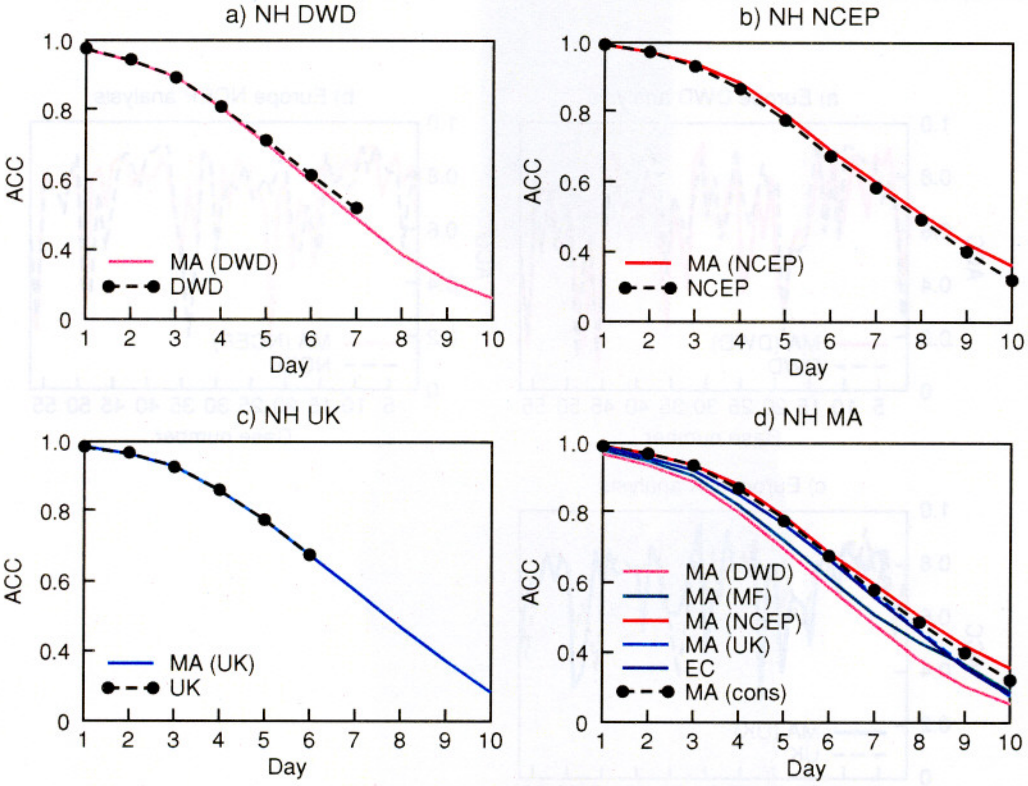

Fig 5. Mean anomaly correlation for forecast days 1 to 10 over 57 cases for 500 hPa height over the Northern Hemisphere. The multi-analysis forecasts (solid lines) are compared with the corresponding operational forecast for a) DWD, b) NCEP and c) UK. All four MA forecasts (including the forecast from the Météo-France analysis) are shown together with the EC control forecast in d)

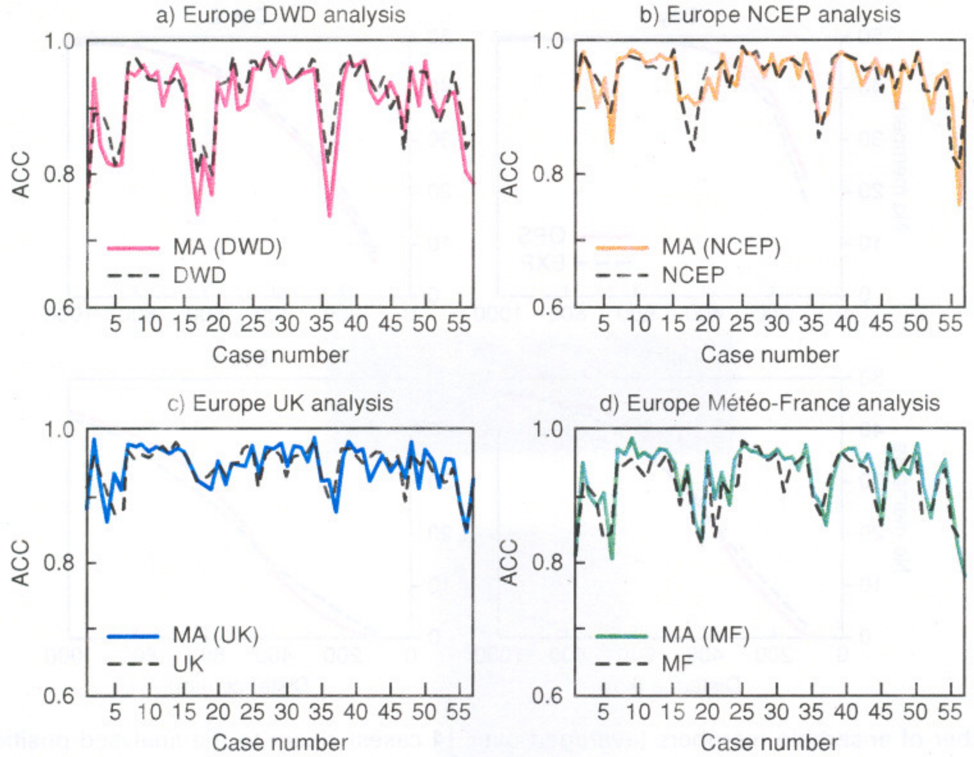

Figure 6. Time series of anomaly correlation at forecast day 3 (57 cases) for 500 hPa height over Europe. The multi-analysis forecasts (solid lines) are compared with the corresponding operational forecast for a) DWD, b) NCEP, c) UK and d) Météo-France.

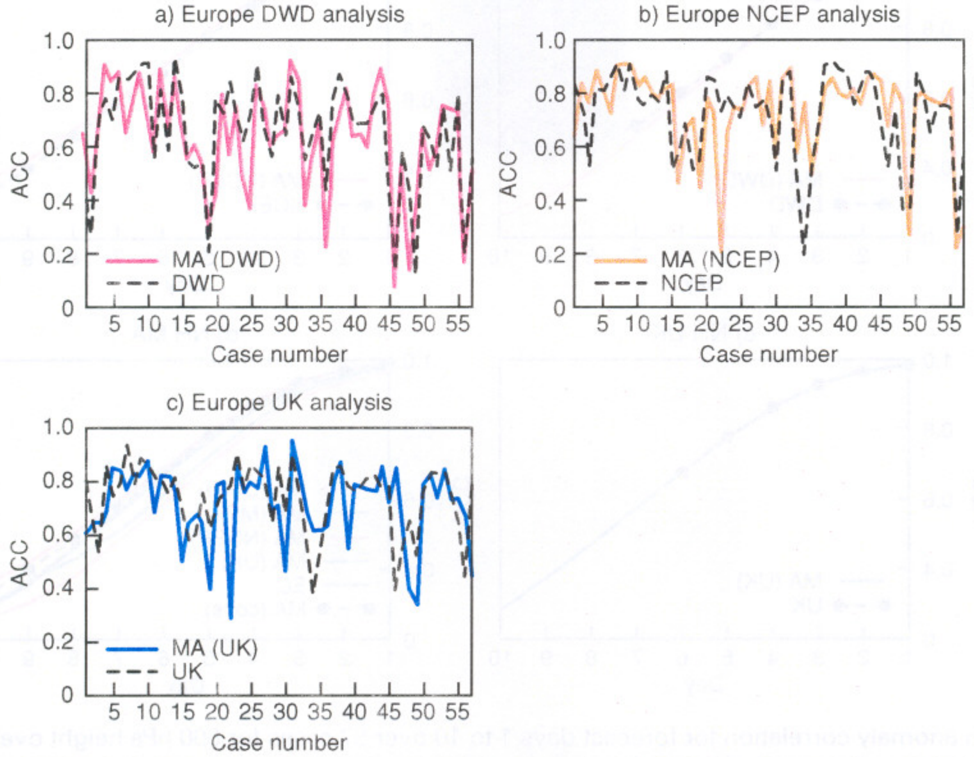

Figure 7. As Figure 6 but for forecast day 6 (Météo-France operational forecasts not available at day 6).

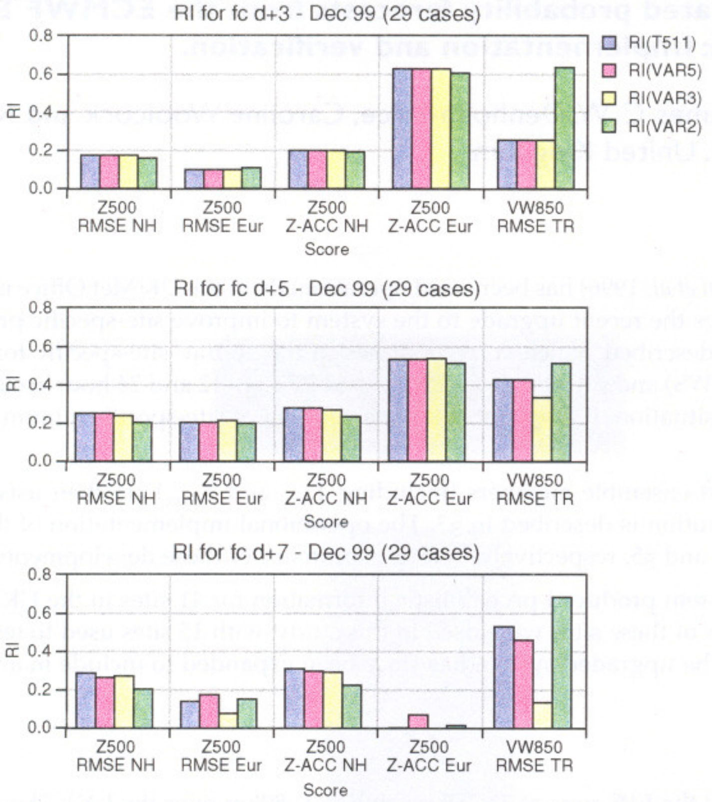

Fig 8. Winter-average relative improvement index RI at forecast day 3 (top), 5 (middle) and 7 (bottom) computed for 5 different scores: RMSE for the 500hPa geopotential height forecast over NH (first group of bars) and Europe (second group of bars), ACC of the 500hPa geopotential height forecast over NH (third group of bars) and Europe (fourth group of bars) and for the RMSE of the vector-wind over the Tropics (last group of bars). Colour bars refer to forecast in configuration TL511 (first bar, mauve), VAR5 (second bar, magenta), VAR3 (third bar, yellow) and VAR2 (fourth bar, green).

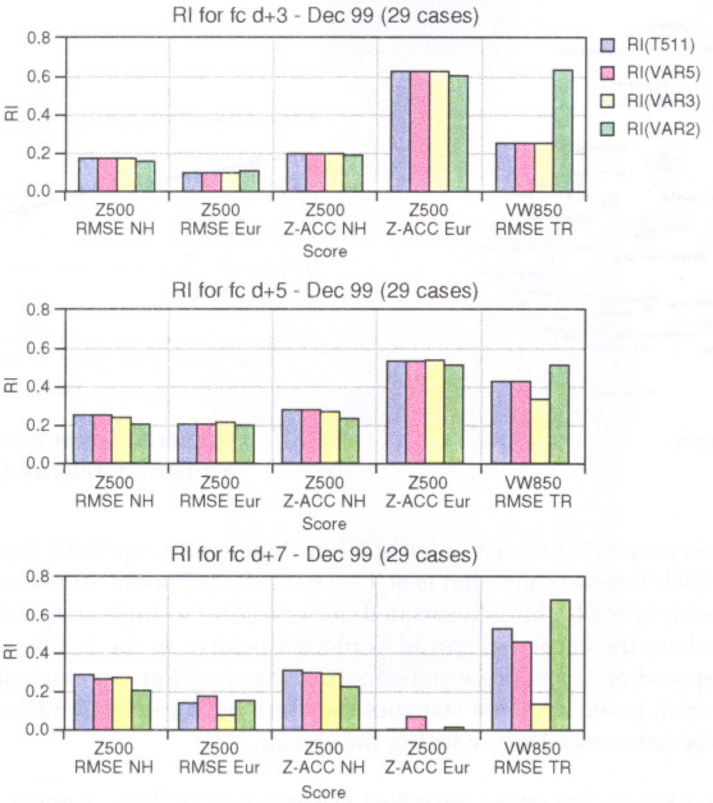

Fig 9. As Figure 8 but for summer-average (August 1999) results.