Ensemble assessment using the TIGGE database

Laurent Descamps, Carole Labadie, Alain Joly, Philippe Arbogast,
CNRM-GAME, Météo-France-CNRS, Toulouse

Ensemble forecasting has now become an important component of Numerical Weather Prediction. Several major meteorological centres have been running operational Ensemble Prediction Systems for years. Here, the purpose is to describe the Météo-France global ensemble forecasting system PEARP (Prévision d'Ensemble ARPEGE) designed for the short-range (from 72 to 108 h) probabilistic prediction and to present an evaluation of PEARP along with other ensembles using the TIGGE database (Bougeault et al., 2010).

The PEARP system

The operational version of PEARP is based on the ARPEGE model with an horizontal spectral truncation of T538 and a stretching factor of 2.4. The finer horizontal resolution is 15 km over France. There are 65 levels on the vertical up to a height of 50 km. The ensemble size is 35 members including a control member.

The initial perturbations of PEARP are built by combining the Météo-France ensemble data assimilation system AEARP (Berre et al., 2007) running at a coarser resolution (6 members, T399, no stretching) with singular vectors computed over different areas and with different optimization times and norms.

"Multi-physics" approach is used to represent model uncertainties. 10 different physical parametrization sets, including the ARPEGE operational physical package, have been chosen (Descamps et al, 2011).

Assessment of PEARP and other TIGGE ensembles

We present here an objective assessment of PEARP and four ensemble prediction systems: the UKMO ensemble prediction system (MOGREPS), the Canadian Meteorological Center (CMC) ensemble prediction system, the NCEP Global Ensemble Forecast System (GEFS) and the ECMWF ensemble prediction EPS. So far, most of the studies using the TIGGE database primarily have focused on medium-range prediction. The evaluation is here dedicated to short and early medium-range forecast.

Evaluation is provided for two synoptic variables, 500 hPa geopotential height, and 850 hPa temperature. Both variables are interpolated to a 1.5° latitude-longitude regular global grid using the interpolation routines provided by the TIGGE data portal (see the TIGGE portal at http://tigge.ecmwf.int) and run over a one-month period (September 2010).

A bootstrap resampling technique (Efron and Tibshirani (1993)) is applied to estimate confidence intervals (5%-95%) for the different scores. Our procedure is the same as the one used in Candille et al. (2007). We recompute the scores 10000 times with a sample of realizations randomly extracted, with replacement, from the original dataset.

A perfectly reliable ensemble and the observations are supposed to have the same climatology. In other words, the rank histogram, also known as the Talagrand diagram, is supposed to be flat. The delta score measures the departure from flatness. Fig 1 displays delta scores for the five global ensembles considered in this study. The results present common features in the sense that the reliability increases with lead-time for all the ensembles and for both variables considered here. For both variables CMC ensemble appears as the most reliable one, followed by PEARP except at very short ranges for the 500 hPa geopotential height.

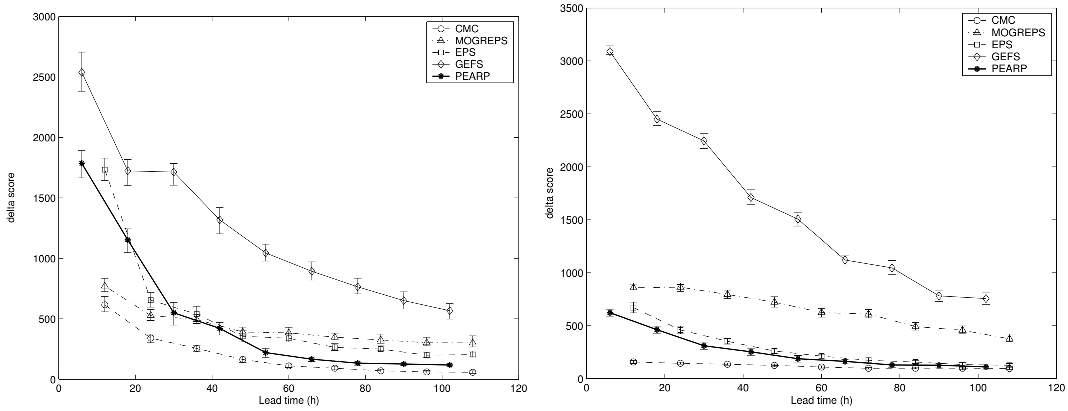

Figure 1: Global “delta” scores against radiosounding observations for a 1-month period for 500 hPa height (left) and 850 hPa temperature (right)

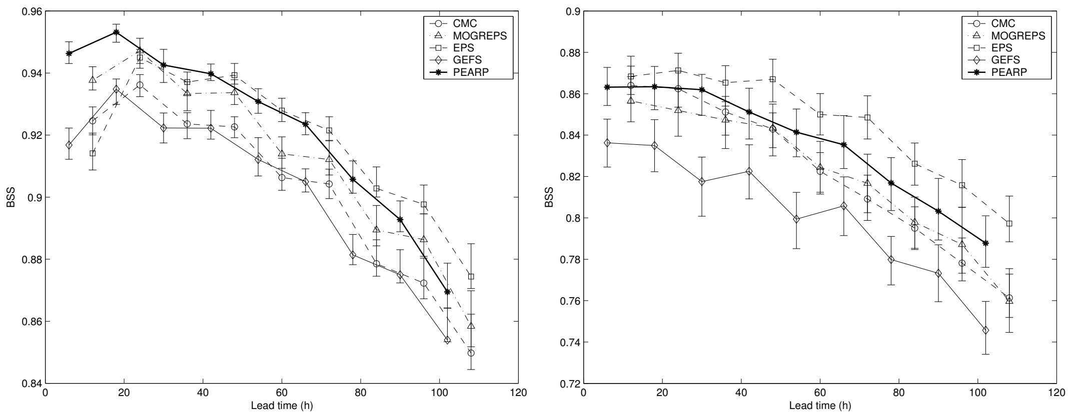

Figure 2: Global Brier Skill Scores scores against radiosounding observations for a 1-month period and for 500 hPa height (left) and 850 hPa temperature (right)

The Brier score measures a distance between the ensemble pdf and the observations. Although Brier score account for both the reliability and the resolution, it is dominated by the resolution term. A perfect ensemble has a Brier Skill Score (BSS) equal to 1. Fig 2 presents the BSS for the same ensembles and over the same period. Four among the five ensembles are very close with respect to the magnitude of the error bars. ECMWF ensemble seems to outperform the other ones except at very short ranges where PEARP behaves well.

References

Berre, L., O. Pannekoucke, G. Desroziers, S. E. Stefanescu, B. Chapnik, and L. Raynaud, 2007: A variational assimilation ensemble and the spatial filtering of its error covariances: increase of sample size by local spatial averaging. ECMWF Workshop on Flow-dependent aspects of data assimilation, Reading, 151–168.

Bougeault, P., et al., 2010: The Thorpex Interactive Grand Global Ensemble (TIGGE). Bull. Amer. Meteor. Soc., 91, 1059–1072.

Candille, G., C. Côté, P. Houtekamer, and G. Pellerin, 2007: Verification of an ensemble prediction system against observations. Mon. Wea. Rev., 135, 2668–2699.

Descamps, L., C. Labadie, and E. Bazile, 2011: Representing model uncertainty using the multiparametrization method. Proceedings of ECMWF Worshop on Model Uncertainty.

Efron, B. and R. Tibshirani, 1993: An Introduction to the Bootstrap. Chapman & Hall, 436pp.