Application and verification of ECMWF products 2010

METEO-FRANCE Bruno Lacroix

1. Summary of major highlights

- Implementation of a limited area model ALADIN configuration coupled with IFS over France at 00 and 12UTC
- Development of a new production of grid point total cloud cover forecast based on a statistical adaptation using satellite data as predictand

2. Use and application of products

2.1 Post-processing of model output

2.1.1 Statistical adaptation

Millions of local forecasts of weather parameters are produced daily through statistical adaptation of NWP output. Main methods are multiple linear regression (MLR) and linear discriminant analysis (DA). MOS (model output statistics) is generally preferred to PP (perfect prognosis). Kalman filter (KF) is applied when relevant. The production is described in table 1.

Note the new production of grid point total cloud cover forecast based on a statistical adaptation using satellite data as predictand.

Deterministic model T1279

Parameter	Method	Domain	Nbr of Sites	Steps
Tri-hourly 2m Temperature	MLR (MOS) +KF	France	2781	+12h to +180h by 3h
Daily extremes 2m temperature	MLR (MOS) +KF	France	2781	D to D+6
10m Wind Speed	MLR (MOS)	France	861	+12h to +180h by 3h
10m Wind Direction	MLR (MOS)	France	822	D to D+6
Total Cloud Cover	MLR (MOS)/LDA	France	164	+12h to +180h by 3h
Total Cloud Cover	LDA	France	GRID 0.5x0.5	0h to +156h by 3h
Tri-hourly 2m relative Humidty	MLR (MOS) +KF	France	1269	+12h to +180h by 3h
Daily extremes 2m rel. Humidity	MLR (MOS) +KF	France	1269	D to D+6
Tri-hourly 2m Temperature	MLR (MOS) +KF	World	6010	+12h to +180h by 3h
Daily extremes 2m temperature	MLR (MOS) +KF	World	6010	D to D+6

Table 1 : Statistical adaptations for the deterministic high resolution model

EPS

Statistical adaptation is applied to individual ensemble runs (table 2). Methods are the same as for the deterministic model output but pseudo-PP (statistical equations computed during the first 24 hours then applied to the other corresponding steps) is preferred to MOS. VAREPS is used and Météo-France provides local forecast (temperatures) up to 14 days.

EPS Ensemble mean and individual members

Parameter	Method	Domain	Nbr of Sites	Steps
Tri-hourly 2m Temperature	MLR (pPP) +KF	France	2761	+12h to +360h by 3h
Daily extremes 2m temperature	MLR (pPP) +KF	France	2761	D to D+14
10m Wind Speed	MLR (pPP)	France	792	+6h to +240h by 3h +246 to +360 by 6h
Tri-hourly 2m relative Humidty	MLR (MOS) +KF	France	1146	0h to +240h by 3h +246 to +360 by 6h
Daily extremes 2m rel. Humidity	MLR (MOS) +KF	France	1146	D to D+14
Tri-hourly 2m Temperature	MLR (pPP) +KF	World	3338	+12h to +360h by 3h
Daily extremes 2m temperature	MLR (pPP) +KF	World	3338	D to D+14

Table 2 : Statistical adaptations for the EPS

EPS Distribution

Calibration is applied to the EPS distribution in order to optimize reliability. Operationally, a calibration based on rank diagrams is used for 10m wind speed and total precipitations. Bayesian Model Averaging (BMA) calibration is under development and will be used for temperatures at the end of the year.

Monthly forecast

Statistical models are also applied to the monthly forecasts up to 32 days (table 3). These locally corrected forecasts allow to couple electricity consumption models.

Parameter	Method	Domain	Nbr of Sites	Steps
Tri-hourly 2m Temperature	MLR (pPP) +KF	France	1056	+0h to +768h by 3h
Daily extremes 2m temperature	MLR (pPP) +KF	France	1056	D to D+31

Table 3 : Statistical adaptations for the monthly forecasts

2.1.2 Physical adaptation

Limited area model

The Limited Area Model ALADIN is operated over western Europe (figure 1) by Météo-France as a dynamical adaptation coupled to the IFS forecasts. It is compared to its driving model against surface stations observations. The ALADIN-ECMWF has no assimilation cycle and corresponds to a dynamical adaptation of the ECMWF forecasts by a higher resolution model (7.5 km against 16 km) .

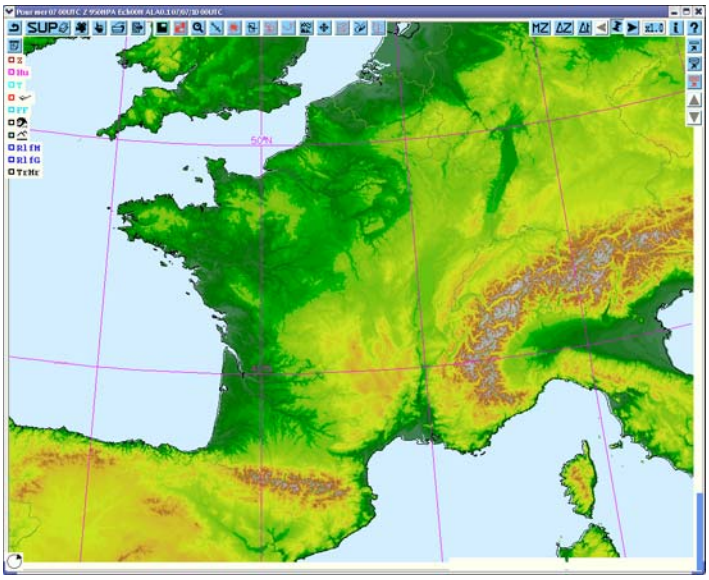


Figure1: orography map of ALADIN-ECMWF

The verification is performed by computing scores for the surface parameters measured by the 240 surface stations of the domain FRANCE extending from 38°N to 53°N and from 8°W to 12°E. Root mean squared errors and bias are computed for the different parameters and lead times and averaged over the entire simulation domain. The forecasted value of the parameter is taken at the nearest point from the observation. The errors are accumulated during one and a half month and scores are presented in function of the lead time. The ALADIN-FRANCE results are also added to help to explain the differences between ALADIN-ECMWF and the IFS model. For the wind at 10 m AGL (figure 2), the physics near the surface improves the scores (about 10 %) for all lead times during this period.

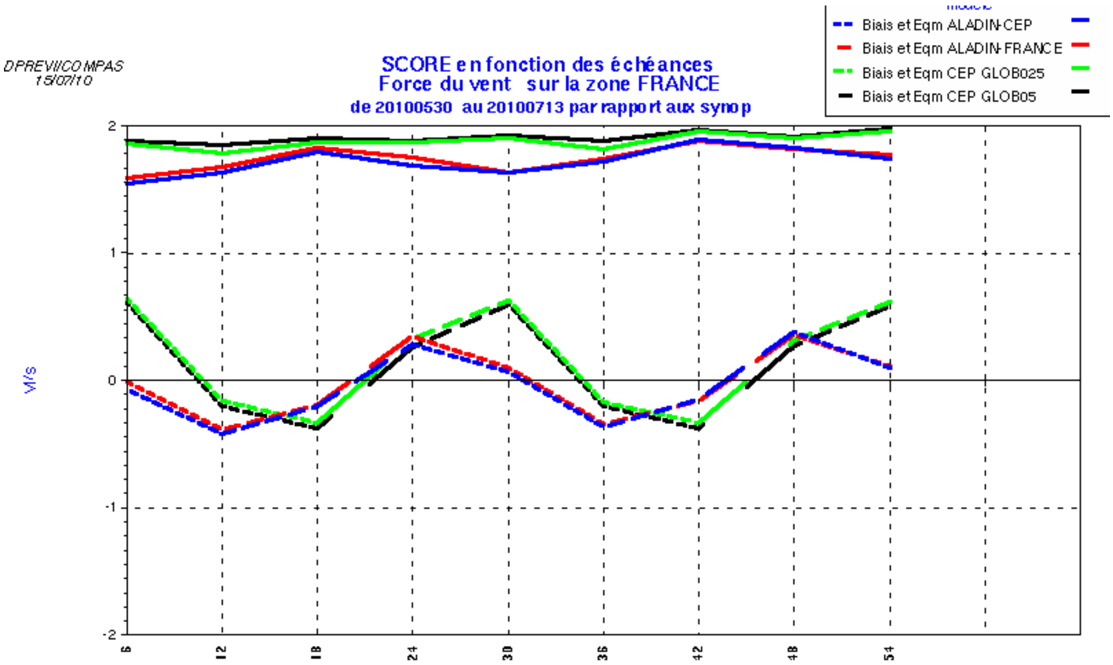


Figure 2: RMS (full lines) and bias (dotted lines) against the surface data observations in the domain FRANCE (see text for details) in m/s for the wind at 10 m AGL for ALADIN-ECMWF (bleu), ALADIN-FARNCE (red), IFS at resolution 0.5 ° (black) and 0.25° (green). The comparison is performed from 30/05/2010 to 13/07/2010. The errors are plotted in function of the lead time from 0 to 54 hours.

Pollutant transport and dispersion forecast

For the long-range dispersion forecast, Météo-France uses three operational tools to assess impacts in case of an accidental release:

- an air mass trajectories software, describing the evolution of a neutrally buoyant particle in the wind field forecasted by ARPEGE or IFS,

- a semi-lagrangian off-line dispersion model, MOCAGE-accident. MOCAGE-accident is a specific configuration of MOCAGE (cf. section on air quality forecast), which takes into account point sources and relevant sink processes. The possibility is given to the forecaster in charge to use ARPEGE or IFS forecasts as meteorological forcings. In its backward version, MOCAGE-accident can also be used for backtracking.

At local scale, the system PERLE focuses on the local description of the atmospheric pollutant cloud at local and regional scales, in the vicinity of the radionuclide or chemical release. It is based on meso-scale non hydrostatic models for meteorological fields, Meso-NH or AROME, coupled to a lagrangian particle model for the dispersion, LPDM (University of Colorado). In the case of Meso-NH, initialization and boundary conditions are provided by ALADIN, itself relying on ARPEGE or IFS fields. At present, PERLE is thus used only over the Metropolitan France in operations.

Developments in progress aim to give the possibility to use IFS forecasts for the initial and boundary conditions of Meso-NH in the system PERLE; it is expected to be part of the operational set-up in the end of 2010 or early 2011.

The interest of using the outputs of a meteorological ensemble prediction system as forcings for the transport and dispersion models is also under investigation. It could bring a valuable information about the uncertainties in the dispersion fields associated with uncertainties on the meteorology. In the framework of the European project PREVIEW, the 51 members of the EPS have been provided on the period of one release of the European Tracer EXperiment (1994) and used to feed MOCAGE-accident as well as other models involved in the project. A scientific publication has been submitted.

Air quality forecast

MOCAGE multi-scale Chemistry and Transport Model was developed at Météo-France for both research and operational applications in the field of environmental modelling. MOCAGE considers simultaneously the troposphere and stratosphere at the planetary scale. In addition, it is possible within MOCAGE to zoom down to the regional scale over limited-area sub-domains, the model providing its own time-dependent chemical boundary conditions. Depending upon applications, MOCAGE can run in both on-line, coupled to a general circulation model for climate studies for instance, or off-line modes, forced by meteorological analyses or forecasts. The off-line configuration can use ARPEGE, ALADIN, AROME or IFS operational Numerical Weather Prediction products.

In the context of the partnership consortium "Prév'Air" in charge of the pollution monitoring for France, the operational version of MOCAGE provides daily air quality (gaz and particles) forecasts, using ALADIN and ARPEGE forecasts. In parallel, air quality MOCAGE forecasts relying on IFS operational forecasts are also daily running in the context of the EU-funded MACC project; MOCAGE is one of the six pre-operational air quality suites. 0-72h forecasts are displayed on the project website maintained by Météo-France.

2.1.3 Derived fields

Including post-processing of EPS output e.g. clustering, probabilities

2.2 Use of products

Cyclone forecasts :

CXML file with RSMC position and calibrated EPS accuracy is produced and vizualized (see Figure 3) thanks to Synergie Cyclone (public product will be issued soon) :

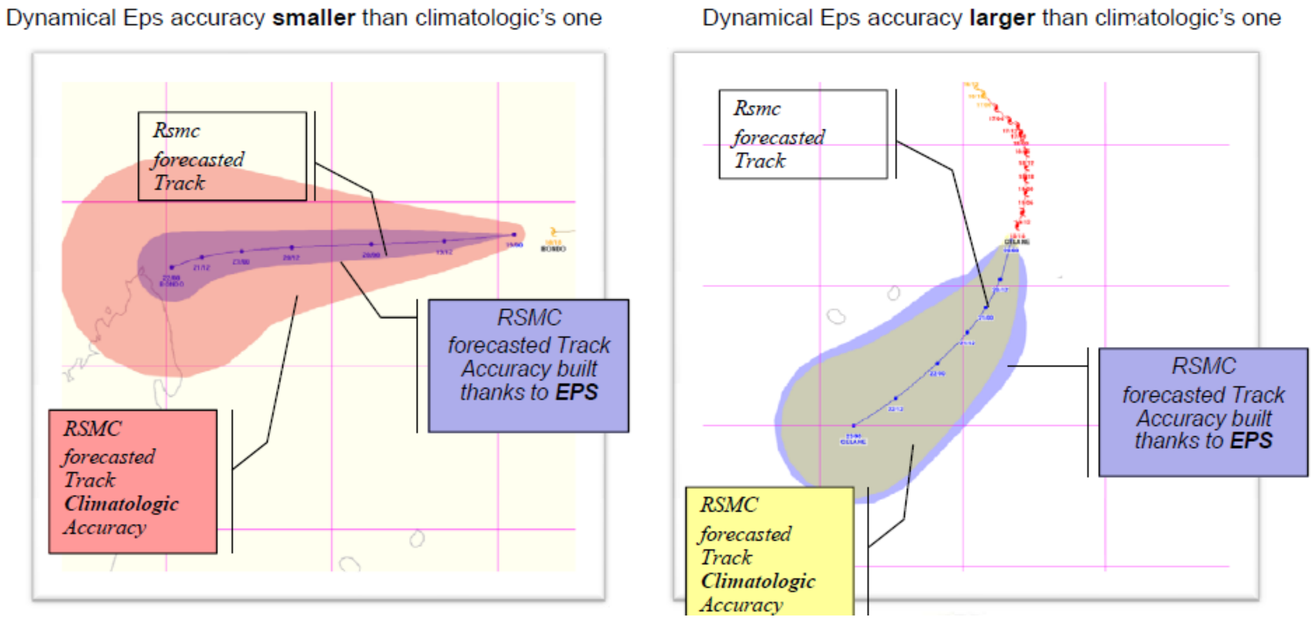


Figure 3 : forecasted track and accuracy

Strike probabilities and individual trajectories from EPS are also used in cyclone tracking (Figure 4) :

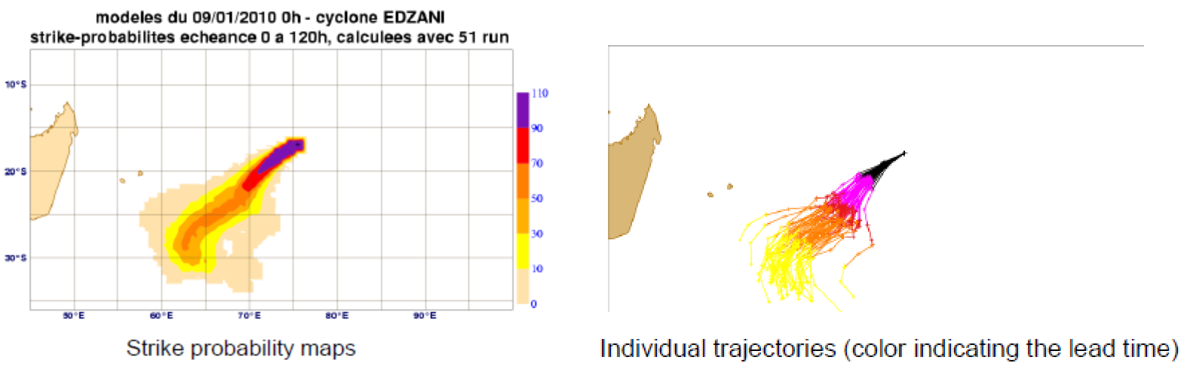


Figure 4 : EPS Strike probability (left) and Individual trajectories (right) for cyclone EDZANI

3. Verification of products

3.1 Objective verification

3.1.1 Direct ECMWF model output (both deterministic and EPS)

3.1.2 ECMWF model output compared to other NWP models

see annexe 1 for a full article on comparison between EPS and ARPEGE ensemble system

Cyclone forecasts :

The ECMWF model is used for cyclone tracking. A comparison has been done for different models in terms of direct position error for 3-days forecasts. The figure 5 below shows the results of this comparison for the season 2009-2010 over the South West Indian Ocean and for different models : ECMWF, UKMO, a non-stretched version of ARPEGE, ALADIN, GFS, NOGAPS, AVN and the american consensus forecast. Numbers of cases are indicated. The ECMWF (red curve) model presents the best skill at each step and confirms its leadership in cyclone tracking.

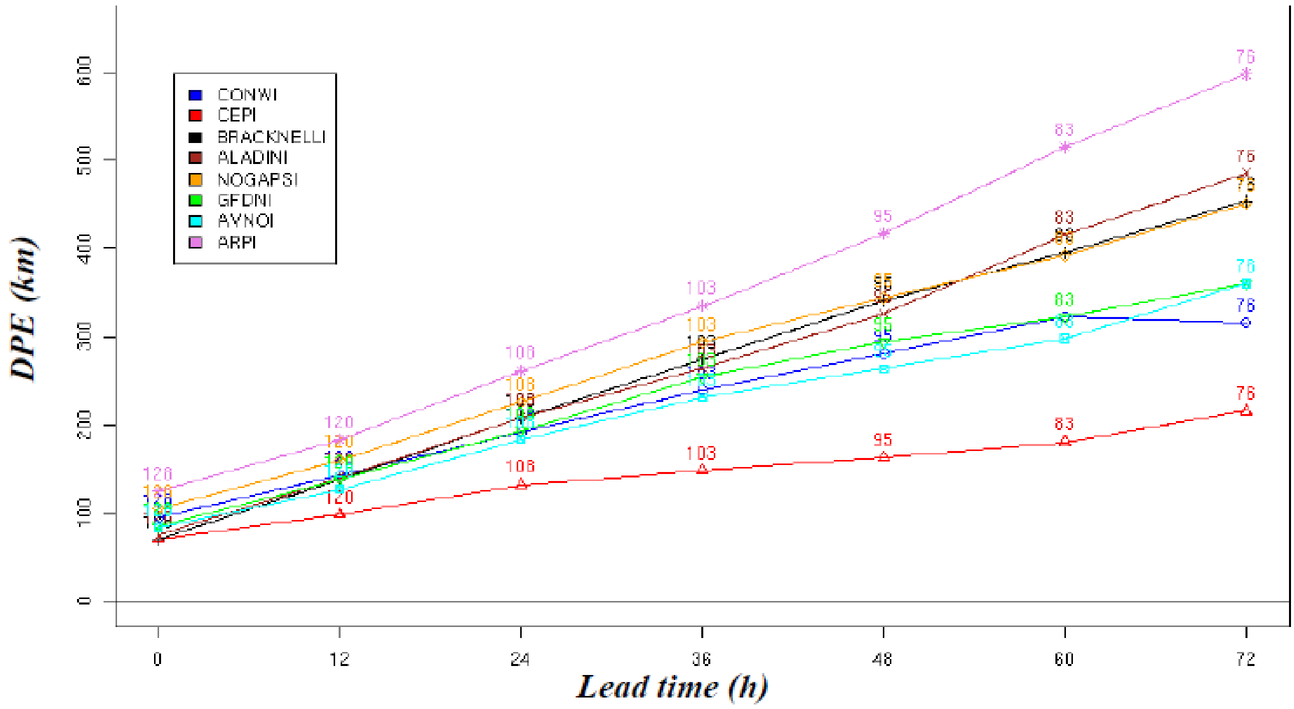


Figure 5 : direct position error for the season 2009-2010 over the South West Indian Ocean

3.1.3 Post-processed products

3.1.4 End products delivered to users

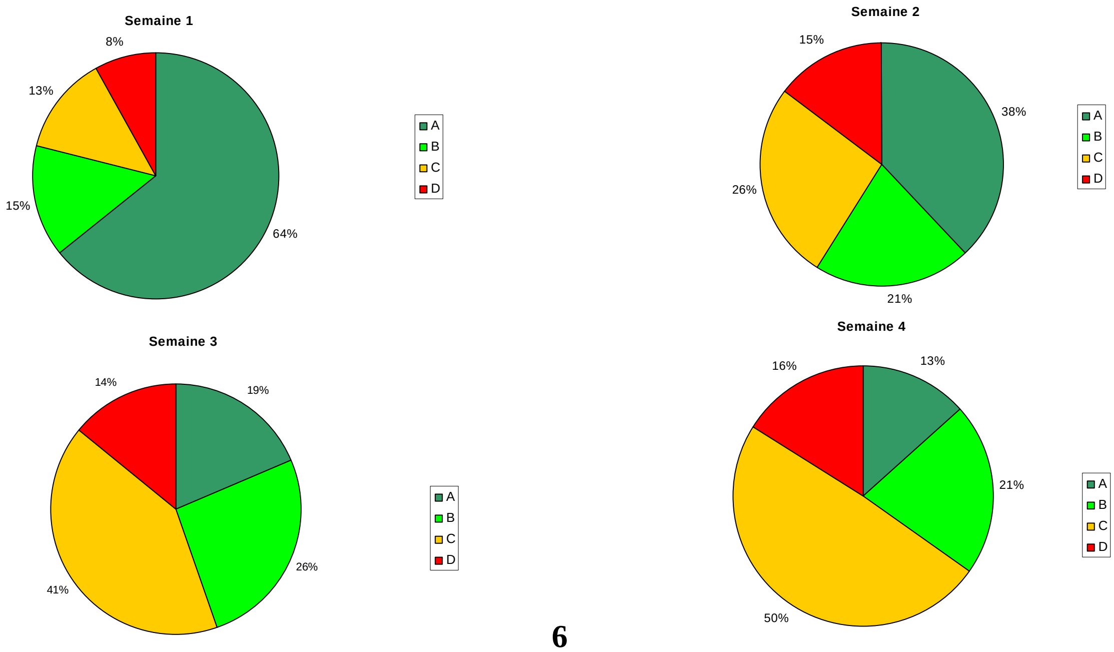
3.2 Subjective verification

3.2.1 Subjective scores (including evaluation of confidence indices when available)

Monthly forecast verification

The monthly forecasts of 2m-temperature anomalies have been assessed by the forecasters since November 2004. From this time to april 2010, a sample of 284 elements is available.
For every week, the marks vary from A to D with the following signification :
A : good localisation and intensity of the anomaly
B : slight differences (localisation and/or intensity) between observed and forecast anomaly
C : forecast anomaly and nothing observed, or (more frequently) observed anomaly and nothing forecast
D : observed anomaly is opposite to the forecast one.

The figure 6 plots the proportion of each mark for week 1 to week 4, over the period November 2004-april 2010.



 France

Figure 6 : Repartition of subjective notations for monthly T2m anomaly over France (sample size=284).

The forecasts are very good for week 1 and good for week 2. For the week 3 and 4, there are more bad forecasts than good ones. This is mainly due to the important number of C marks, which often correspond to the case where there is no signal in the forecast and an observed anomaly. If we put out these cases, the number of good forecasts becomes around 70% for week 3 and 55% for week 4.
Note that the proportion of bad forecasts (D marks) is very similar from week 2 to week 4.
The evolution of the proportion of each mark has been plotted on figure 7, with a 12 months running average. No clear tendancy appears over the period for week 1 to 3, but there is a slight improvement for week 4.

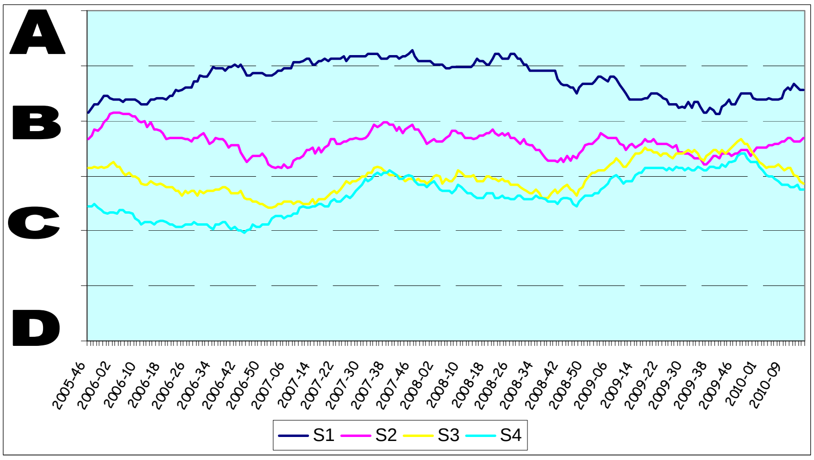


Figure 7 : Proportion of each mark, 12 months running average.

3.2.2 Synoptic studies

4. References to relevant publications

annexe 1: Comparison of ECMWF EPS and PEARP over Europe

Laurent Descamps(1), Carole Labadie(1), Eric Bazile (1), Francis Pouponneau (2), Nicole Girardot (2) and Joël Stein(2)

Météo-France
(1) CNRM/GMAP/RECYF
(2) DPREVI/COMPAS/COM
42, Avenue G. Coriolis
31057 Toulouse Cedex France
E-mail : joel.stein@meteo.fr

Abstract : The probabilistic forecasts of two different ensemble forecasting systems (EFS), the “Prévisions d’Ensemble ARPEGE” (PEARP) and the Ensemble Prediction System (EPS) of ECMWF are compared using different probabilistic scores. They are computed over Europe for both ensembles separately and a direct comparison is performed for scalar scores. A bootstrap method is used in order to check if the observed differences are significant.

1 Introduction

This paper presents the comparison of two Ensemble Forecasting Systems (EFS), the “Prévisions d’Ensemble ARPEGE” (PEARP) of Météo-France and the Ensemble Prediction System (EPS) of ECMWF (Leutbecher and Palmer 2008). The PEARP system is based on the ARPEGE model and is operational since June 2004 (Descamps *et al* 2009). It has a variable horizontal resolution with a maximum of 23km over France (see Figure 1). Important upgrades have been implemented at the end of 2009. Ensemble size has been increased from 11 to 35 members. A new initialization procedure, mixing singular vectors and perturbations coming from an Ensemble Data Assimilation System (AEARP, Assimilation d’Ensemble ARPEGE (Berre *et al 2010)*), has been implemented for as 3DVAR FGAT assimilations. A ‘Multi-physics’ approach is used in order to take into account ‘model error’. A change from 3DVAR to 4DVAR for AEARP occurred in April 2010.

Verification procedure

The different fields derived from the members of the EFS are interpolated on a regular latitude-longitude grid of 0.5 degree. The geographical domain, named EURAT5, is centered over France and extends from latitude 20°N to 72°N and from longitude 32°W to 42°E (see Figure A1). The PEARP horizontal resolution over this domain varies between 23 and 30 km (Figure 1). The interpolation scheme is based on a spline function which uses the 12 nearest points of the computational grid for the PEARP. For ECMWF EPS, the fields are computed by the ECMWF procedures on a global regular latitude-longitude grid of 0.5° of resolution and used as input of the verification package.

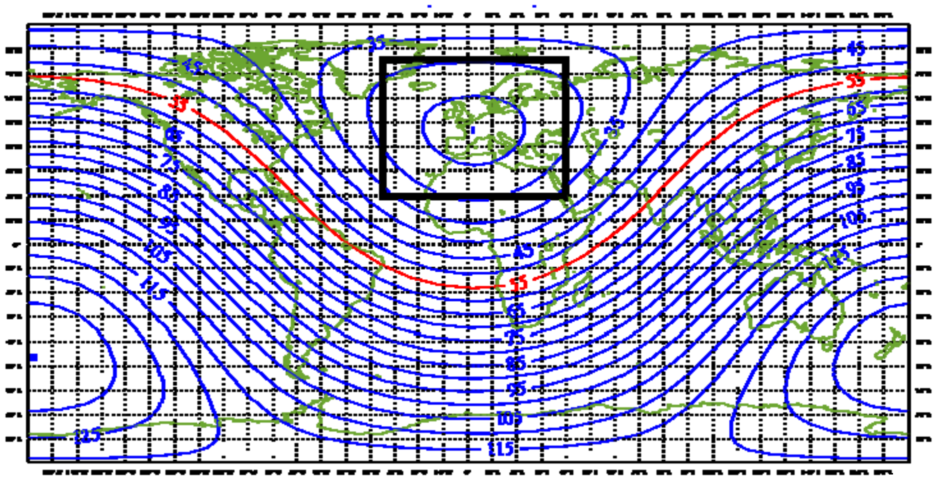


Figure A1 : Horizontal resolution of PEARP (km, contour every 5 km). The numerical equator (55 km) is in red. The black rectangle corresponds to the verification domain EURAT5 used in this study.

A set of scores is computed from the individual outputs of both ensembles. The scores are partially computed every day and stored in a data base in a form which allows accumulating a temporal sample of a sufficient size, in general three months. The following scores are presented for a given 3 months period:

- Fiability diagram for given validity hour and event
- ROC diagram for given validity hour and event
- Delta score as a function of the lead time for a given event
- Rank histograms for given validity hour and event.

A second set of plots collects the temporal series of scalar measures of performance still evaluated for a three months period. These scores are :

- Brier skill scores against the climatological forecasts deduced from the ERA-INTERIM reanalysis or from the temporal sample
- The decomposition of the Brier skill score against the climatological forecast deduced from the temporal sample
- Brier skill scores against the climatological forecasts deduced from the temporal sample with the confidence interval at the 10 % level
- ROC area (ROCA) with the confidence interval at the 10 % level
- Delta score with the confidence interval at the 10 % level

A last temporal evolution plots at the time the monthly dispersion of the ensemble (standard deviation of the ensemble around its mean) and the monthly root mean squared error of the mean of the ensemble.

The confidence intervals are computed by the following bootstrap algorithm:

1. generate a random temporal series of dates chosen with replacement among the 90 dates of the temporal sample.
2. for this list of dates, compute the score
3. iterate steps 1 and 2, 1000 times

4. rank the 1000 scores
5. store the 50th and the 950th value of the score

All these plots are available for PEARP and EPS and can be compared for any trimester ending after December 2008.
Supplementary verification plots are computed to compare more easily the EPS and PEARP. Only 3 scalar scores: Brier skill score against the sample climatological forecast, the ROC area and the Delta score are computed. The plots are drawn either as a function of the lead time for a given trimester and event or as a temporal serie for a given lead time and event. The significativity of the difference between the scores of both ensemble is tested by a bootstrap algorithm which looks like the previous one but with the following differences:

- in step 2, compute the score for model 1 and model 2 and store the difference of scores
- in step 4 and 5, replace the scores by the differences of scores
- add step 6 which compare 0 with the quantiles Q5 and Q95. If 0 is between Q5 and Q95, the difference is not significant at the 10 % level and significant otherwise.

A few examples

In this part, a few examples of scores computed using the procedure described in part 2 are shown. The 18h UTC run of PEARP and the 12h UTC run of ECMWF EPS are compared (same validity time). For figures that present score evolution against lead-time, the indicated lead-time is the PEARP one. ECMWF lead-time has a 6h positive shift from the PEARP one(for example 0h lead-time for PEARP is 6h lead-time for ECMWF EPS).

Figure A2 shows the time evolution of ROCA score (from September 2008 to May 2010) for the event ‘850 hPa temperature anomaly is greater than 1 times the climatological standard deviation (std)’ and PEARP 36h lead-time (42h lead-time for EPS). It can be seen that the December upgrade of PEARP (described in section 1) had a strong positive impact with results close to ECMWF EPS. The same positive impact is measured for all scores and all variables (not shown). The last upgrade in April leads to damage of this score relative to EPS but both scores remain very close. This probably indicates that further tuning of the amplitude of the initial perturbations coming from AEARP is required for the 4DVAR configuration.

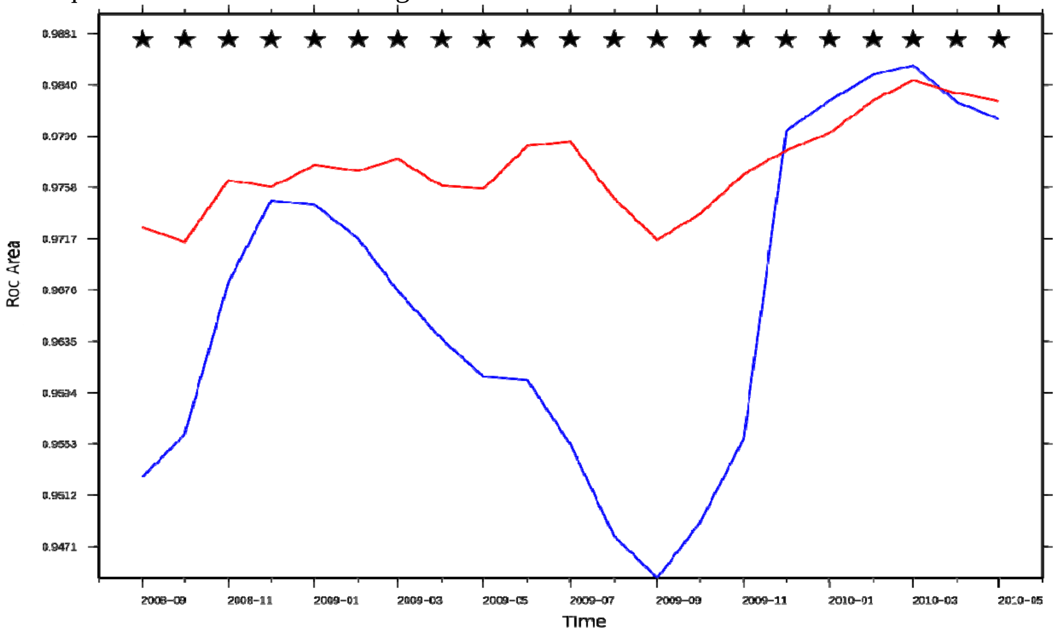


Figure A2 : Time evolution of the Area under the ROC curve for the event ‘850 hPa temperature anomaly is greater than 1 times the climatological std’. PEARP result (blue curve) is for 36h lead-time and ECMWF EPS result (red curve) is for 42h lead-time. Black stars indicate that the observed differences are significant at the10% level.

Figure A3 shows the Brier Skill Score for the event ‘Wind at 10 m AGL is greater than 5 m/s’. The PEARP shows better results that ECMWF EPS. For this parameter and for most of the scores and lead-times, PEARP has better results than ECMWF EPS. Nevertheless, the quality of the forecast remains poor because only very small and even negative values are reached by both ensembles when surface parameters are verified against real observations. The different upgrades of PEARP only lead to minor changes for this score.

 France

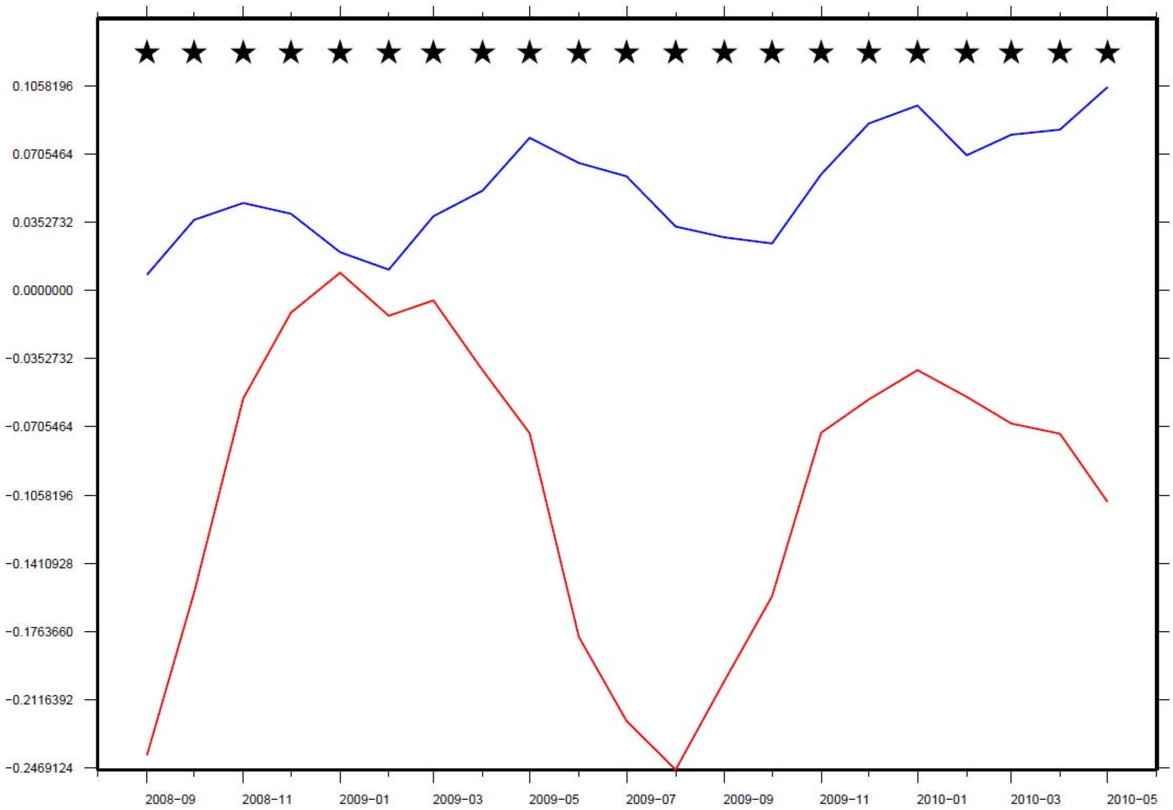


Figure A3 : Evolution of the Brier Skill Score for the event ‘wind at 10 m AGL is greater than 5 m/s’ for PEARP (blue curve) and ECMWF EPS (red curve) for the domain EURAT5. The score is computed over a 3-months period and the observations come from the surface stations. Black stars indicate that the observed differences are significant at the 10% level.

Conclusion

A comparison of the two EFS of Météo-France and ECMWF is performed every month at Météo-France in order to document the weakness and the benefits of the different ensembles. Results show that the upgrade implemented at the end of 2009 have a strong positive impact on PEARP scores. Generally speaking, PEARP has scores close to those of ECMWF EPS but the comparison needs to be carried on to get more stable conclusions. Further results will be therefore presented next year with a more representative sample of weather regimes for theses ensemble and including new scores (CRPS, etc..).

Berre L and Desroziers G. 2010 : Filtering of background error variances and correlations by local spatial averaging. Monthly Weather Review. to appear.

Descamps L. , Labadie C., Joly A., Bazile E. and Nicolau J. 2009: PEARP, the Météo-France Ensemble Prediction System. WWRP/Third THORPEX International Science Symposium. Monterey, USA

Leutbecher, M., Palmer, T.N.P., 2008 : Ensemble Forecasting. , J. Comp. Phys. , 227, 3515-3539.