



## **Evaluation of probabilistic quality and value of the ENSEMBLES multi-model seasonal forecasts: comparison with DEMETER**

Andrea Alessandri (1), Andrea Borrelli (2), Antonio Navarra (2), Alberto Arribas (3), Michel Déqué (4), Philippe Rogel (5), and Antje Weisheimer (6)

(1) Agenzia nazionale per le nuove tecnologie, l'energia e lo sviluppo economico sostenibile (ENEA), C.R. Casaccia, UTMEA-CLIM, Rome, Italy (a.alessandri73@gmail.com), (2) Centro Euro-Mediterraneo per i Cambiamenti Climatici, Bologna, Italy, (3) Met Office, Exeter, UK, (4) Météo France, CNRS/GAME, Toulouse, France, (5) CERFACS/URA1875, Toulouse, France, (6) European Centre for Medium-Range Weather Forecasts, Reading, UK

The performance of the new multi-model seasonal prediction system developed in the framework of the ENSEMBLES EU project is compared with the results from the previous project DEMETER. The comparison is carried out over the five Seasonal Prediction Systems (SPSs) that participated in both projects. Since DEMETER the contributing SPSs have improved in all aspects with the main advancements including the increase in resolution, the better representation of sub-grid physical processes, land, sea-ice and greenhouse gas boundary forcing and the more widespread use of assimilation for ocean initialization.

The ENSEMBLES results show an overall enhancement for the prediction of anomalous surface temperature conditions. However, the improvement shows considerable space-time variations and often tends to be quite small. In the tropics, ENSEMBLES improves systematically the sharpness and the discrimination attributes of the forecasts. Enhancements of the ENSEMBLES resolution attribute are also reported in the tropics for the forecasts started February 1<sup>st</sup>, May 1<sup>st</sup> and November 1<sup>st</sup>. Our results indicate that, in ENSEMBLES, an increased portion of prediction signal from the single-models effectively contributes to amplify the multi-model forecasts skill. On the other hand, a worsening is shown for the multi-model calibration over the tropics compared to DEMETER.

Significant changes are also shown in northern middle latitudes, where the ENSEMBLES multi-model discrimination, resolution and reliability improve for February, May and November starting dates. However, the ENSEMBLES multi-model decreases the capability to amplify the performance with respect to the contributing single-models for the forecasts started in February, May and August. This is at least partly due to the reduced over-confidence of the ENSEMBLES single-models with respect to the DEMETER counterparts.

Provided that they are suitably calibrated beforehand, it is shown that the ENSEMBLES multi-model forecasts represent a step forward for the potential economical value they can supply. A warning for all potential users concerns the need for calibration due to the degraded tropical reliability compared to DEMETER. Besides, the superiority of re-calibrating the ENSEMBLES predictions through the discrimination information is shown.

Concerning the forecasts started in August, ENSEMBLES exhibits mixed results over both tropics and northern middle latitudes. In this case, the increased potential predictability compared to DEMETER appears to be balanced by the reduction in the independence of the SPSs contributing to ENSEMBLES. Consequently, for the August start dates no clear advantage of using one multi-model system instead of the other can be evidenced.