Project no. GOCE-CT-2003-505539

Project acronym: ENSEMBLES

Project title: ENSEMBLE-based Predictions of Climate Changes and their Impacts

Instrument: Integrated Project

Thematic Priority: Global Change and Ecosystems

Deliverable D5.24: Scientific publication on the mechanics of multi-model combination

Due date of deliverable: February 2009
Actual submission date: March 2009

Start date of project: 1 September 2004 Duration: 60 Months

Organisation name of lead contractor for this deliverable
Federal Office of Meteorology and Climatology (MeteoSwiss)

Final version 5 March 2009

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	PU
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the Consortium (including the Commission Services)	

A scientific publication on the mechanics of multi-model combination has been submitted to and published in the Quarterly Journal of the Royal Meteorological Society.

Reference:

A. P. Weigel, M. A. Liniger and C. Appenzeller, 2008: Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts? *Quart. J. Roy. Meteor. Soc.*, **134**, 241-260.

Abstract:

The success of multi-model ensemble combination has been demonstrated in many studies. Given that a multi-model contains information from all participating models, including the less skilful ones, the question remains as to why, and under what conditions, a multi-model can outperform the best participating single model. It is the aim of this paper to resolve this apparent paradox.

The study is based on a synthetic forecast generator, allowing the generation of perfectly-calibrated single-model ensembles of any size and skill. Additionally, the degree of ensemble under-dispersion (or overconfidence) can be prescribed. Multi-model ensembles are then constructed from both weighted and unweighted averages of these single-model ensembles.

Applying this toy model, we carry out systematic model-combination experiments. We evaluate how multi-model performance depends on the skill and overconfidence of the participating single models. It turns out that multi-model ensembles can indeed locally outperform a best-model approach, but only if the single-model ensembles are overconfident. The reason is that multi-model combination reduces overconfidence, i.e. ensemble spread is widened while average ensemble-mean error is reduced. This implies a net gain in prediction skill, because probabilistic skill scores penalize overconfidence. Under these conditions, even the addition of an objectively-poor model can improve multi-model skill. It seems that simple ensemble inflation methods cannot yield the same skill improvement.

Using seasonal near-surface temperature forecasts from the DEMETER dataset, we show that the conclusions drawn from the toy-model experiments hold equally in a real multi-model ensemble prediction system.