Geophysical Research Abstracts,
Vol. 10, EGU2008-A-03527, 2008
SRef-ID: 1607-7962/gra/EGU2008-A-03527
EGU General Assembly 2008

Weather and seasonal ensemble forecasts: Are multi-models better than recalibrated single models?

A. P. Weigel, **M. A. Liniger** and C. Appenzeller
Federal Office of Meteorology and Climatology MeteoSwiss, Zurich, Switzerland

Multi-model ensemble combination (MMEC) has become an accepted technique to improve probabilistic forecasts at short to long-range time scales. The success of MMEC has been demonstrated in many studies. MMEC techniques typically widen ensemble spread, thus improving the dispersion characteristics and the reliability of the forecasts (which often tend to be overconfident and hence unreliable). This raises the question as to whether the same effect could be achieved in a cheaper way by simply inflating single model ensemble forecasts such that they become reliable. In this study the "climate conserving recalibration" (CCR) technique will be introduced and compared to MMEC techniques.

With a simple conceptual toy model we show that both CCR and MMEC successfully improve forecast reliability. The difference between these two methods is that CCR inevitably "dilutes" the potentially predictive signal, while with MMEC the signal is fully retained - at least if an "ideal" multi-model setup is considered (i.e. if an infinite number of models is combined and if the model errors are independent). Therefore, from a conceptual point of view, MMEC is to be preferred.

Unfortunately, real multi-models are not "ideal": firstly, the number of participating single models is usually finite, and secondly, the model errors are not fully independent. Consequently, in a real forecasting setting, CCR is more efficient in improving reliability than MMEC, as will be shown with seasonal forecast data from the DEMETER / ENSEMBLES database. The adverse effect of signal dilution due to the recalibration, on the other hand, is a function of the signal strength itself and is therefore only visible in a few high-skill regions. Indeed, it turns out that single models, if recal-

ibrated using CCR, can on average have comparable skill as multi-model ensembles, depending on the skill metric applied and its sensitivity on reliability. The implications of these findings, particularly for an operational prediction context, will be discussed.