Geophysical Research Abstracts, Vol. 9, 04324, 2007
SRef-ID: 1607-7962/gra/EGU2007-A-04324

Probabilistic verification of weighted multi-models

A.P. Weigel and M.A. Liniger and C. Appenzeller

Federal Office of Meteorology and Climatology, MeteoSwiss, Zurich, Switzerland
(andreas.weigel@meteoswiss.ch)

In recent years probabilistic ensemble forecast systems have become a powerful tool for the quantification of weather and climate risks. While probabilistic forecasts are consequent and desirable in principle, it is not trivial to verify them such that the full information content is considered. Indeed, the formulation of appropriate skill scores is still an area of ongoing research, in particular for ensemble systems.

Among the most widely used probabilistic skill scores are the Brier and Ranked Probability Skill Scores (BSS and RPSS, respectively), which are based on a quadratic metric applied in probability space. As skill scores, they quantify the degree to which a given ensemble prediction system outperforms a (typically climatological) reference strategy. From earlier studies it is known that the BSS/RPSS are substantially negatively biased for small ensemble sizes. This flaw imposes major problems on the verification of ensemble predictions, especially in the context of seasonal forecasts and multi-model approaches, where large ensemble sizes are not yet standard.

In this contribution a new view of the BSS/RPSS is presented. In a first step, we show that the bias of these skill scores can be removed if the effects of finite ensemble size are adequately considered in the climatologic reference scores. This leads to the formulation of a new, bias-less "version" of the BSS/RPSS (henceforth referred to as BSSD and RPSSD, respectively) that quantifies the potential skill of an ensemble prediction system. Its performance is demonstrated in synthetic and real case examples, and the corresponding significance levels are evaluated. The bias itself is identified as the “intrinsic unreliability” of the ensemble prediction system.

In a second step, the RPSSD, which so far could only be applied to single model ensembles, is generalized to weighted multi-model ensemble forecasts. This generalized RPSSD has the same formal structure as in the single model case, if an “effective en-

semble size" characterizing the multi-model is introduced. This is of practical importance for multi-model assessment studies, where the consequences of varying effective ensemble size need to be clearly distinguished from the true benefits of multi-model combination.

The presentation concludes with a discussion of the consequences of such a bias removal and the relevance for the verification of real world forecasting systems. It is suggested to use the RPSSD as a complement rather than a replacement for the classical RPSS.