The debiased Brier and ranked probability skill scores

A.P. Weigel, 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.

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.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. Its performance is demonstrated in a synthetic and a real case example, and the corresponding significance levels are evaluated. The interpretation of the bias, consequences of bias removal and its practical relevance will be discussed. A central conclusion of this study is that increasing the number of ensemble members does not increase prediction skill per se; rather, the statistical significance of the skill scores is enhanced.