Geophysical Research Abstracts, Vol. 9, 07188, 2007
SRef-ID: 1607-7962/gra/EGU2007-A-07188

Evaluation of Fuzzy Scores for Quantitative Precipitation Forecast Verification using a Testbed Approach

F. Ament (1), T. Baehler (1) and E. E. Ebert (2)

(1) MeteoSwiss, Switzerland (felix.ament@meteoswiss.ch), (2) Bureau of Meteorology Research Centre, Australia

Recently developed mesoscale numerical weather prediction models have the potential to resolve small scale precipitation structures like convective cells or squall lines. However, the predictability of such features is limited and they are often simulated not exactly at the right position in space or time. Consequently, traditional verification scores based on matching forecast and observation point-by-point have problems to reflect the value of mesoscale forecasts.

Due to the demand of scores that evaluate the structure of quantitative precipitation forecast (QPF) while being rather insensitive to positioning errors, numerous "fuzzy" scores have been proposed recently: They have all in common to consider spatial or temporal neighbourhoods surrounding the point interest instead of focusing on a single grid point. The size of the neighbourhood can be varied in order to evaluate the forecast performance on various spatial scales. However, the interpretation of the forecast data within a neighbourhood differs among the scores and different aspects are emphasised. Applying various fuzzy techniques to the same forecast will thus end up with partly contrasting results and it depends on the application to choose the most appropriate score.

We will present a testbed to determine the characteristics of fuzzy scores. The general idea is to monitor the reaction of the scores to typical, well-defined QPF errors: In the testbed framework, an idealised or real radar derived precipitation field is used as observational reference. Next, a virtual forecast is generated by applying typical perturbation to the reference, like e.g. a systematic bias, white noise or spatial translations. After verifying this virtual data set, the degradation of the score can be used as a measure of sensitivity to a certain type of error. The significance of the results is enhanced by performing multiple realisations of perturbation based on a variety of reference fields. Additionally, the variance of degradations of these verifications indicates the reliability of the scores. We will discuss testbed results for 12 different fuzzy scores: For example, the simplest perturbation, which is changing nothing, yields surprising results. Almost half of the analysed scores deviate from perfect values at coarser scales, although forecast and observations agree perfectly. This finding in combination with the results for other perturbations may be used as a guideline to select an appropriate score for a certain application.