A new non-Gaussian evaluation method for ensemble forecasts based on analysis rank histograms

JAN D. KELLER*1,2 and ANDREAS HENSE1

1Meteorological Institute, University of Bonn, Germany
2German Meteorological Service (DWD), Offenbach, Germany

(Manuscript received July 9, 2010; in revised form March 2, 2011; accepted March 21, 2011)

Abstract
The verification of ensemble forecasts is an active field of research. Today, more and more weather forecasting centers implement ensemble forecasts in their operational weather prediction cycles, giving rise to the demand for reliable verification methods especially designed for probabilistic forecasting. Beside well-known scores (such as the Brier score or the continuous ranked probability score) analysis rank histograms or Talagrand diagrams are very popular amongst the ensemble forecasting community. However, a shortcoming of this verification method is its graphical character. In this paper, we present a score which can easily be derived from analysis rank histograms. The so-called β-score holds the advantage of expressing the graphical character of the histogram, and therefore the character of the ensemble spread, in just one single number. We show results of a β-score analysis of operational ensemble forecasts from different centers for different regions and forecast lead times. In addition to the β-score, we present the so-called β-bias which quantifies the shift of analysis rank histograms to lower or higher values and therefore a relative measure of ensemble forecast bias. Results of the β-bias are also shown in this paper.

1 Introduction
Forecast verification is the process of assessing the quality of a numerical prediction. As such a verification is needed to set the forecast’s relevance into relation to other forecasts; it is an inherent part of modern numerical weather prediction (NWP) systems. Hence, as methods and techniques of weather forecasting are subject to a permanent evolution, verification tools have to be adapted, discarded or new methods have to be developed to account for new demands.

Over the last decades a strong evolution process in NWP could be observed, leading the focus to ensemble forecasting. Hence, the statistical nature of the forecast is altered from a deterministic to a probabilistic forecast and as a matter of course, this development essentially changes the requirements of forecast verification.

A verification is the estimation and analysis of the common probability density function between forecast and observation (MURPHY and WINKLER, 1987), a verification tool should account for this. Verification scores are a commonly used tool to assess forecast quality and are therefore an integral part of the verification process. Such a score is a single value which represents some aspect of forecast quality and is determined from forecast and observation data. Following MURPHY and WINKLER this should be done in relation to the respective probability density functions.

Traditional numerical environmental prediction, especially NWP, produced an output consisting of one single number per time interval and grid point at most under the premise of the perfect model assumption. Various well-established verification methods for this kind of forecast have been developed for time series or spatial