



## Trends in the predictive performance of raw ensemble weather forecasts

Stephan Hemri (1,2), Michael Scheuerer (3), Florian Pappenberger (2,4,5), Konrad Bogner (2,6), and Thomas Haiden (2)

(1) Heidelberg Institute for Theoretical Studies (HITS), Heidelberg, Germany, (2) European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK, (3) Physical Sciences Division, NOAA/ESRL, Boulder, Colorado, USA, (4) College of Hydrology and Water Resources, Hohai University, Nanjing, China, (5) School of Geographical Sciences, University of Bristol, Bristol, UK, (6) Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Birmensdorf, Switzerland

Over the last two decades the paradigm in weather forecasting has shifted from being deterministic to probabilistic. Accordingly, numerical weather prediction (NWP) models have been run increasingly as ensemble forecasting systems. The goal of such ensemble forecasts is to approximate the forecast probability distribution by a finite sample of scenarios. Global ensemble forecast systems, like the European Centre for Medium-Range Weather Forecasts (ECMWF) ensemble, are prone to probabilistic biases, and are therefore not reliable. They particularly tend to be underdispersive for surface weather parameters. Hence, statistical post-processing is required in order to obtain reliable and sharp forecasts.

In this study we apply statistical post-processing to ensemble forecasts of near-surface temperature, 24-hour precipitation totals, and near-surface wind speed from the global ECMWF model. Our main objective is to evaluate the evolution of the difference in skill between the raw ensemble and the post-processed forecasts. The ECMWF ensemble is under continuous development, and hence its forecast skill improves over time. Parts of these improvements may be due to a reduction of probabilistic bias. Thus, we first hypothesize that the gain by post-processing decreases over time. Based on ECMWF forecasts from January 2002 to March 2014 and corresponding observations from globally distributed stations we generate post-processed forecasts by ensemble model output statistics (EMOS) for each station and variable. Parameter estimates are obtained by minimizing the Continuous Ranked Probability Score (CRPS) over rolling training periods that consist of the  $n$  days preceding the initialization dates.

Given the higher average skill in terms of CRPS of the post-processed forecasts for all three variables, we analyze the evolution of the difference in skill between raw ensemble and EMOS forecasts. The fact that the gap in skill remains almost constant over time, especially for near-surface wind speed, suggests that improvements to the atmospheric model have an effect quite different from what calibration by statistical post-processing is doing. That is, they are increasing potential skill. Thus this study indicates that (a) further model development is important even if one is just interested in point forecasts, and (b) statistical post-processing is important because it will keep adding skill in the foreseeable future.