Royal Netherlands
Meteorological Institute
Ministry of Infrastructure and the Environment

A comparison between raw EPS output, (modified) BMA and extended LR using ECMWF EPS precipitation reforecasts

Maurice Schmeits* (schmeits@knmi.nl) and Kees Kok*

A comparison has been made [1] between raw ensemble output, (modified) Bayesian model averaging (BMA) [1, 2] and extended logistic regression (LR) [3], using a 20-year (i.e. autumns of 1982-2001) ECMWF ensemble reforecast data set of precipitation and a 20-year data set of a dense precipitation observation network in the Netherlands.

1. Introduction

A previous study [2] has indicated BMA and conventional LR to be successful in calibrating multi-model ensemble forecasts of precipitation for a single forecast projection, but a more elaborate comparison between these methods has not yet been made. This study compares the raw ensemble output, (modified) BMA and extended LR for single-model ensemble reforecasts of precipitation, namely from the ECMWF ensemble prediction system (EPS).

2. Data sets, statistical methods and predictand definitions

The data sets used in this study [1] are precipitation observations from the observation network of volunteers in the Netherlands and precipitation data from a reforecasting experiment with the ECMWF EPS system.

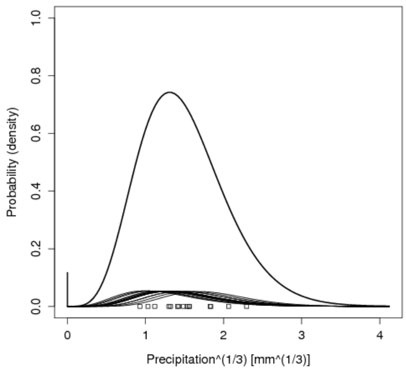

Figure 1: BMA-fitted pdf of 24-h accumulated mean precipitation in the grid box ($3.5 - 4.5^\circ$ E, $51.5 - 52.5^\circ$ N) on 25 September 1988 0800 UTC for a training window length w of 247 cases (after [2]). The thick vertical line at 0 represents the BMA estimate of the probability of no precipitation, and the top solid curve is the BMA pdf of the precipitation amount given that it is nonzero. The bottom curves are the components of the BMA pdf, namely the weighted contributions from the bias-corrected ensemble members, and the squares represent the ensemble member precipitation forecasts.

The BMA method is described in [2] and is illustrated in Fig. 1. A modified version of BMA [1] has also been used, in which a simple bias correction is applied instead of a correction based on model output statistics (MOS), as well as a modified formulation for the probability of precipitation (POP). For the extended LR method we refer to [3].

Both the BMA predictand and the extended LR predictand is defined as the probability density function (pdf) of the area-mean 24-h accumulated precipitation amount at 0800 UTC for a $1^\circ \times 1^\circ$ grid box. There are 12 grid boxes [1], which have been pooled.

3. Results using cross-validation

The continuous ranked probability skill score (CRPSS) of the BMA system with respect to the raw EPS has been computed (Fig. 3 of [1]). Only for the 30-h forecast projection the CRPSS is positive, indicating that the BMA system has more skill than the raw ensemble. From the 54-h projection on, the BMA system is as skilful as or less skilful than the raw ensemble, the latter being an undesirable property of a statistical post-processing method based on that raw ensemble.

This bad skill for the longer forecast projections is caused by a decreased spread of the bias-corrected ensemble members, leading to a post-processed ensemble that is underdispersed. The reason is that MOS is used for the bias-correction of the individual ensemble members, leading to a regression of the individual ensemble members towards the climatological mean.

If a simple bias correction is applied instead, together with a modified POP formulation, the performance of BMA is significantly improved for e.g. the 78-h (Fig. 2) and 126-h projections (Fig. 4c of [1]). On the other hand, the modified BMA system is about as skilful as the conventional BMA system for the 30-h projection (Fig. 4a of [1]).

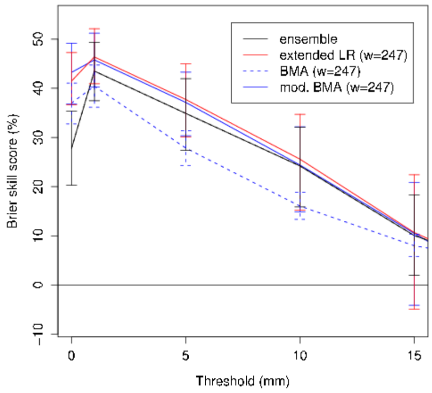

Figure 2: Brier skill score (%) of area-mean precipitation forecasts, with respect to sample (i.e. the autumns of 1982-2001) climatology, as a function of the precipitation threshold (mm) for the 78-h projection (from [1]); the vertical bars indicate the 90% block bootstrap confidence intervals and the training window length $w = 247$.

4. Conclusions

Surprisingly, BMA is less skilful than the raw EPS output from forecast day 3 onward. This is due to the bias correction in BMA, which applies MOS to individual ensemble members. As a result, the spread of the bias-corrected ensemble members is decreased, especially for the longer forecast projections. Here a simple bias correction has been applied instead. Besides, the POP equation in BMA has also been changed. These modifications to BMA lead to a significant improvement in the skill of BMA for the longer projections (e.g. Fig. 2). The difference in skill between the raw EPS, extended LR and modified BMA is generally not statistically significant.

5. References

[1] M. J. Schmeits and C. J. Kok. Mon. Wea. Rev., 138:4199–4211, 2010.

[2] J. M. Sloughter, A. E. Raftery, T. Gneiting, and C. Fraley. Mon. Wea. Rev., 135:3209–3220, 2007.

[3] D. S. Wilks. Meteorol. Appl., 16:361–368, 2009.

*Royal Netherlands Meteorological Institute (KNMI), P.O. Box 201, 3730 AE De Bilt, The Netherlands.