Calibrating ensemble forecasts of quantitative precipitation: An emprical comparison

Alexander Henzi [1] Johanna F. Ziegel [1] Tilmann Gneiting [2]

[1] Institute of Mathematical Statistics and Actuarial Science, University of Bern, Switzerland

[2] Heidelberg Institute for Theoretical Studies and Karlsruhe Institute of Technology, Germany

alexander.henzi@stat.unibe.ch johanna.ziegel@stat.unibe.ch tilmann.gneiting@h-its.org

Introduction

- Ensemble prediction systems have been very successful, but remain subject to biases and dispersion errors, thus fail to be calibrated.
- Statistical postprocessing improves raw ensemble forecasts.
- Comparison of raw and postprocessed ECMWF ensemble forecasts for quantitative precipitation at four major European airports in 2007 to 2017.

Statistical postprocessing of ensemble forecasts

Goal:

Calibration of ensemble forecasts by modeling the distribution of the precipitation amount Y conditional on the ensemble members $X_1, \ldots, X_d$:

$$F(y) = P(Y \le y \mid X_1, \ldots, X_d).$$

In practice:

- Define a suitable class $\mathcal{F}$ of admissible cumulative distribution functions (CDFs).
- Estimate $F \in \mathcal{F}$ using training data (n days of past forecasts and observations):

$$\hat{F} = \arg\min_{F \in \mathcal{F}} \sum_{i=1}^{n} L\big(F(y_i; x_{i,1}, \ldots, x_{i,d}), y_i\big),$$

where L is a loss function.

Major tasks:

- Find a suitable class $\mathcal{F}$: How to model the mixed discrete-continuous distribution of precipitation?
- Select a loss function L: Use proper scoring rules (Gneiting and Raftery, 2007), typically

$$L(F, y) = \mathrm{CRPS}(F, y) := \int_{-\infty}^{\infty} (F(z) - 1\{y \le z\})^2 \, dz.$$

- Define appropriate training dataset: How large should n be? Seasonal selection of training data?

Popular calibration techniques

- Ensemble Model Output Statistics (EMOS): Based on left-censored generalized extreme value (GEV_0) distributions (Scheuerer, 2014) or censored, shifted gamma distributions (Scheuerer and Hamill, 2015).
- Bayesian Model Averaging (BMA, Sloughter et al., 2007).

Isotonic Distributional Regression (IDR): A new nonparametric calibration technique

- Isotonic Distributional Regression (IDR, Henzi, 2018): A recently developed calibration technique.
- Nonparametric: Only assumption is that the observed precipitation accumulation increases, as forecasts increase.
- Optimal in-sample fit with respect to a wide range of scoring functions, including the CRPS.

Learn more about IDR at the presentation by Johanna Ziegel on Thursday, 10:00.

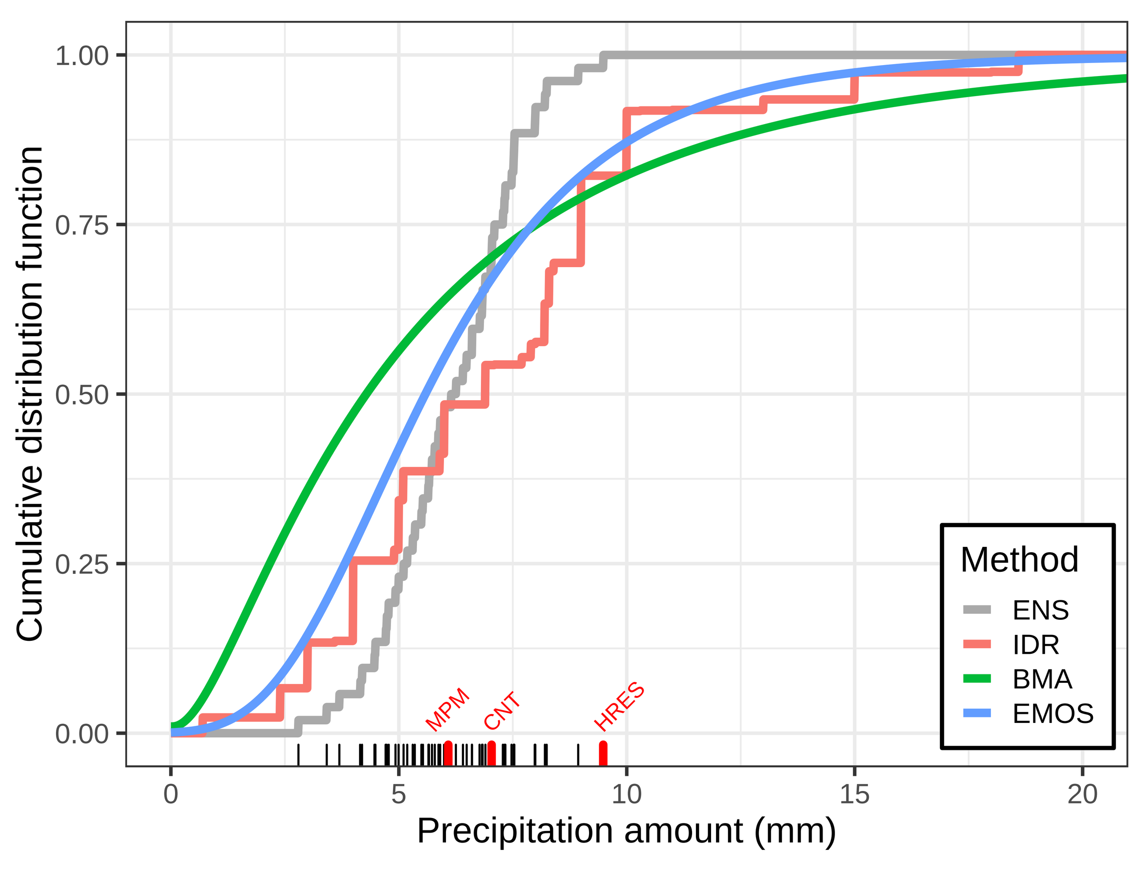

Figure 1: *Raw ECMWF ensemble (ENS) and postprocessed (IDR, BMA, EMOS) 24h forecasts for the precipitation amount at Zurich on 2016/02/21. Marks at bottom: Raw ensemble members with high resolution (HRES), control (CNT) run and mean of perturbed members (MPM).*

Data

- Test period: 2007 – 2017.
- ECMWF 52-member IFS forecast consisting of high resolution (HRES), control (CNT) and 50 perturbed members.
- Forecasts for gridbox of size $0.25^\circ \times 0.25^\circ$ containing the station, lead times 24 to 120 hours.
- Airport station observations (Frankfurt, Zurich, London Heathrow, Brussels).
- IDR, BMA and EMOS (GEV_0) postprocessed forecasts based on HRES, CNT and mean of perturbed members (MPM)
- EMOS additionally uses the spread of the perturbed members as a predictor.

Verification measures

Probabilistic forecasts should be consistent with the observed precipitation amount (calibration) and, on that basis, have concentrated predictive distributions (sharpness).

Tools:

- Reliability diagram and probability integral transform (PIT) histogram to assess calibration: Does an event with a predicted probability of p occur at a rate of p?
- Brier score (BS) to assess probability of precipitation (PoP) forecasts.
- Continuous ranked probability score (CRPS) as a measure of overall predictive performance, covering both calibration and sharpness.

Choice of training period

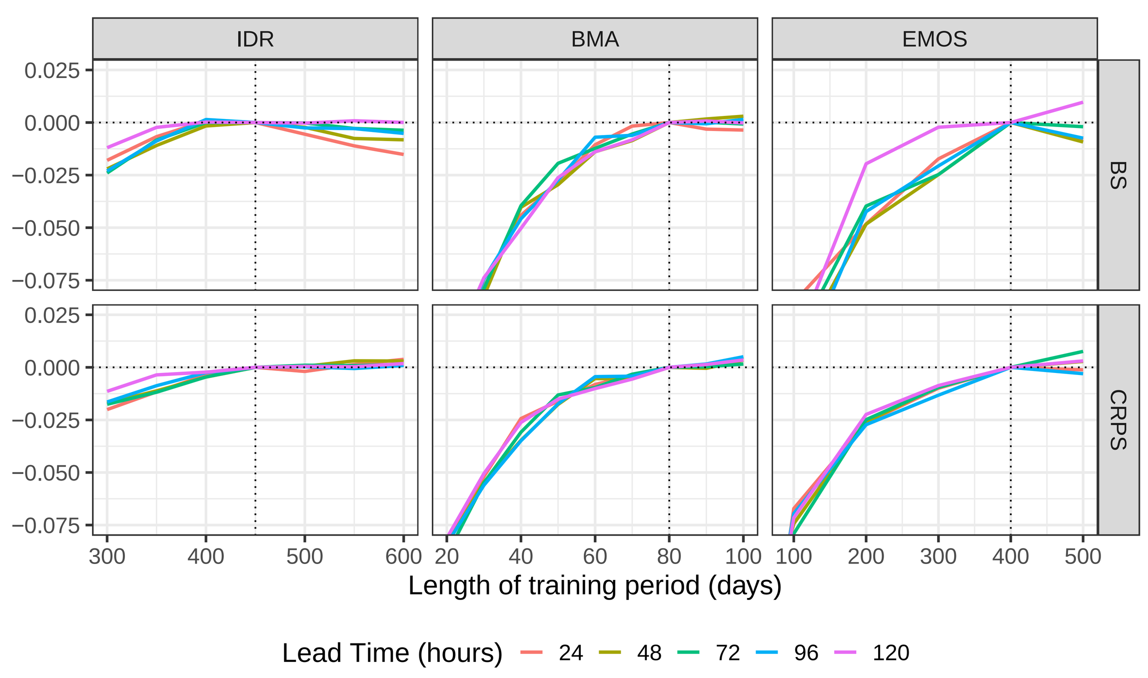

Figure 2: *BS and CRPS skill score relative to forecasts using a training period of 450, 80 and 400 days for IDR, BMA and EMOS, respectively (averages over all four stations).*

- Optimal training period: 450 days for IDR, 80 days for BMA, 400 days for EMOS.
- Seasonal selection of training data does not improve IDR, but slightly benefits EMOS (results shown are without seasonal selection).

Results

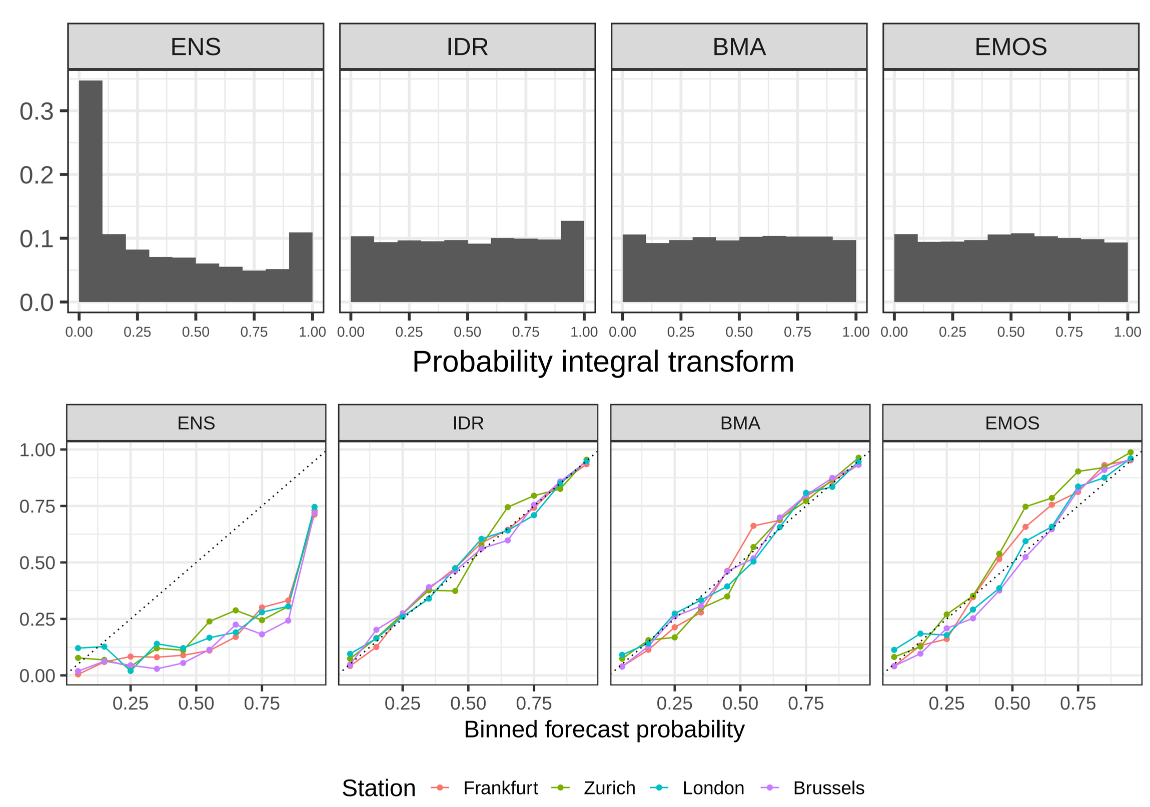

Figure 3: *PIT histograms for raw and postprocessed ECMWF ensemble forecasts (top) and reliability diagrams for PoP forecasts (bottom), at a lead time of 48 hours.*

- Statistical postprocessing corrects the underdispersion of the raw ECMWF ensemble.
- Postprocessed forecasts are calibrated with reliable PoP.

Figure 4: *Brier skill score of postprocessed PoP forecasts.*

- With postprocessing, Brier score for PoP forecasts decreases by up to 60%.
- IDR yields highest improvement.

Figure 5: *CRPS skill score of postprocessed forecasts.*

- Postprocessing reduces CRPS by up to 17%.
- EMOS outperforms BMA and IDR, due to its superiority at higher thresholds.

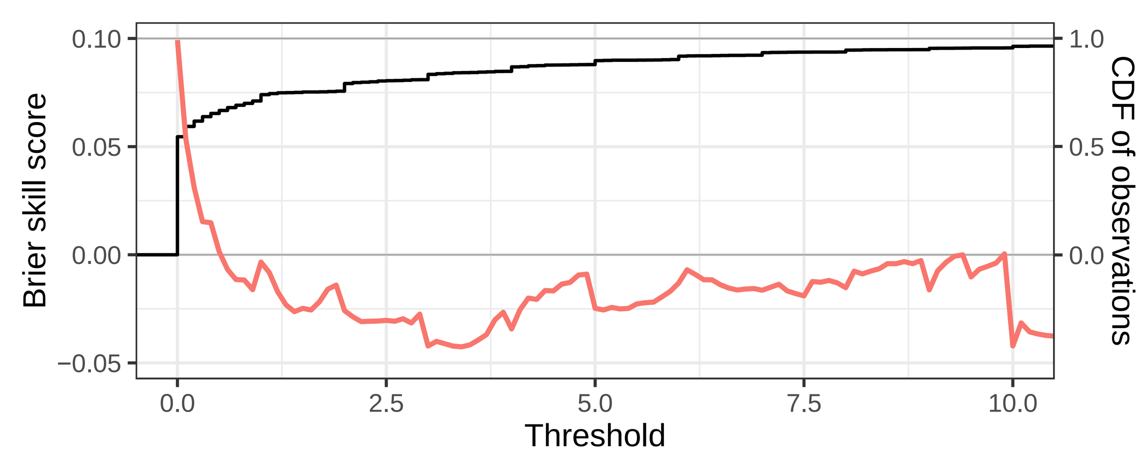

Figure 6: *Brier skill score of IDR relative to EMOS as a function of the threshold value for 24h forecasts at Frankfurt (in red), along with CDF of observed precipitation amount.*

- Performance of EMOS based on gamma distributions (not shown) is similar to using GEV_0.

Discussion

- Statistical postprocessing improves ECMWF ensemble forecasts of quantitative precipitation.
- EMOS achieves the lowest CRPS values, but IDR performs better for low precipitation amounts, despite using less information.
- IDR is an ideal, easy to implement reference method.
- Readily applicable to TIGGE and S2S.

References

Gneiting, T. and Raftery, A. E. (2007), 'Strictly proper scoring rules, prediction, and estimation', *Journal of the American Statistical Association* **102**(477), 359–378.

Henzi, A. (2018), Isotonic distributional regression (IDR): A powerful nonparametric calibration technique, Master's thesis, University of Bern.

Scheuerer, M. (2014), 'Probabilistic quantitative precipitation forecasting using ensemble model output statistics', *Quarterly Journal of the Royal Meteorological Society* **140**(680), 1086–1096.

Scheuerer, M. and Hamill, T. M. (2015), 'Statistical postprocessing of ensemble precipitation forecasts by fitting censored, shifted gamma distributions', *Monthly Weather Review* **143**(11), 4578–4596.

Sloughter, J. M., Raftery, A. E., Gneiting, T. and Fraley, C. (2007), 'Probabilistic quantitative precipitation forecasting using Bayesian model averaging', *Monthly Weather Review* **135**(9), 3209–3220.