**To what extent does the diagnosis of multiple grid-box weather types add value in post-processing ensemble rainfall forecasts?**

Fatima Pillosu1,2, Tim Hewson2

1 University of Reading, Reading, UK

2 ECMWF, Reading, UK

**Abstract**

**Plain Language Summary**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Introduction

# Data

## Raw ECMWF ensemble

## ecPoint post-processing technique

## Rainfall observations

# Methods

Reliability (i.e., when x% probability is forecast is the event observed on x% of occasions?) and discrimination ability (i.e., the ability to distinguish between event and non-event) are desirable properties of ensemble forecasts (Wilks, 2019). While calibration deals with the meaning of probabilities, discrimination appraises the existence of a signal in the forecasts when an event materializes and its absence in the opposite situation (Ben Bouallègue & Richardson, 2022). To have value, post-processing techniques must improve both properties in the raw NWP forecasts. The performance of one year of retrospective forecasts from the three systems, the raw NWP, ecPoint-MultipleWT, and ecPoint-SingleWT, was compared using 12-hourly rainfall observations from the datasets described in section 2.3 as truth. Although raw NWP model output does not pertain to point values, it is very common to verify it using them, as in ECMWF’s two headline measures for precipitation (Haiden et al., 2023).

## Estimation of forecast reliability: Brier Score – Reliability Component

The reliability component of the Brier score is used to assess the reliability of the examined forecasts (Jolliffe & Stephenson, 2011).

## Estimation of “real” forecast resolution: ROC curves and area under the ROC curve computed with the trapezoidal approximation

The area under the relative operating characteristic (ROC) curve (AROC) is used as a summary measure of the forecasts’ discrimination ability (Jolliffe & Stephenson, 2011). The ROC plots the hit rate (HR) versus the false alarm rate (FAR) of an event for incremental decision thresholds. The ROC curve is defined by the line joining successive ROC points, where each point corresponds to results for increasing decision thresholds, from the top right to the bottom left corners of the plot.

The decision variable is the number of members exceeding the event threshold. Thus, the issued forecasts can take values in [0, 1/M, 2/M, …, 1] for an ensemble of size M. As a consequence, the resulting ROC curves are based on M+1 points (i.e., 52 for the raw ENS, and 100 for ecPoint-MultipleWT and ecPoint-SingleWT).

The area under the straight lines formed by connecting the M+1 points [including the (0,0) and (1,1) points] of the ROC plot correspond to the AROC with the so-called trapezoidal approximation (AROCT). This nomenclature come from the fact that the area is estimated considering straight lines between two consecutive points of the plot and so as a sum of trapeziums.

For rare events, there is a tendency for the points on the ROC to cluster toward the lower left corner of the graph (Casati et al., 2008). When computing AROCT, a straight line is drawn between the last meaningful point on the ROC curve and the top-right corner to close the ROC curve, giving the impression that part of the curve is missing. How much of the curve is missing depends on the lowest category, defined here by the ensemble size and the base rate of the event.

## Estimation of “potential” forecast resolution: Binormal approximation for the computation of ROC curves and area under the ROC curve

To draw a full ROC curve, one can apply the so-called binormal model (Harvey et al., 1992). The fitting of the ROC curve with the binormal model is based on the assumption that HR and FAR are integrations of normal distributions (i.e., a signal and a noise distribution), respectively. Harvey et al. (1992) provide a close-form for the computation of the AROC. The fitting of the HR and FAR requires a Z-transformation based on the unit normal distribution. For this reason, the result AROC is denoted as AROCZ.

When applied to ensemble-derived probability forecasts for rare events, this approach consists effectively in an extrapolation to a hypothetical continuous decision variable based on the limited set of decision thresholds materially assessable. Because such a decision variable may not be achievable in practice, AROCZ is sometimes considered as a measure of the potential discrimination that could be achieved for an unlimited ensemble size (Ben Bouallègue & Richardson, 2022; Bowler et al., 2006).

# Results

Figure 2 displays results for three 12 h accumulation thresholds: 0.2 mm (“dry or not”), 10 mm (“wet”), and 50 mm (“extreme, with flash flood potential”).

## Forecast reliability

## “Real” and “potential” forecast resolution

# Case Study: Extreme rainfall and flash floods in China in July 2021

# Discussion

## Forecast reliability

## “Real” and “potential” forecast resolution

# Conclusions

# Figures

A diagram of a computer

Description automatically generated with medium confidence

Figure 1 – The top-left area of the figure shows the mathematical equation defining the rainfall errors between observations and forecasts (i.e., Forecast Error Ratio, FER). The top-right area of the figure shows the distribution of such errors (i.e., Mapping Function, MF). The MFs are used in the ecPoint methodology to post-process the raw rainfall forecasts. The distributions can be split into sub-sets based on a specific grid-box weather type (WT, e.g., mainly large-scale rainfall and low steering winds). If the top distribution with no splits is used in the post-processing, the post-processed rainfall forecasts are called single-WT ecPoint-Rainfall forecasts (ecPoint\_SingleWT, MF within the black circle). If the top distribution is split into sub-sets based on a certain number of grid-box WTs, the post-processed rainfall forecasts are called multiple-WT ecPoint-Rainfall forecasts (ecPoint\_MultipleWT, MF within the grey rectangle).

A group of graphs showing different types of data

Description automatically generated

Figure 2 – Brier score, reliability component for verifying rainfall events (VRE) >= 0.2 mm/12h, 10 mm/12h, 25 mm/12h, and 50 mm/12h (from top-left to bottom-right). Errors are given with a confidence level of 99%.

A group of graphs showing different types of data

Description automatically generated

Figure 3 – “Real” and “Binormal” areas under the ROC curve (AROC and AROCz) for verifying rainfall events (VRE) >= 0.2 mm/12h, 10 mm/12h, 25 mm/12h, and 50 mm/12h (from top-left to bottom-right). Errors are given with a confidence level of 99%.

# References

Ben Bouallègue, Z., & Richardson, D. S. (2022). On the ROC area of ensemble forecasts for rare events. *Weather and Forecasting*, *37*(5), 787–796. https://doi.org/10.1175/waf-d-21-0195.1

Bowler, N. E., Pierce, C. E., & Seed, A. W. (2006). STEPS: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP. *Quarterly Journal of the Royal Meteorological Society*, *132*(620), 2127–2155. https://doi.org/10.1256/QJ.04.100

Casati, B., Wilson, L. J., Stephenson, D. B., Nurmi, P., Ghelli, A., Pocernich, M., Damrath, U., Ebert, E. E., Brown, B. G., & Mason, S. (2008). Forecast verification: Current status and future directions. *Meteorological Applications*, *15*(1), 3–18. https://doi.org/10.1002/met.52

Haiden, T., Janousek, M., Vitart, F., Ben-Bouallegue, Z., & Prates, F. (2023). Evaluation of ECMWF forecasts, including the 2023 upgrade. *ECMWF Technical Memoranda*, *911*, 1–60.

Harvey, L. O., Hammond, K. R., Lusk, C. M., & Mross, E. F. (1992). The Application of Signal Detection Theory to Weather Forecasting Behavior. *Monthly Weather Review*, *120*(5), 863–883. https://doi.org/10.1175/1520-0493(1992)120

Jolliffe, I. T., & Stephenson, D. B. (2011). Forecast Verification: A Practitioner’s Guide in Atmospheric Science. In *Forecast Verification* (2nd Editio). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119960003

Wilks, D. S. (2019). Statistical Methods in Atmoshperic Sciences. In *Statistical Methods in the Atmospheric Sciences* (Fourth Edi). Elsevier Inc. https://doi.org/https://doi.org/10.1016/C2017-0-03921-6