Evaluating the post-processing of the European Flood Awareness System's medium-range streamflow forecasts

Gwyneth Matthews[1] | Hannah Cloke[1,2,3,4,5] | Sarah L. Dance[1,6] | Christel Prudhomme[5,7,8]

1. Department of Meteorology, University of Reading | 2. Department of Geography and Environmental Sciences, University of Reading | 3. Department of Earth Sciences, Uppsala University | 4. Centre of Natural Hazards and Disaster Science | 5. European Centre for Medium-Range Weather Forecasts | 6. Department of Mathematics and Statistics, University of Reading | 7. Department of Geography, Loughborough University | 8. UK Centre for Ecology and Hydrology

University of Reading

UKRI Engineering and Physical Sciences Research Council

ECMWF

1. Introduction

The European Flood Awareness System (EFAS) produces medium-range (up to 15 days) streamflow forecasts as part of the European Commission's Copernicus Emergency Management Services. At locations with historic and near real-time discharge observations, the forecasts are post-processed. We evaluate the post-processing method at 522 stations to identify areas for improvement and to quantify the benefits of post-processing.

Key Questions

- **Does the post-processing provide improved forecasts?**
- **What affects the performance of the post-processing method?**

2. Post-processing technique

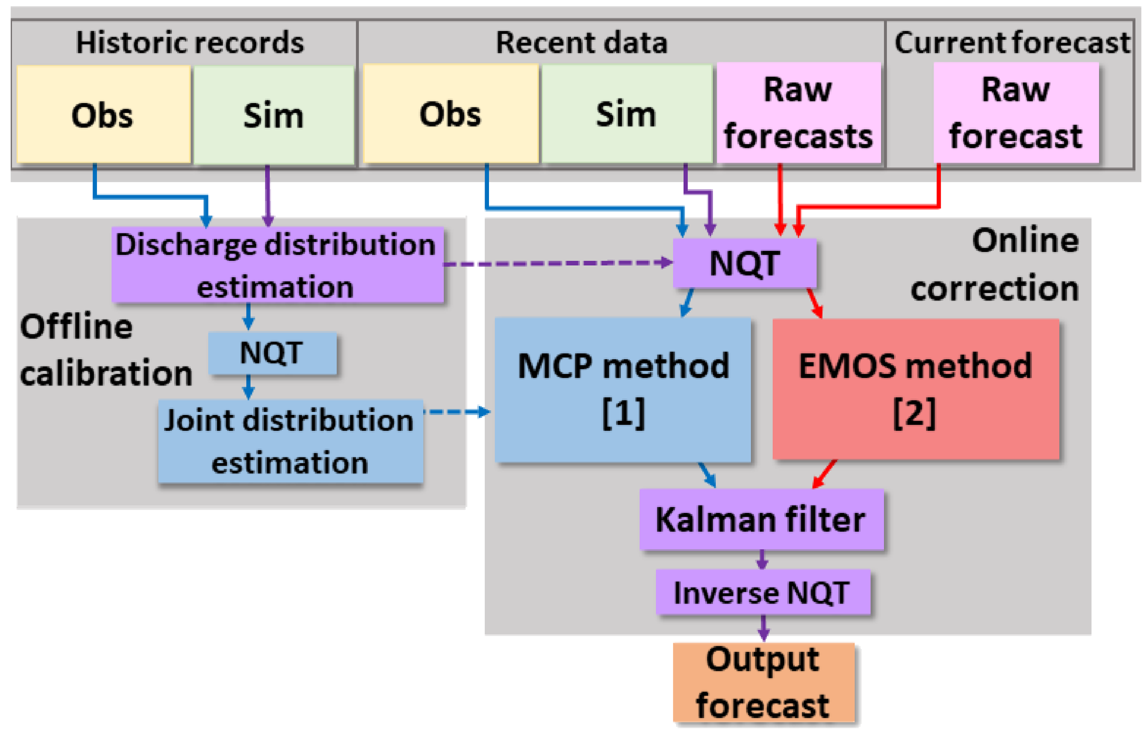

Figure 1: Post-processing method for a station. Input data are separated by time and data type. Colour of arrows and boxes show which uncertainties the data and methods are used to quantify. **Blue: Hydrological. Red: Meteorological. Purple: Both.**

3. Evaluation Strategy

Post-processing method is evaluated by comparing the raw forecasts with the post-processed forecasts.

- 2 years of twice-weekly ensemble reforecasts (**208 forecasts**)
- Evaluation uses daily discharge **observations**.
- Skill scores use **raw forecast as benchmark**:

$$\textit{Skill score} = \frac{score_{pp} - score_{raw}}{score_{perf} - score_{raw}}$$

Stations selection criteria:

1) No overlap between the calibration timeseries and the evaluation period.
2) At least 95% of the observations for the evaluation period.

4. Results

4.1. Forecast median

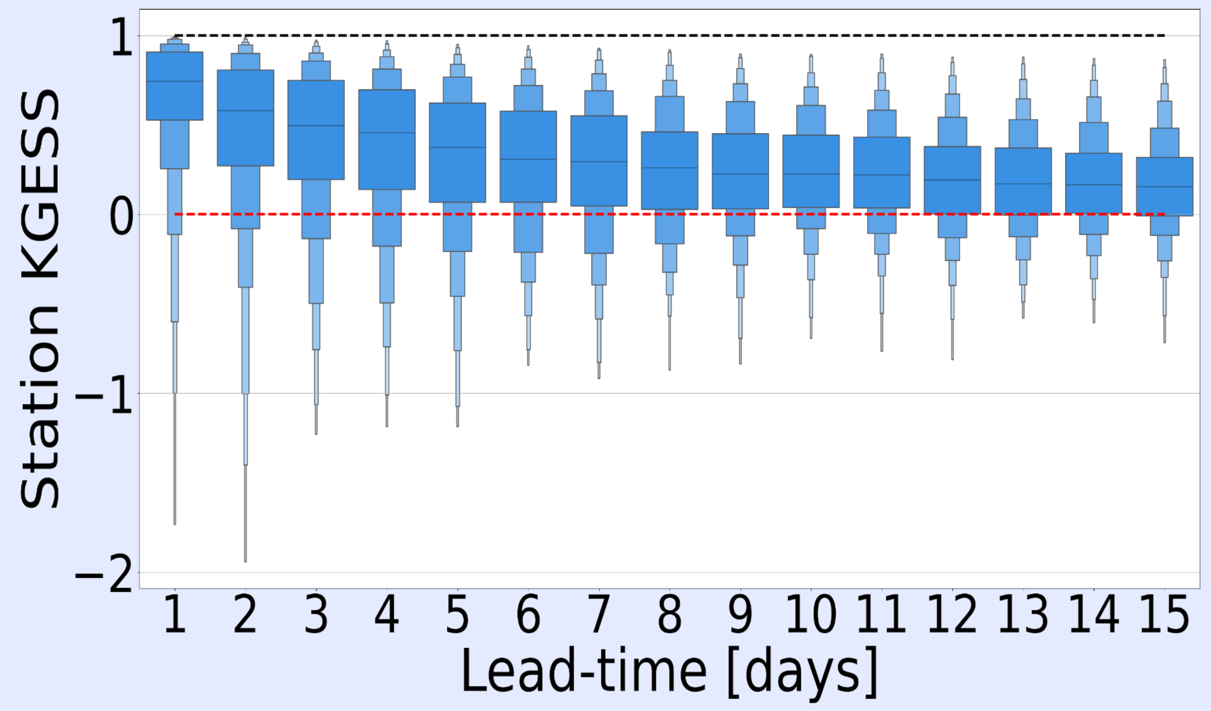

Figure 2: Station modified Kling-Gupta Efficiency Skill Score (KGESS) value against lead-time. **Black line:** Station KGESS=1 (perfect score). **Red line:** Station KGESS=0 (below which forecasts medians are degraded). Outliers not shown.

- **Forecast median improved at majority of stations (~75% at a lead-time of 15 days.**
- **Decrease in magnitude of improvement at longer lead-times for most stations.**
- **Increase in lowest KGESS value with lead-time due to decrease in skill of benchmark.**
- **Greatest improvement is to the correlation component of the modified Kling Gupta Efficiency score.**
- **Bias and variability ratio components improved at most stations but are often over-corrected.**

4.2. Forecast distribution

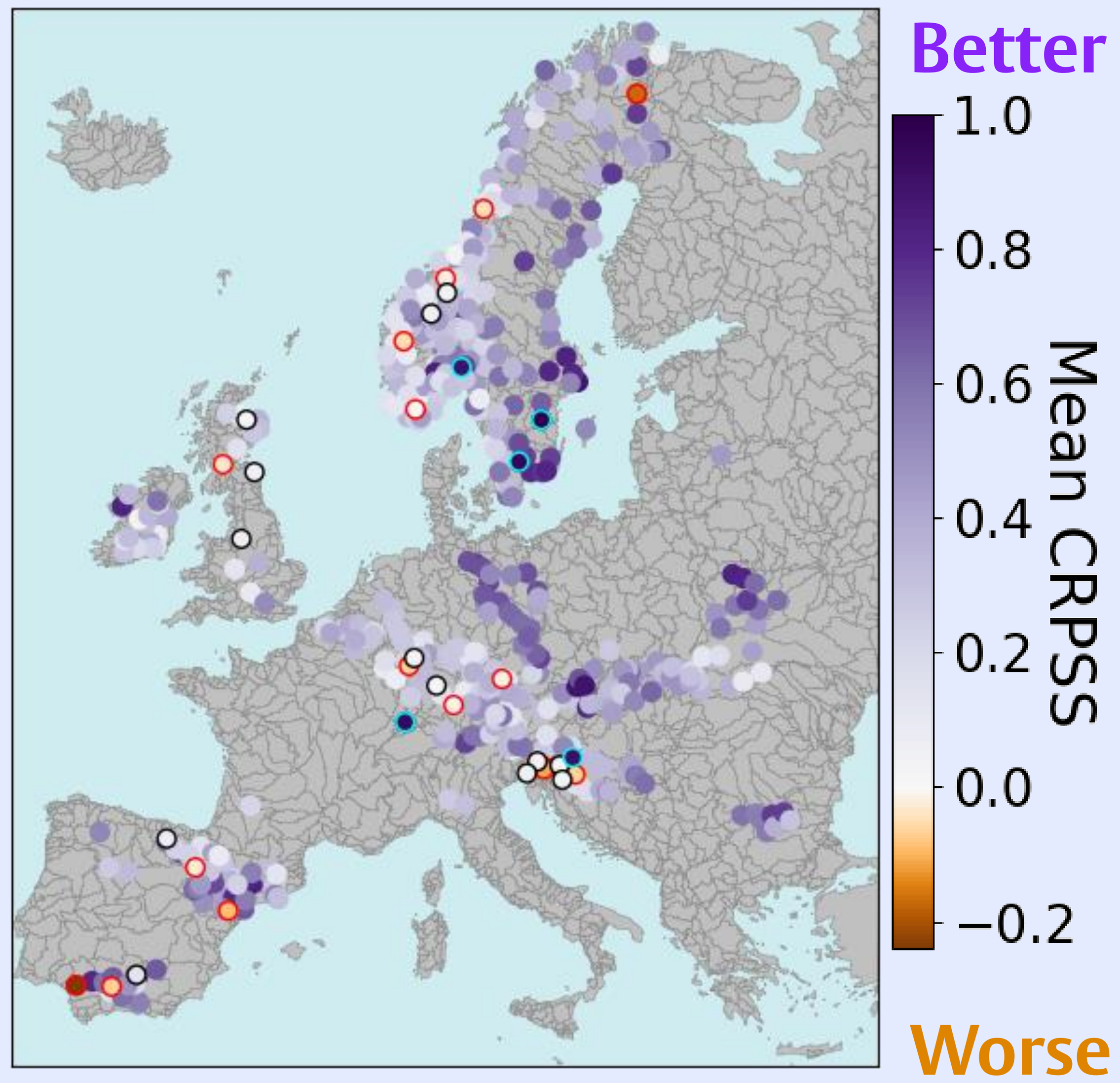

Figure 3: Continuous Ranked Probability Skill Score (CRPSS) averaged over lead-times of 6-10 days. **Red**: Mean CRPSS < 0. **Black**: Mean CRPSS > 0 but a CRPSS < 0 at one or more lead-times. **Blue**: Mean CRPSS > 0.9.

- **Most stations are improved by post-processing (only 16 stations have negative CRPSS) but range of CRPSS is large (-0.24 to 0.99)**
- **Degraded stations tend to be:**
  - **Near mountainous regions (e.g. west of the Scandinavian Peninsula), or**
  - **In flashy catchments (e.g. Southern Spain).**
- **Catchments with lower hydrological model skill are improved more.**

4.3. Impact of catchment characteristics

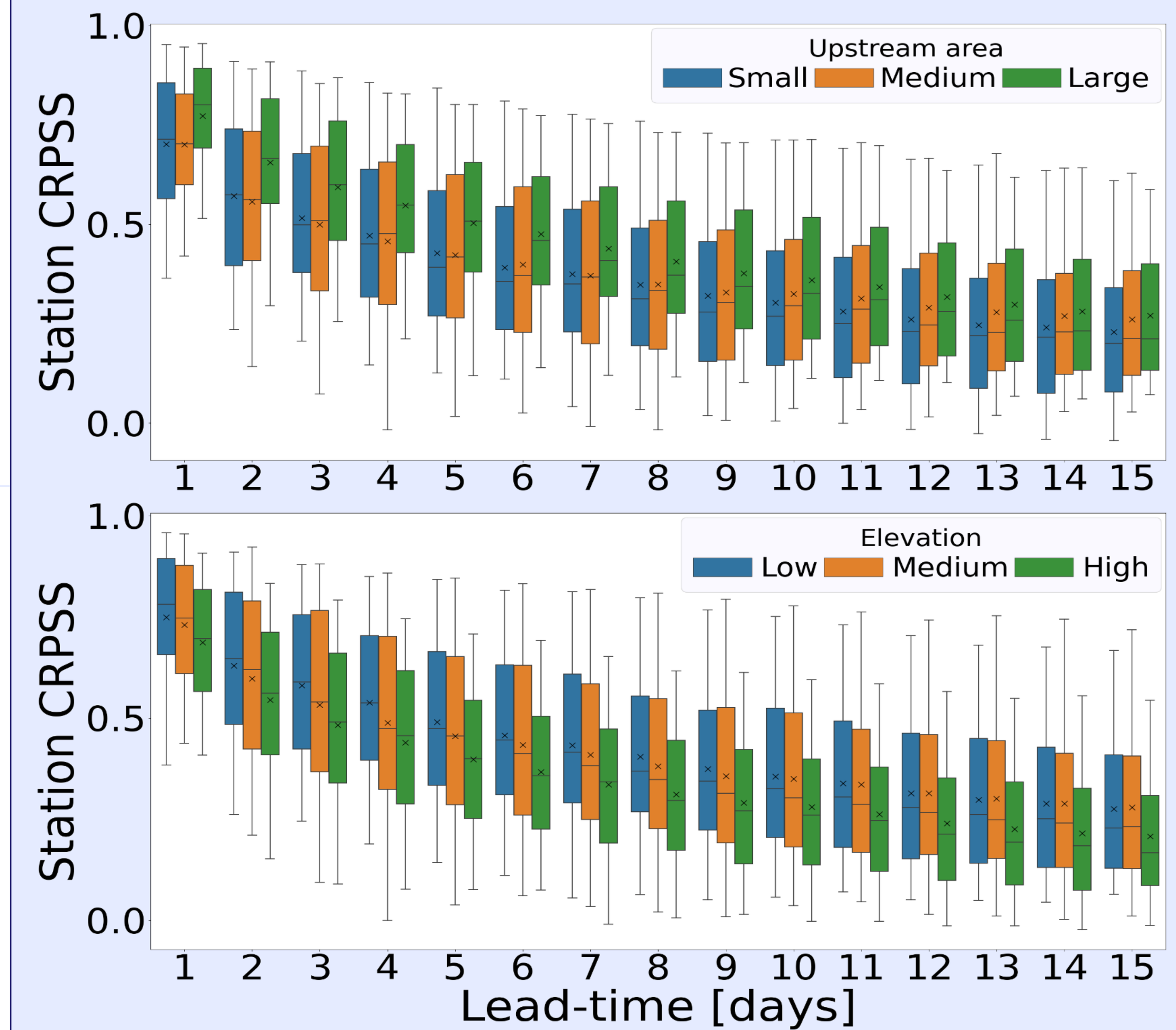

Figure 4: Station CRPSS value against lead-time. Coloured box shows the interquartile range. Whiskers extend to 5th and 95th percentiles. Outliers not shown.

- **Large catchments (green, top) are improved more due to their slower responses. Recent observations are more informative about the discharge in the forecast period.**
- **High catchments (green, bottom) are improved less due to a decrease in response time and an increase in the meteorological errors.**

Key Results

- **Post-processing improves the skill of the streamflow forecasts at the majority of stations.**
- **The improvement decreases at longer lead-times.**
- **The effectiveness of post-processing largely depends on the response time of the catchments.**
- **Hydrological model errors are corrected more than errors in the meteorological forcings.**

References

1. Todini, E. (2008). A model conditional processor to assess predictive uncertainty in flood forecasting. International Journal of River Basin Management, 6(2), 123-137.
2. Gneiting, T, et al. (2005). Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. Monthly Weather Review, 133(5), 1098-1118.
3. Barnard, C, et al. (2020). Reforecasts of river discharge and related data by the European Flood Awareness System, version 4.0, Copernicus Climate Change Service (C3S) Climate Data Store (CDS). (4th March 2021).10.24381/cds.c83f560f

Contact information

- Department of Meteorology, University of Reading, Whiteknights, RG6 6AH
- Email: g.r.matthews@pgr.reading.ac.uk