

# RainForests: A novel Machine Learning approach to calibrating rainfall forecasts

Benjamin Owen\*, Belinda Trotta, Jiaping Liu, Thomas Gale, Anja Schubert, Gary Weymouth  
Bureau of Meteorology, Australia

\*benjamin.owen@bom.gov.au

## The RainForests method

Our approach to forecast rainfall probabilities calibrated at the spatial scale of rain gauges is based on the ECPoint method of Hewson & Pillosu (2021)\*, with the following key differences:

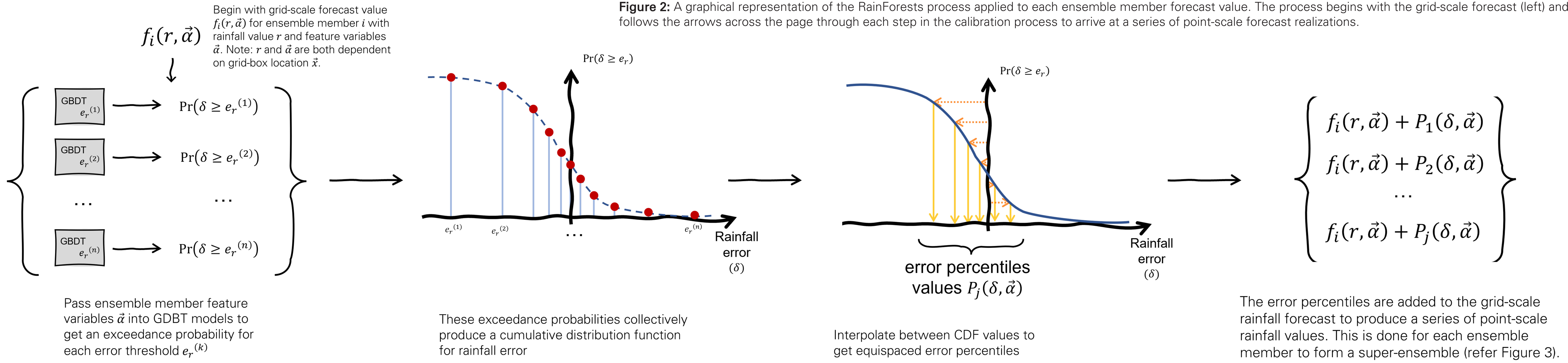
- we use gradient-boosted decision trees (GBDT) in place of the manually trained decision tree
- our observation dataset uses both rain gauges and gauge-calibrated radar data (refer Figure 1)
- we use additive error rather than the forecast error ratio, allowing for calibration when forecast rainfall is zero

The adjacent figures show the calibration process diagrammatically:

- Figure 2: calibration steps applied to each ensemble realization
- Figure 3: how the per member sub-ensembles collectively form the full calibrated ensemble

Probability distributions are evaluated from the calibrated ensemble output for each NWP model and then blended in probability space.

\*Hewson, T.D., Pillosu, F.M. A low-cost post-processing technique improves weather forecasts around the world. Commun Earth Environ 2, 132 (2021)



## How does it compare?

To benchmark the RainForests method we compare to our current production system, the Gridded Operational Consensus Forecast (GOCF).

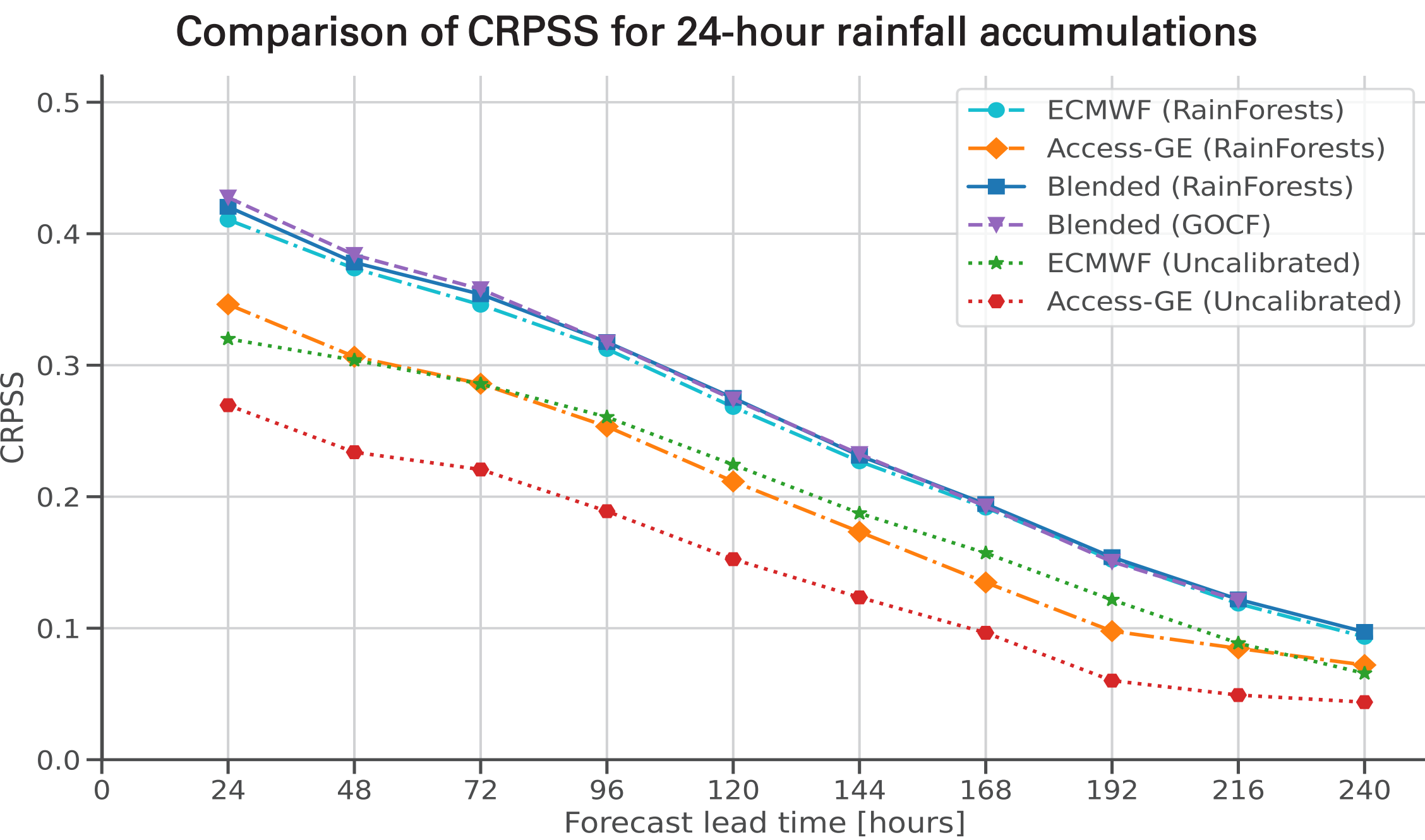


Figure 4: Comparison of the continuous rank probability skill score for 24-hour rainfall accumulations for our new rainfall system (dark blue) against our current system GOCF (purple). Overall skill of the new system is consistent with GOCF with the exception of the very earliest lead-times. Here we also show the uncalibrated and RainForests calibrated ensemble forecasts to highlight the improvement thanks to calibration.

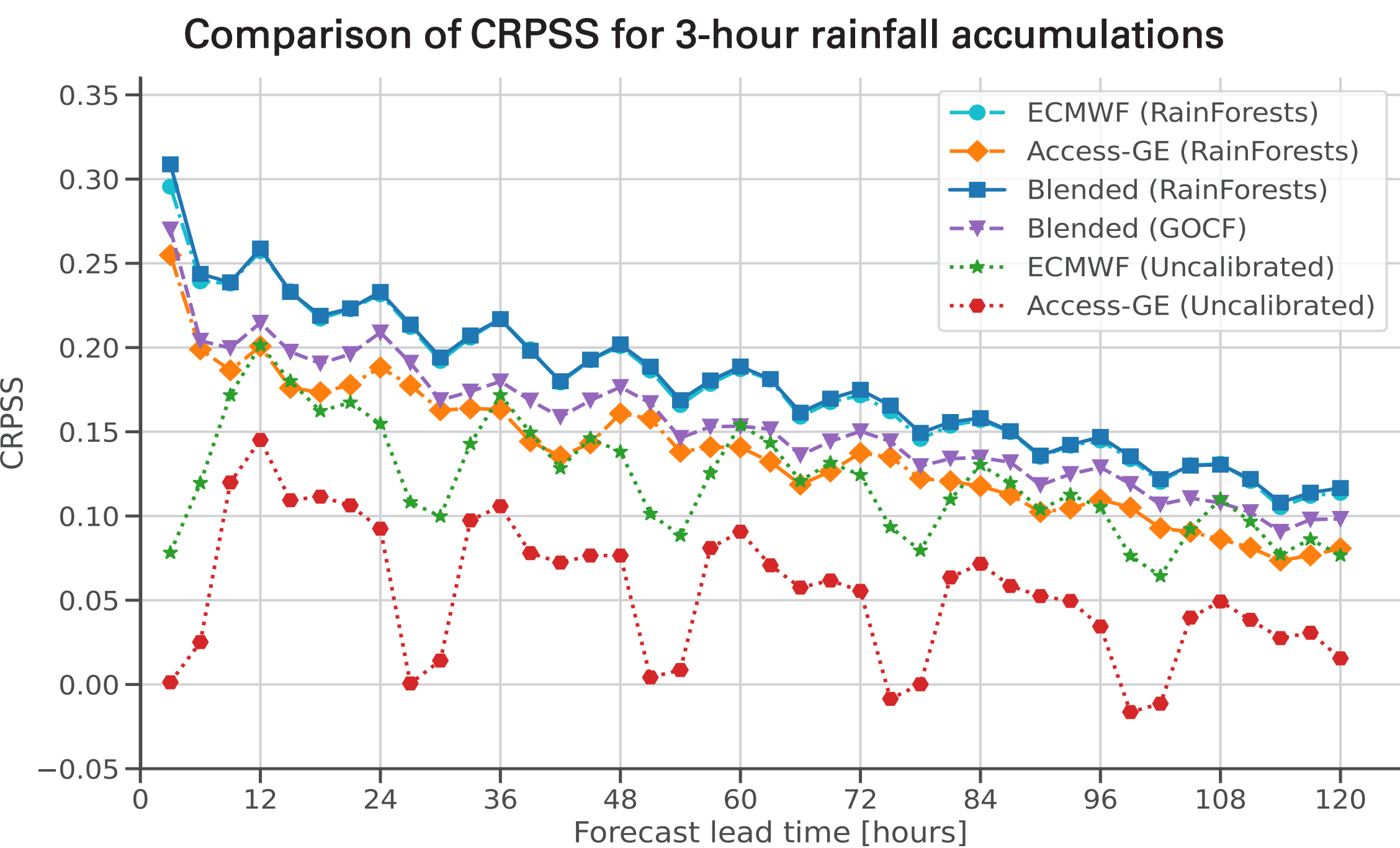


Figure 5: The figure above compares the continuous rank probability skill score for 3-hour rainfall accumulations for our new rainfall system (dark blue) against our current system GOCF (purple). The ECMWF ensemble calibrated using the RainForests method (light blue) alone out-performs GOCF, with further improvement coming through the blending of calibrated model outputs.

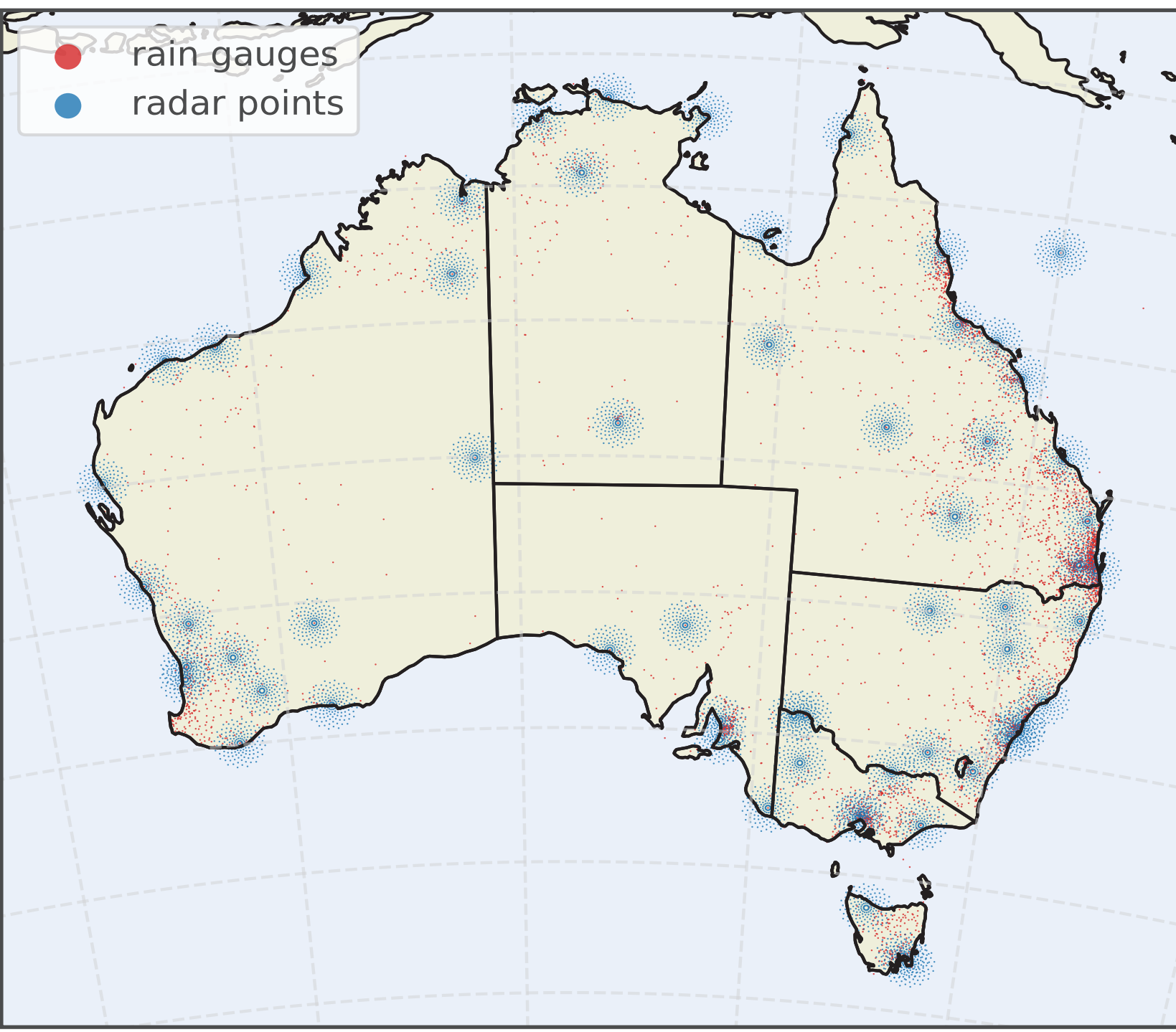


Figure 1: The rain gauge (red) and Radar (blue) observations used in training the GBDT models used in calibrating the forecasts.

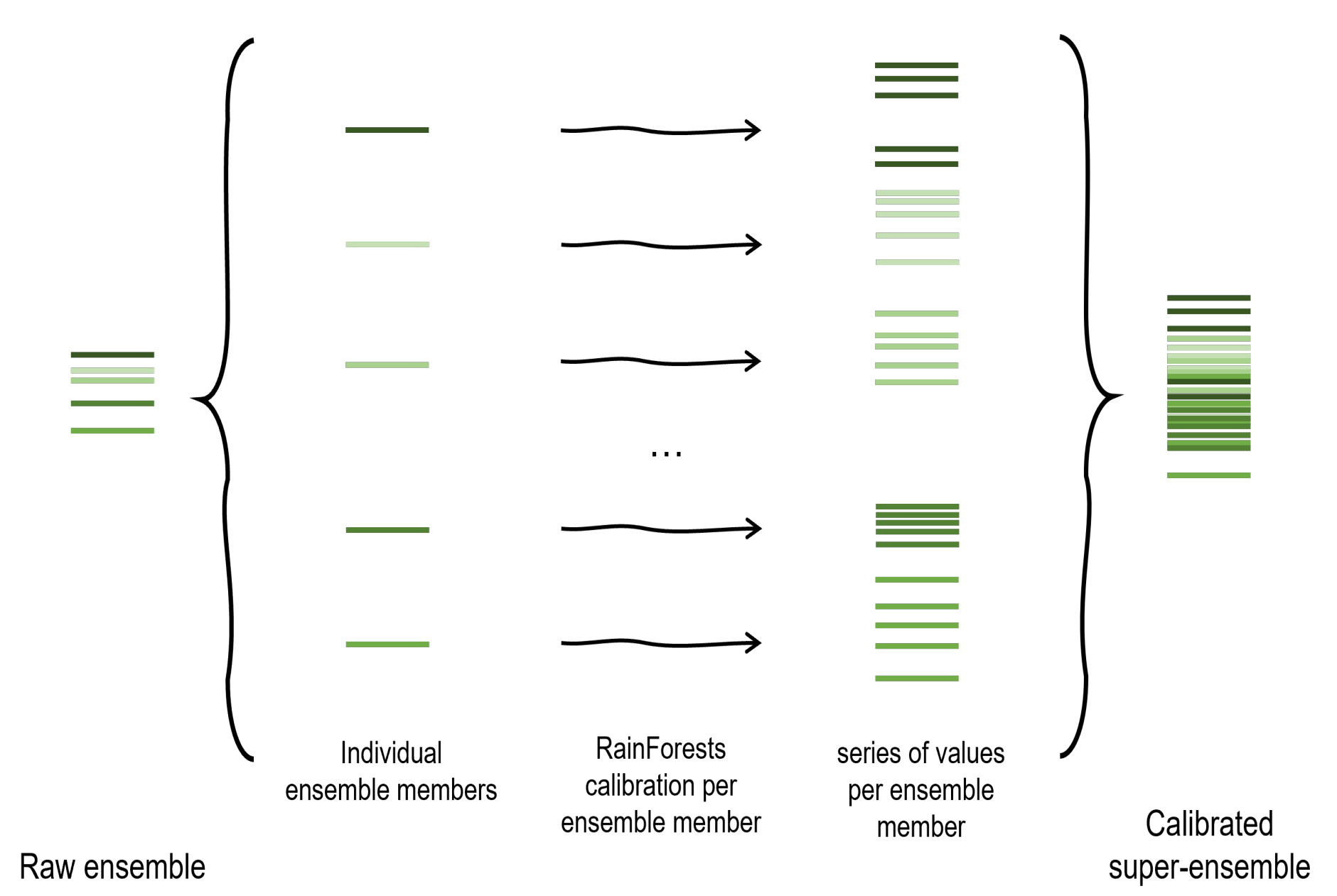


Figure 3: A graphical representation of how we calibrate the full ensemble forecast. Each ensemble member is calibrated separately to give a series of forecast realizations (refer Figure 2). These are combined into a super-ensemble from which a probability distribution is evaluated in the same manner one would from the raw ensemble.

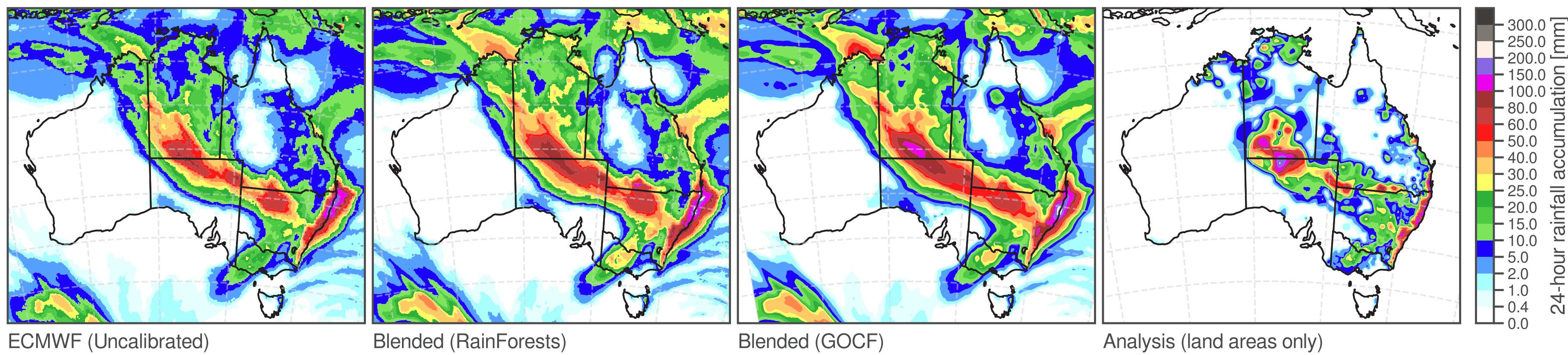
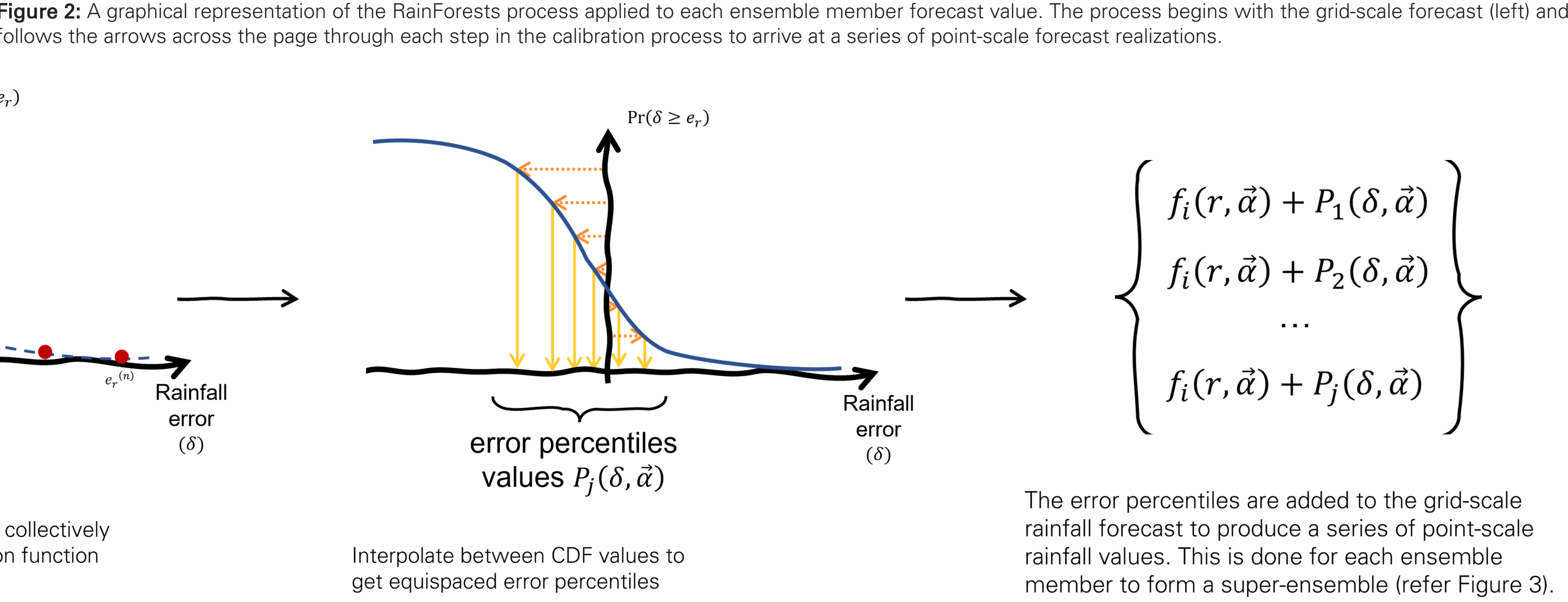


Figure 7: A comparison of the gridded output for the 90th percentile. The forecasts are for lead-day +1 for the 24-hour period ending 2021-03-22 00:00Z; this day saw significant rainfall across large parts of the east coast and central interior of Australia. The rainfall analysis (right) is shown to highlight areas that observed extreme rainfall on the day.

For this comparison we undertook verification at rain gauge sites across Australia over the year 2021 (see figures 4, 5 & 6).

Figure 7 (above) compares gridded output for a significant rainfall event in March 2021.

### What we found:

- Comparable skill for 24-hour accumulations (see Figure 4)
- **Significant skill improvement for 3-hour accumulations** (see Figure 5)
- **Better consistency** between 24-hour and 3-hour forecasts (see figure 6)

### Further Advantages:

- Per model calibration allows for the generation of calibrated ensemble members via Ensemble Copula Coupling or similar methods
- New system is more extensible and easier to maintain than our current system
- Integration into IMPROVER allows for simple inclusion of additional processing steps into the rainfall processing chain
- Applicable to other diagnostics beyond rainfall

### Daily and sub-daily consistency comparison

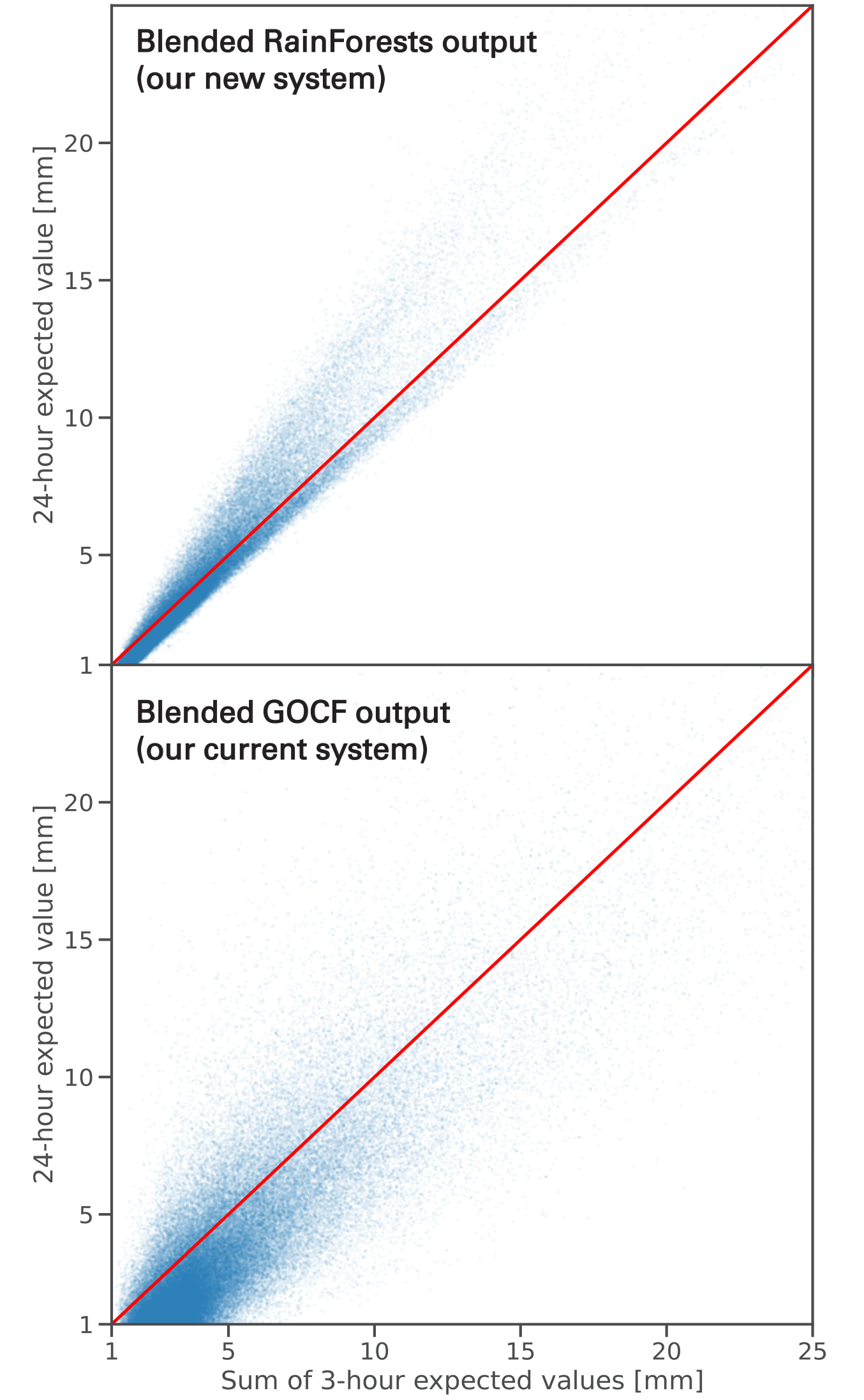


Figure 6: The figure above compares the sum of expected values for 3-hour forecasts with the corresponding 24-hour forecast expected value; ideally these should be equal. The RainForests calibration (top) exhibits good consistency and tight banding, while the GOCF calibration (bottom) shows more spread and a greater proportion below the diagonal.

**Note:** There is a slight offset from the diagonal (present in both datasets) which is the result of insufficient resolution in the rainfall thresholds at low rainfall values (below 0.2 mm).