

# Assessing operational upgrades to an ensemble prediction system

Seamless Global Modelling workshop, 3-6 June 2025, Bristol

Warren Tennant

## What is the purpose of the ensemble?

### Site-specific likely weather

Typically shown as a *meteogram* with the likelihood of meteorological events, e.g. *rain*, *cloud*, *temperature*, shown as percentages. It is designed to provide an overview of expected conditions. **Figure 1** shows clear and dry conditions, turning cloudy with increased chance of rain and stronger winds in the second half of the forecast period.

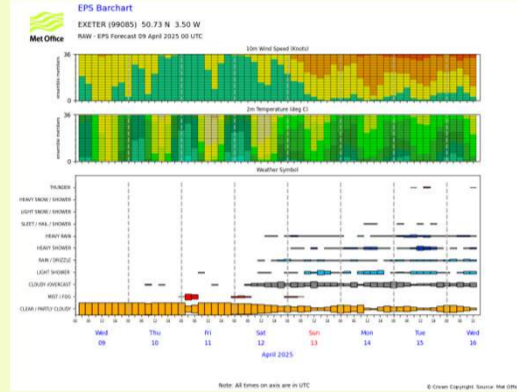
### Probability of an event

Maps showing the probability of a specific event occurring. **Figure 2** shows precipitation amount exceeding a threshold of 10mm in 24-hours.

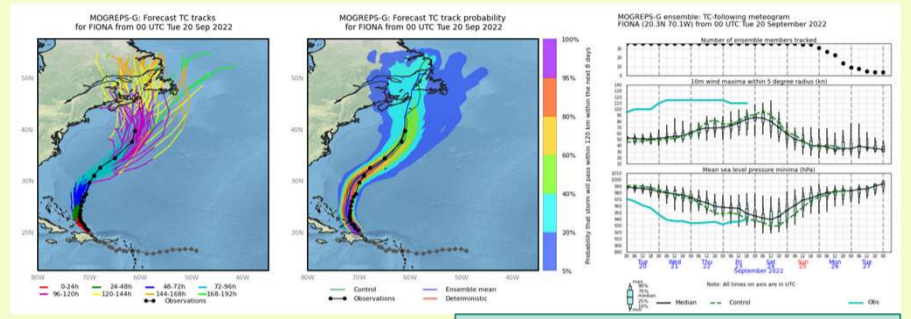
### Alternate scenarios

Plumes of possible tropical cyclone tracks and indications of intensity of the system. **Figure 3** shows an example of a tropical cyclone forecast.

**Figure 1: Meteogram of MOGREPS-G raw forecast probabilities of parameters as shown for a specific site**



**Figure 2: MOGREPS-G forecast map of the probability (%) of 24-hour precipitation exceeding 10mm. Care needs to be taken to avoid confusing this with where maximum rainfall is likely to fall.**



**Figure 3: Plumes of ensemble-member tropical cyclone tracks (left), probability plume of the same (centre) and box-whisker plots of central wind-speed and pressure (right). The solid cyan line shows the observations and previous forecasts to help gauge forecast performance of this system so far, top panel shows number of members included in the forecast plots.**

## MOGREPS-G Summary

### Configuration

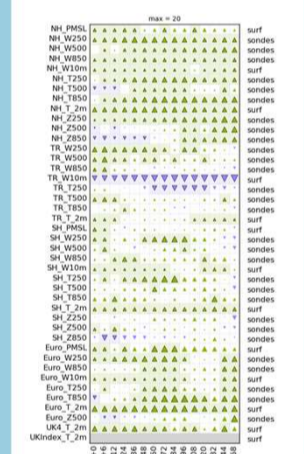
Global model with grid-spacing at 20km, 70 hybrid-height levels with 80km top. Ensemble consists of 8-day forecasts of a control and 17 perturbed members every 6 hours. Uses Ensemble of 4D-Ensemble-Var assimilation (44 members). Model error representation with SKEB and SPT, Additive Inflation, SST and land-surface perturbations.

### PS47 Upgrade

During 2025 we will be upgrading the global coupled-model configuration from GC4 to GC5. This includes improvements to most physics schemes leading to an overall improvement in RMSE and CRPS evaluation. **Figure 4** shows the top-level objective assessment of CRPS changes for all lead-times and various parameters in different regions against observations. Only near-surface winds in the tropics are degraded, but this is understood, and is from stronger winds which were chosen to help improve tropical cyclone intensity.

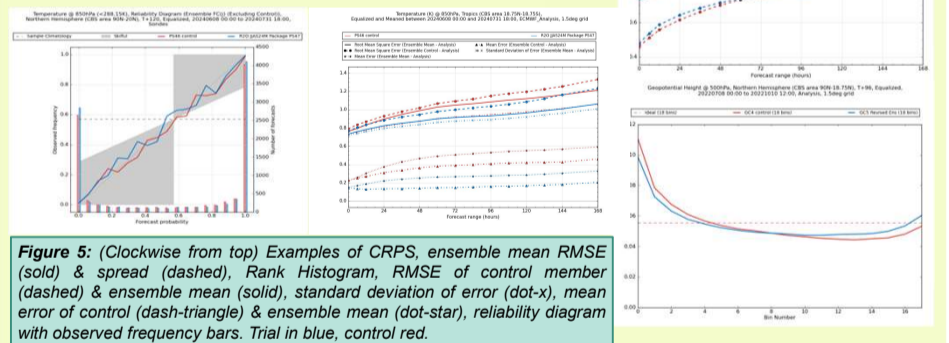
The decision on whether to accept an upgrade into operations relies on these scorecards being positive overall and that any degradation is fully understood and acceptable to customers.

**Figure 4: Scorecard of PS47 package trial. Green (purple) triangles indicate improvement (degradation) for each parameter and lead-time vs PS46.**

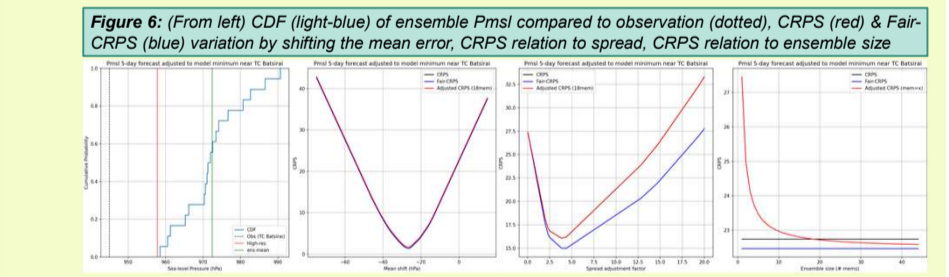


## Evaluating Upgrades

- Consider a range of metrics of mean and variance errors, ensemble spread characteristics
- Understand issues such as increased bias or changes in spread-error
- Figure 6** shows the effects of mean error, ensemble size and spread on CRPS. Biggest gains from reducing mean error, then spread



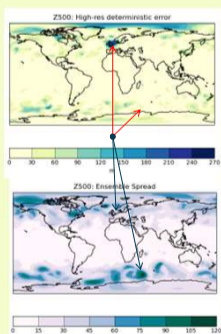
**Figure 5: (Clockwise from top) Examples of CRPS, ensemble mean RMSE (solid) & spread (dashed), Rank Histogram, RMSE of control member (dashed) & ensemble mean (solid), standard deviation of error (dot-x), mean error of control (dash-triangle) & ensemble mean (dot-star), reliability diagram with observed frequency bars. Trial in blue, control red.**



**Figure 6: (From left) CDF (light-blue) of ensemble Pmsl compared to observation (dotted), CRPS (red) & Fair-CRPS (blue) variation by shifting the mean error, CRPS relation to spread, CRPS relation to ensemble size**

## Spread-error relationship

A well-calibrated ensemble should have a good match between spread and error when averaged over time and space (see **Figure 5**). However, **Figure 7** shows that this is not always the case when considering a point at a specific time. This is something lacking in ensemble evaluation and needs further attention.

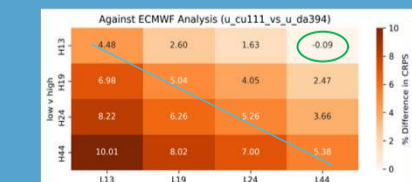


**Figure 7: Map of 500hPa error (top) and ensemble spread (bottom). Arrows highlight specific areas where spread and error do not correlate well.**

## “Good” ensemble?

An improvement can be interpreted in different ways depending on the purpose of the ensemble. Generally, it should be reliable (forecast probability matching the observed frequency), have good resolution (distinguish between different events) and have low mean/RMS error. A large ensemble at coarse grid-spacing can be competitive with one at small fine grid-

spacing (see **Figure 8**). However, the latter would suffer higher errors, less physical realism and poorer event resolution.



**Figure 8: Table of relative CRPS between low & high-resolution forecasts of different ensemble sizes**

## Key Points

- Improve unperturbed model mean and variance error
- Ensure ensemble spread leads to a reliable ensemble with good Brier Score resolution in time and space
- A higher model resolution is often better than more ensemble members
- CRPS remains a good overview metric for evaluation purposes