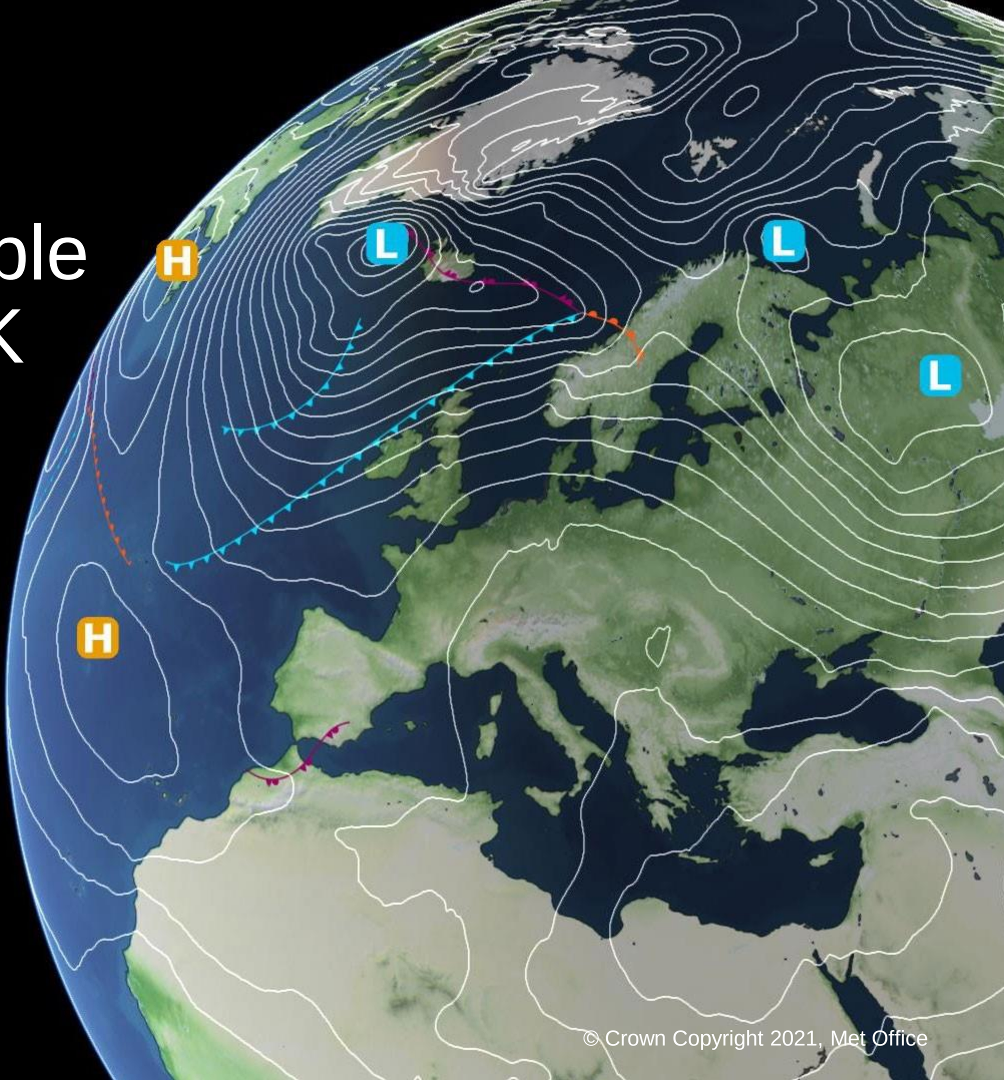
Met Office

Towards an improved understanding of ensemble spread in MOGREPS-UK

Anne McCabe, Nigel Roberts, Marion Mittermaier, Aurore Porson, David Flack, Mike Bush, Steve Willington, David Walters

EWGLAM September 2021



Outline

- The problem of a ‘lack of spread’ in MOGREPS-UK
- Sensitivity tests
- Using the dispersive Fraction Skill Score (dFSS) to evaluate perturbation growth
- Does the ensemble follow the control member?
- Summary and future work
- Extra slides: ensemble SAL

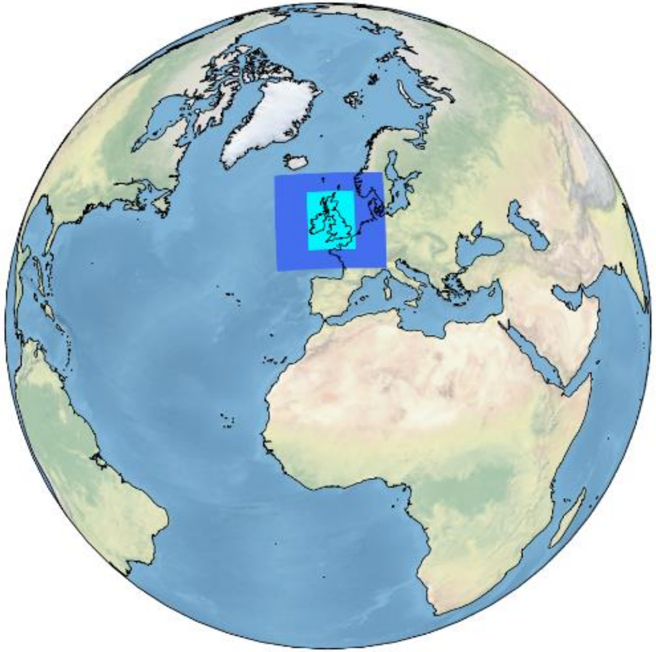
Met Office

The problem of a 'lack of spread' in MOGREPS-UK

- Over the last few years, updates to MOGREPS-UK have resulted in improvements in the objective verification statistics, such as the spread-skill ratio, Brier score, etc
- However, these improvements have not been felt by the operational meteorologists. Main criticisms are:
  - not enough spread
  - MOGREPS-UK 'follows' the deterministic model too closely

  → resulting in a lack of trust in the ensemble
- Here, we take a 'back-to-basics' evaluation of the different ensemble perturbations with the aim of identifying areas of 'weakness' in the ensemble
- This work is part of a larger project looking to tackle the lack of ensemble spread from many different aspects. Other work includes test-beds, subjective verification techniques and the SRNWP-EPS multi-model ensemble work (see Aurore Porson & David Flack's talks in the October workshop).

Research version of MOGREPS-UK (operational up to 2019)



The UK convective scale ensemble

- 12 members
- 54 h forecast length
- Runs 4 cycles per day at 03, 09, 15 and 21 UTC
- 2.2 km resolution on the inner domain, 4 km in the variable resolution zone
- **Boundary conditions** from the Met Office global ensemble, MOGREPS-G
- **Initial conditions** for each ensemble member are provided by the deterministic UK model (UKV) analysis with perturbations from MOGREPS-G added to these
- Stochastic physics from the **Random Parameter (RP)** scheme
- **Stochastic boundary layer perturbations** are applied to all members, including the control, to aid the initiation of convection

In 2019, we moved to a new hourly configuration, which includes 18 members. This new hourly configuration runs to T+120.

Exploring sensitivity to sources of uncertainty

Back-to-basics sensitivity tests

Experiments	Initial Conditions	Lateral Boundary Conditions	Random Parameter (RP) Scheme	Stochastic Boundary Layer (BL) perturbations
IC	Perturbed	As control member	Off	Off
LBC	As control member	Perturbed	Off	Off
RP	As control member	As control member	On	Off
BL Perts	As control member	As control member	Off	On

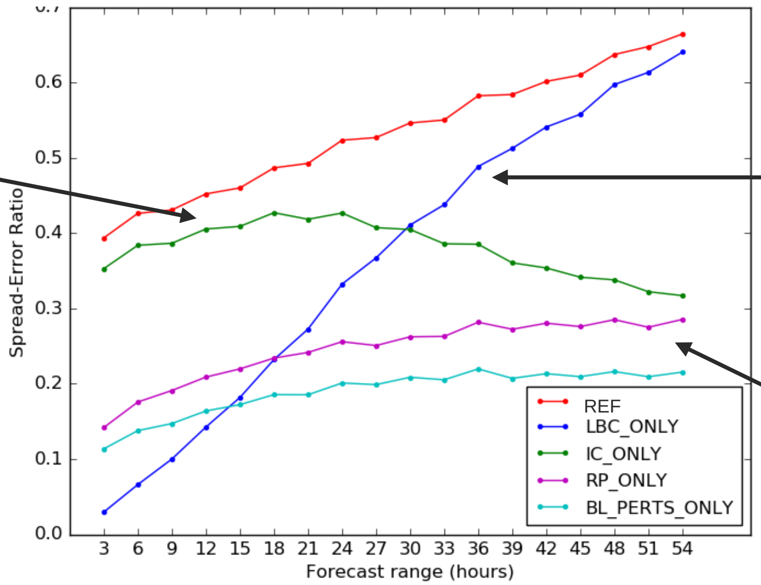
Two month-long trials: summer and winter 2017

Met Office

Grid point spread-error ratio

Initial conditions are most important in the early part of the forecast – they begin to lose significance after 12 hours but dominate over LBC's until T+30.

Surface Temperature (1.5m)

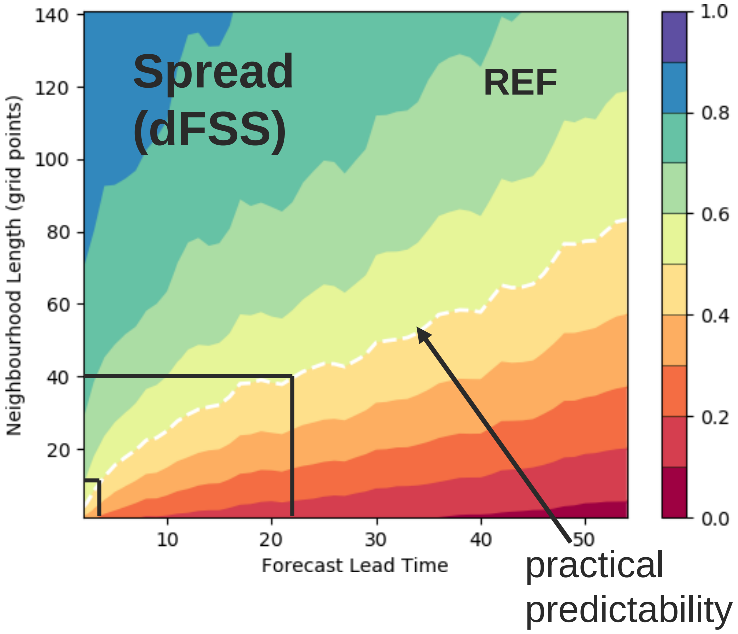


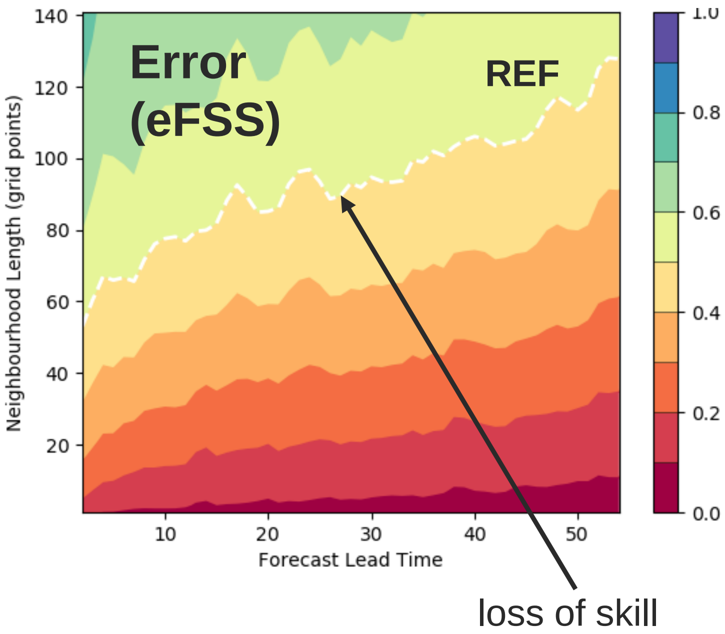
The impact of the LBC's shows a steady increase throughout the forecast

Stochastic physics generates 30 – 40% of the full ensemble spread

Using dispersive FSS (dFSS) to evaluate perturbation growth of the 99th percentile of precipitation forecasts for July 2017 for the REF ensemble

Similar work has been done with a different approach by Frogner et al (2019) & Surcel et al (2015)





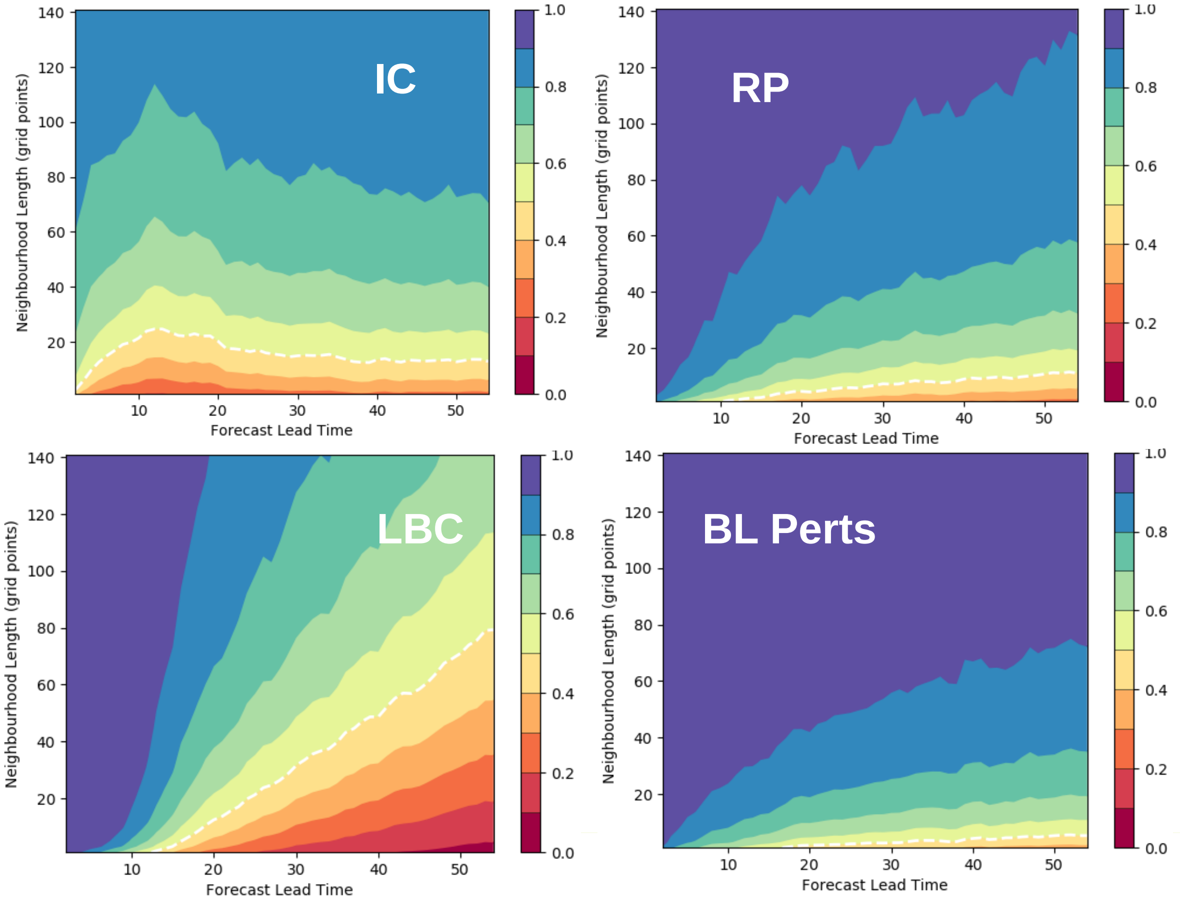
Lower values correspond to ***larger*** spread and ***larger*** error

dFSS references: Dey et al (2014), Roberts (2008), Roberts & Lean (2008)

Met Office

Dispersive FSS for the ensemble experiments

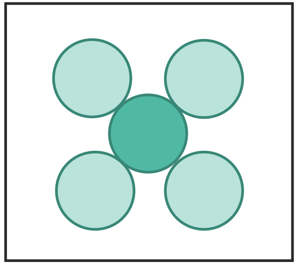
- Initial conditions dominate for the first 24 hours of the forecast and lateral boundary conditions at later times
- IC ensemble fails to upscale from the medium to the large scales
- RP & BL ensembles upscale at a slower rate
- The scale of practical predictability for the RP & BL ensemble is less than for the IC or LBC ensemble



Met Office

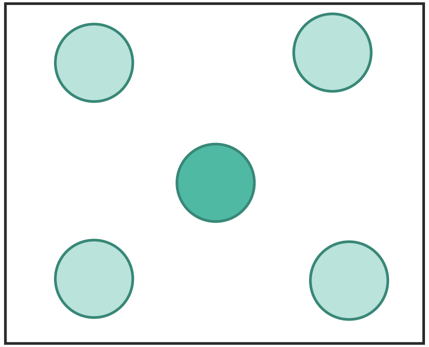
‘Useful’ spread – spatial uncertainty matters

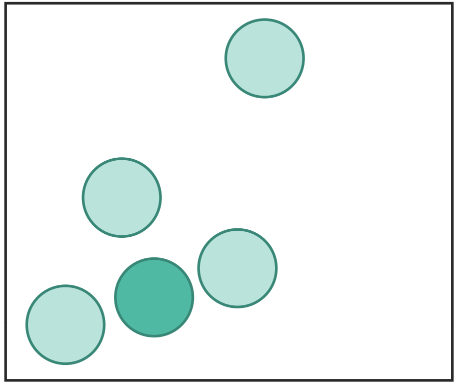
Sometimes, but (crucially) ***not all the time***, we would like the ensemble to …



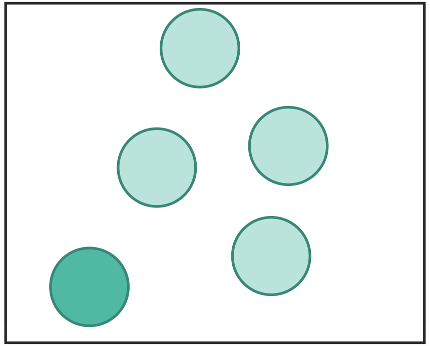
… strongly support the deterministic forecast

… show a large spread of possible outcomes (and quantify the scale of that spread)





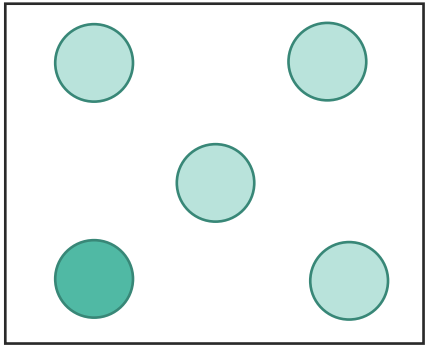
… do something more interesting



Met Office

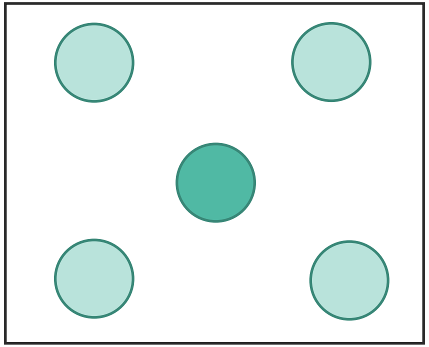
But how do we know how often we forecast these different scenarios?

The dispersive FSS can tell us how spatially spread the members are, but it does not differentiate between these two scenarios:



Standard ‘following the control’ forecast →

← Majority of members shifted away from the control



But the difference is really important to our forecasters

So how often is the control member the central member?

Skok (2016) and **Skok & Roberts (2018)** use the FSS to calculate the mean distance between precipitation features at the neighbourhood where FSS=0.5

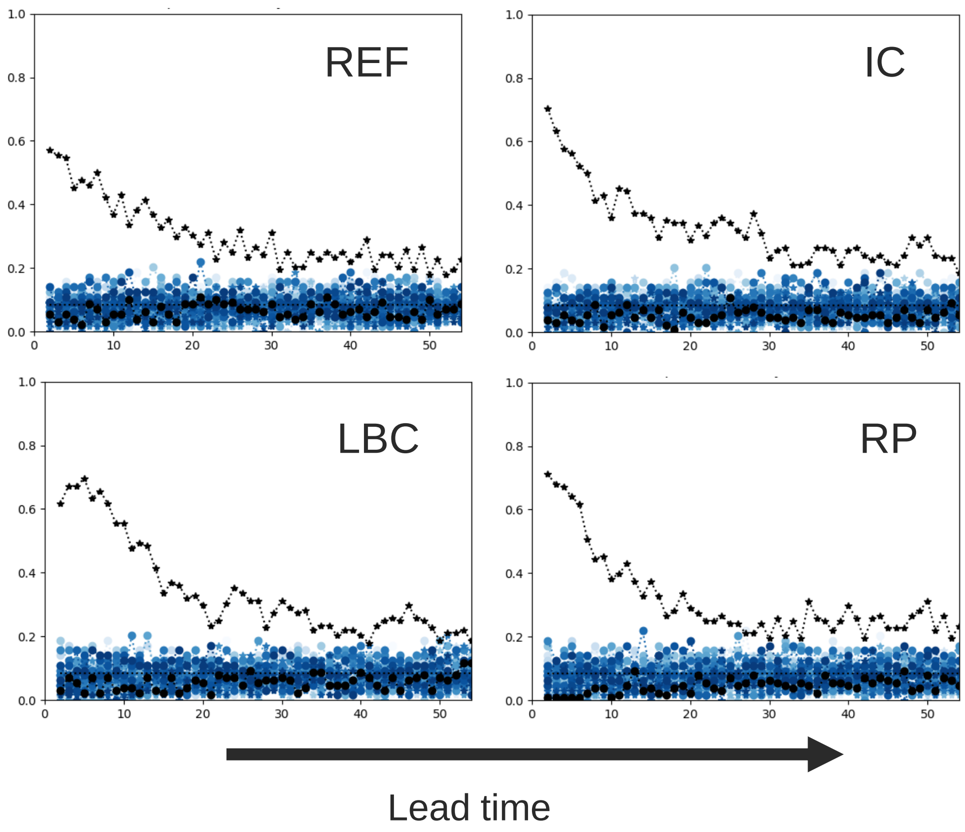
We use this measure to find which ensemble member is the "central member" and which is closest to the observations

Plots show the fraction of times that each member is the central member

The control member is in the centre of the spatial distribution more times than any other member ….

… but is equally likely to be closest to the observations

Fraction of times each member is the central member



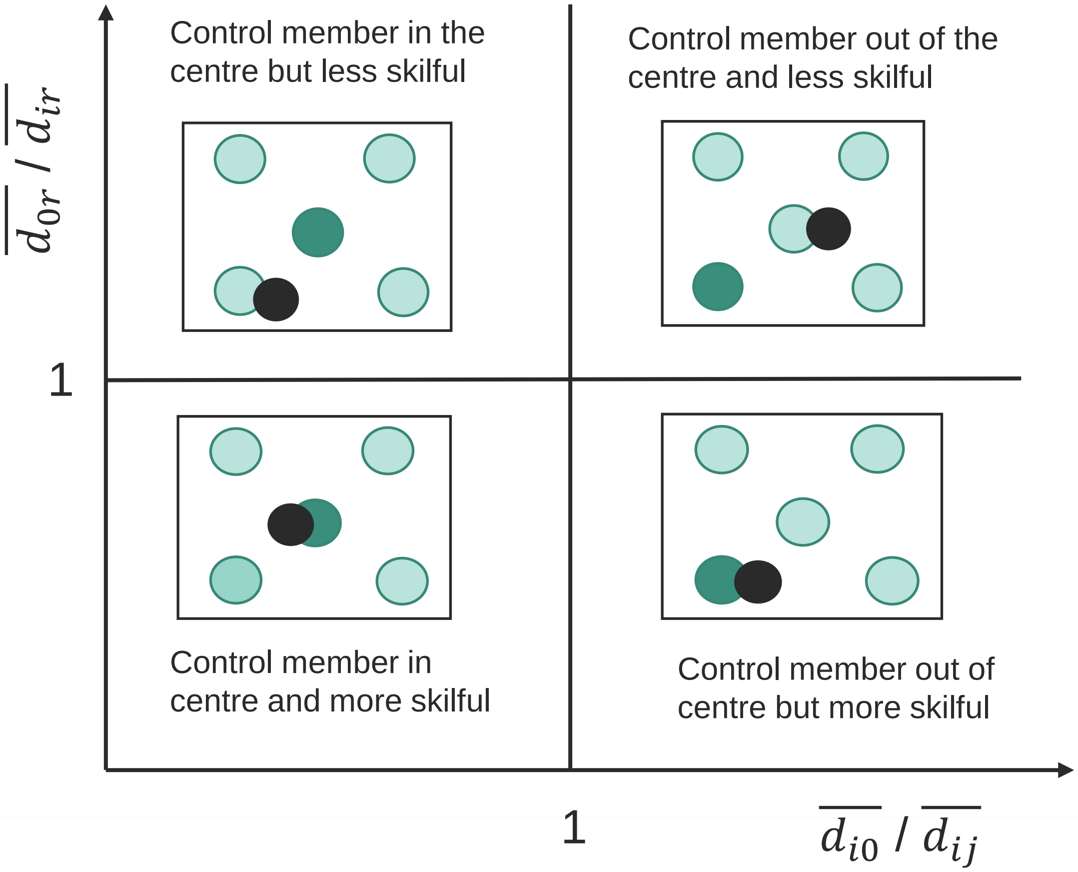
So the control member is most often the central member, but how close are the other members?

We calculate the mean member-control distances: $\overline{d_{i0}}$
and the mean of all member-member distances: $\overline{d_{ij}}$

If $\overline{d_{i0}}$ / $\overline{d_{ij}}$ < 1 then the control member is either in or close to the centre

If $\overline{d_{i0}}$ / $\overline{d_{ij}}$ > 1 then the control member is outside of the centre of the distribution

We compare the mean control member-radar distances, $\overline{d_{0r}}$, with the mean of all member-radar distances, $\overline{d_{ir}}$, in the same way



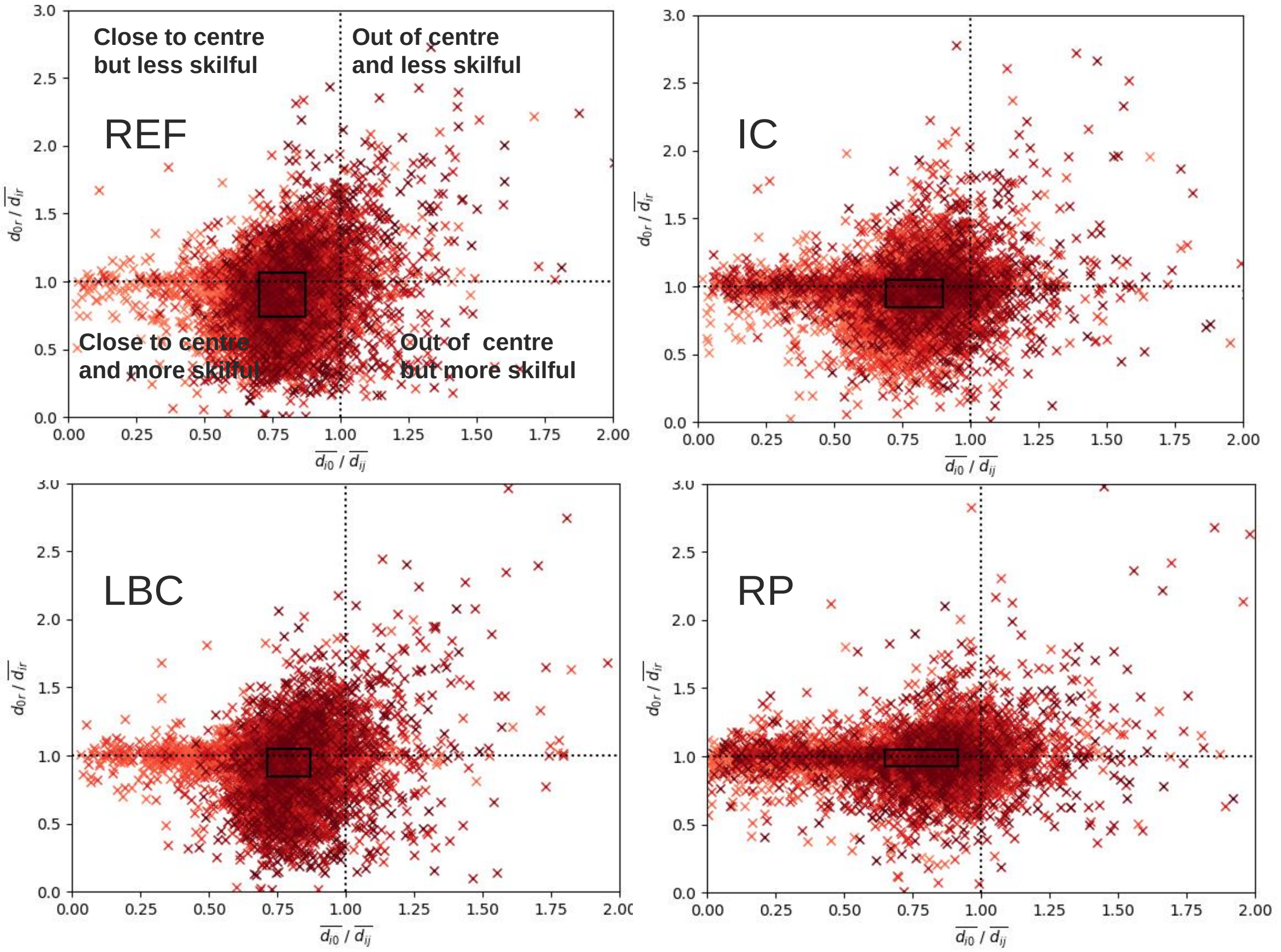
Met Office

These plots show that the control member is closest to the centre most of the time in all of the ensemble experiments

In the REF, LBC and IC ensembles, the control member is also the more skilful the majority of the times

There are still plenty of times when the control member is close to the centre but less skilful – potentially misleading given the ‘double weighting’ of the control member

In the RP ensemble, the control member has a similar chance of being closest to the radar as the other members



Colours are lighter (darker) at earlier (later) lead times

Met Office

Summary

We have taken a critical eye to a research version of MOGREPS-UK with a view to focusing future work.

- All the metrics used here to measure ensemble spread suggest that MOGREPS-UK is under-spread most of the time

- The spread is increasing at a faster rate that the error but has a poor starting point and never quite ‘catches up’

- The IC perturbations fail to upscale from the medium to the larger scales while the stochastic physics has a more local impact on the spatial spread

- All ensembles ‘follow the control member’

Future work

- Focus on improving the initial condition perturbations – do the same problems exist with the operational hourly cycling MOGREPS-UK?

- Think about the interaction between the stochastic physics and initial condition perturbations – could these work together more intelligently?

- Include the ‘central member’ method in future evaluations so we know whether we are improving or worsening this characteristic – does the ensemble still ‘follow the control’ in the hourly cycling configuration?

Met Office

Thank you for listening

Met Office

Extra Slides

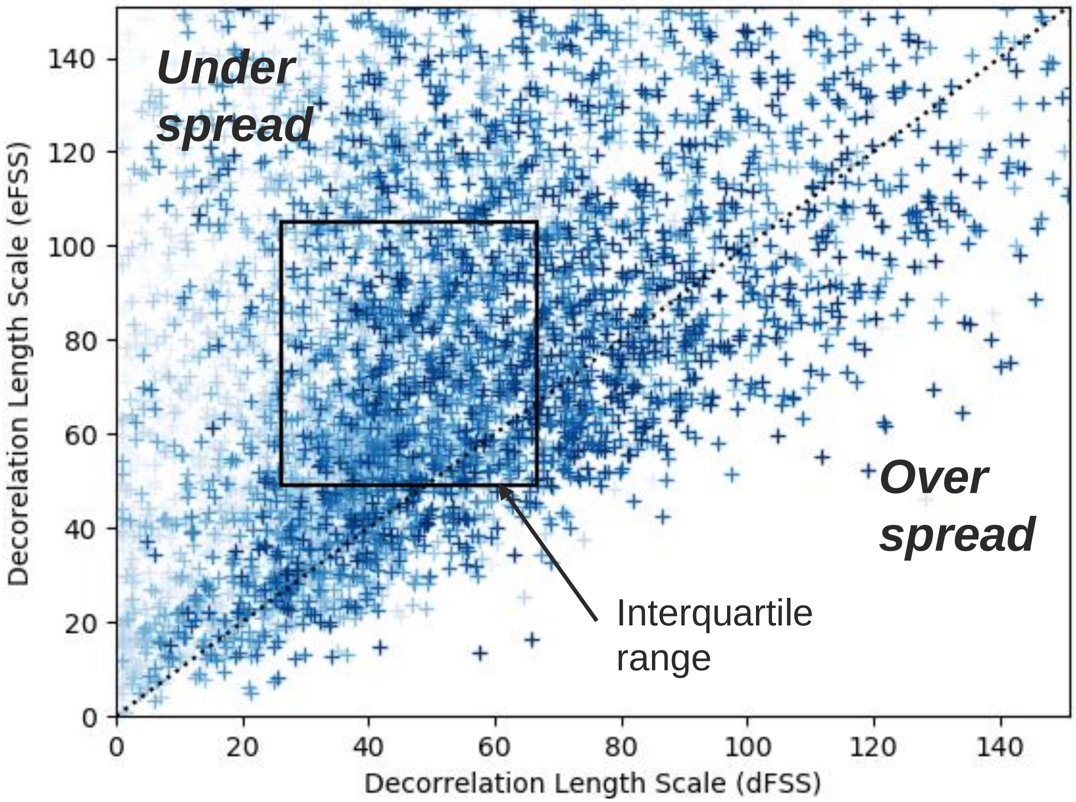
Contents:

- dFSS – eFSS scatter plots for the reference ensemble
- Ensemble SAL

Scatter plots of the scale at which the dFSS and eFSS = 0.5 show how the spatial spread and error compare for every lead time and every cycle

The majority of cycles and lead times are underspread but there are occasions where there is a reasonable match of spread and error

REF ensemble



Colours are lighter (darker) at earlier (later) lead times

Met Office

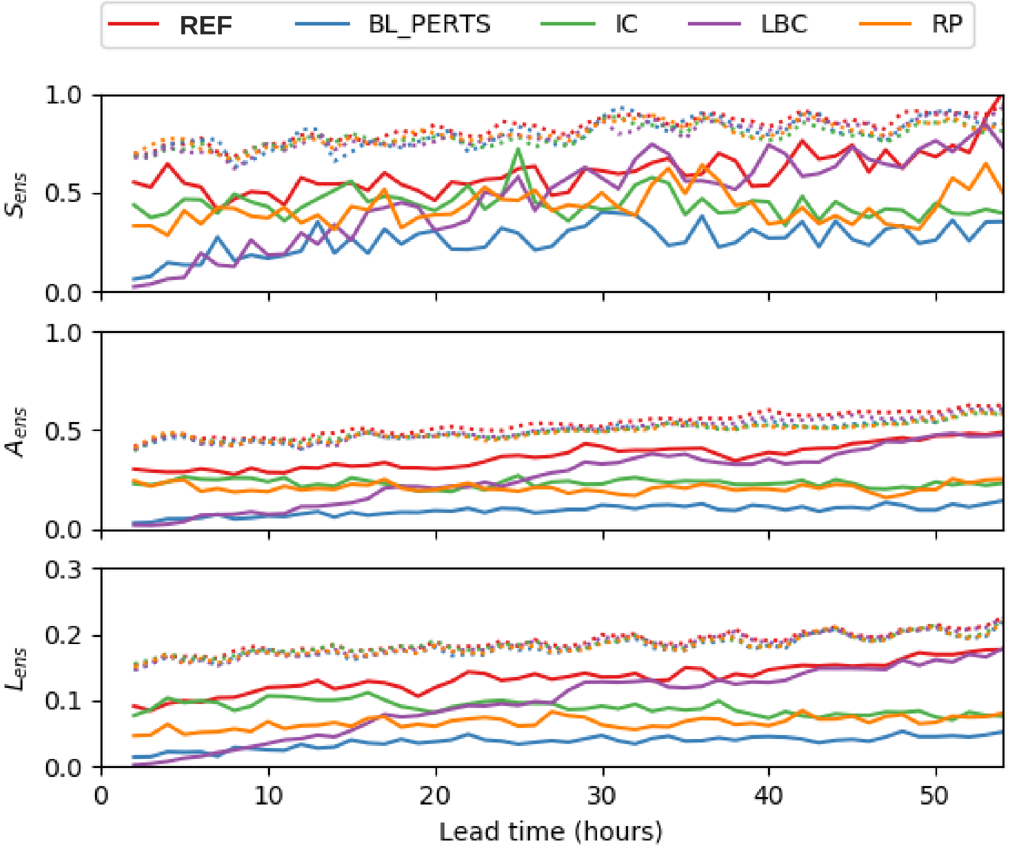
Ensemble SAL

The SAL method compares two fields using the following components:

- Structure (S) – the shape of the objects
- Amplitude (A) – domain averaged precipitation
- Location (L) – the distance between objects

We use an ensemble version of the SAL method to give another measure of spread and error

We calculate the mean of the member-member (spread) and member-radar (error) components and compare the absolute values

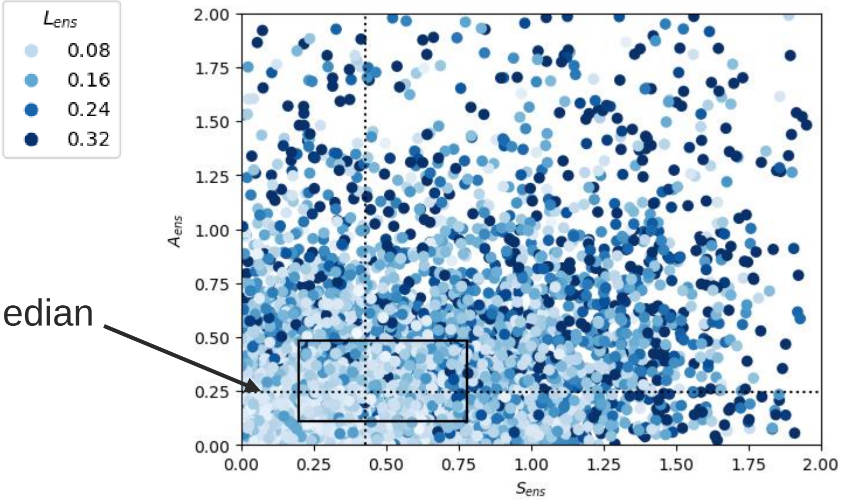


SAL references: Wernli *et al.* 2008, 2009

Ensemble SAL spread and error plots for all cycles and lead times

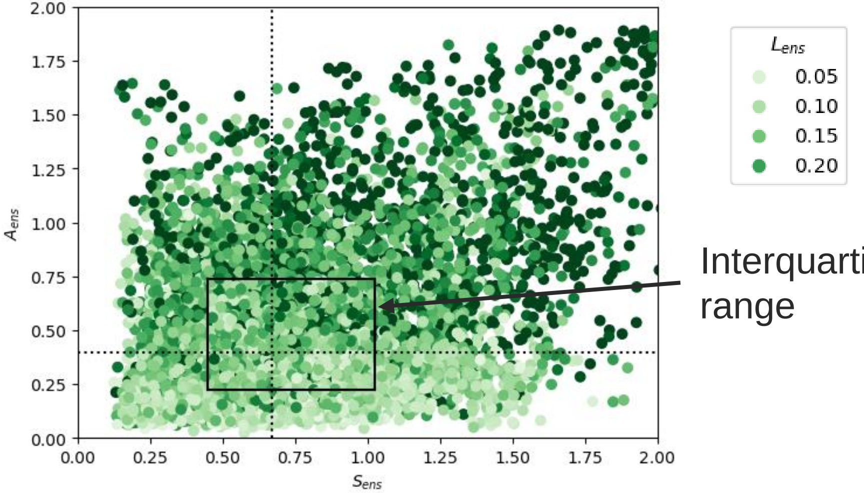
REF ensemble

SAL *spread*



Median

SAL *error*



Interquartile range

There is more variability between the ensemble members and the radar than between the ensemble members themselves

Met Office

Contribution of different sources of uncertainty to the spread of the SAL components

- The LBC ensemble creates the most spread in the structure and amplitude components, particularly when the objects are further apart
- The IC ensemble has slightly more spread in S and A than the RP ensemble but the distribution of points is slightly different
- The BL perts ensemble gives the least amount of spread with more spread in the structural component than the amplitude or location