Assessing the value of clustering convection-permitting ensemble forecasts

Article

Supplemental Material

RESEARCH ARTICLE

Assessing the Value of Clustering Convection-Permitting Ensemble Forecasts

immediate

1 | COMPARISON OF SPATIAL DISTANCE METHODS

In Sec. 2.2.2 of the main text, we used the Precipitation Smoothing Distance (PSD, eq. 1 and 2) to measure the displacements between fields of different members, but we also mentioned the possibility of using other methods. Since the PSD shows large biases between the average displacements for each ensemble, it is instructive to understand whether these biases are a common feature for all smoothing-based displacement metrics.

The simplest of these methods, the FSS Displacement (Skok and Roberts, 2018), operates by finding the neighbourhood at which the FSS (Roberts and Lean, 2008) reaches 0.5, where the FSS is calculated as:

$$\mathrm{MSD}_{(n)}\left(A, B\right) = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[A_{(n)i,j} - B_{(n)i,j}\right]^2 , \tag{1}$$

$$\mathrm{MSD}^{\mathrm{ref}}_{(n)}\left(A, B\right) = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left[A^2_{(n)i,j} + B^2_{(n)i,j}\right] , \tag{2}$$

$$\mathrm{FSS}_{(n)}\left(A, B\right) = 1 - \frac{\mathrm{MSD}_{(n)}\left(A, B\right)}{\mathrm{MSD}^{\mathrm{ref}}_{(n)}\left(A, B\right)} , \tag{3}$$

where MSD is the mean-square difference between features fields, $A_{(n)}$ and $B_{(n)}$ that have been smoothed by a neighbourhood of grid size n, MSD_{ref} is the low-skill climatalogical baseline, and N_x, N_y are the number of grid points in the x and y directions. An FSS of unity indicates identical fractions fields, while a score of zero indicates fields that are completely mismatched. Then, the neighbourhood at which the FSS reaches 0.5, $n_{\mathrm{FSS}=0.5}$, is used to calculate the FSS Displacement (Skok and Roberts, 2018), $DFSS$ as:

$$\mathrm{DFSS} = \left(1 - \mathrm{FSS}_{(n_{\mathrm{FSS}=0.5})}\right) \cdot n_{\mathrm{FSS}=0.5} , \tag{4}$$

which can often be approximated to simply $\mathrm{DFSS} = n_{\mathrm{FSS}=0.5}/2$ since the FSS is close to 0.5. In this form, the FSS Displacement is the average grid-point distance between features in fields $A_{(n)}$ and $B_{(n)}$, and can easily be converted to a physical distance by multiplying by the grid size.

However, Skok and Roberts (2018) found that the displacement between overlapping features could be severely underestimated with the formulation outlined in equation 4. It was found that a simple overlap-adjustment corrects this underestimation by setting any overlaps between members equal to 0 in both fields when calculating distances using any applied smoothing (i.e., for $n > 1$). The overlap-adjusted FSS Displacement is then calculated as:

$$\mathrm{DFSS_{OA}} = (1 - \mathrm{FSS}_{(n=1)}) \cdot \mathrm{DFSS}\,, \tag{5}$$

where $\mathrm{FSS}_{(n=1)}$ is the FSS applied at the gridscale (where overlaps must be preserved to ensure nonzero scores). The DFSS in Equation 5 has been calculated as in equation 4 but with overlaps removed in the binary feature fields used to create the smoothed $A_{(n)}$ and $B_{(n)}$ fields. A full procedure and set of considerations for using the overlap-adjusted DFSS can be found at the end of Skok and Roberts (2018).

While all methods are functionally similar, the main differences between the two DFSS formulations and the PSD are:

1. Unlike the DFSS, the PSD does not require binary data to calculate displacements. Instead, each field is normalised by its area-averaged value to account for intensity biases. However, using non-binary data introduces intensity weighting to the calculation, meaning that the score can no longer be interpreted as a purely spatial displacement. Using the PSD with non-binary fields can be advantageous since the score will provide greater weight to displacements between areas of more impactful weather. Here, however, we are only interested in offset between features, and therefore we will continue to use binary input fields.
2. The PSD uses a circular neighbourhood rather than a (typically) square one as used in the DFSS. While there is no fundamental reason why the DFSS cannot also use a circular neighbourhood, square neighbourhoods are easier to implement. Either way, it is not believed that the neighbourhood shape has a meaningful impact on scores (Roberts and Lean, 2008).
3. Not only are the input fields adjusted to remove overlaps as with the $\mathrm{DFSS_{OA}}$, the PSD explicitly takes account of the overlap fraction, Q, within both the similarity score and the distance estimations.
4. The PSD does not normalise its similarity score by a low-skill climatology as with the DFSS.
5. The PSD compares fields using a mean-absolute difference rather than a mean-squared difference, since this was found to produce more accurate estimates of displacement.

Figure S1 shows the average distances provided by all three methods over 12 h leadtime windows. For references, MOGREPS-UK grid spacing is 2.2 km while MOGREPS-G grid spacing is approximately 20 km over the region tested. All methods show that the distances estimated by MOGREPS-G are larger than those for MOGREPS-UK on their native grids. This bias is likely reflective of the additional convective-scale detail introduced in MOGREPS-UK, where the floor for feature distances is smaller than for the coarser global ensemble. The increased similarity between MOGREPS-G trends and those of MOGREPS-UK coarse-grained onto the same grid is supportive of this conclusion, although some differences remain.

Comparisons of the different methods shows that the FSS Displacement displays consistently smaller distances, consistent with the expectation that including overlapping features reduces estimates (Skok and Roberts, 2018). Note that the overlap-adjusted FSS Distance for MOGREPS-UK was not calculated due to the additional computational time complexity required to reach the FSS threshold. While procedures do exist to reduce the computational costs associated with neighbourhooding (e.g. Faggian et al., 2015), we did not anticipate this would yield important insights.

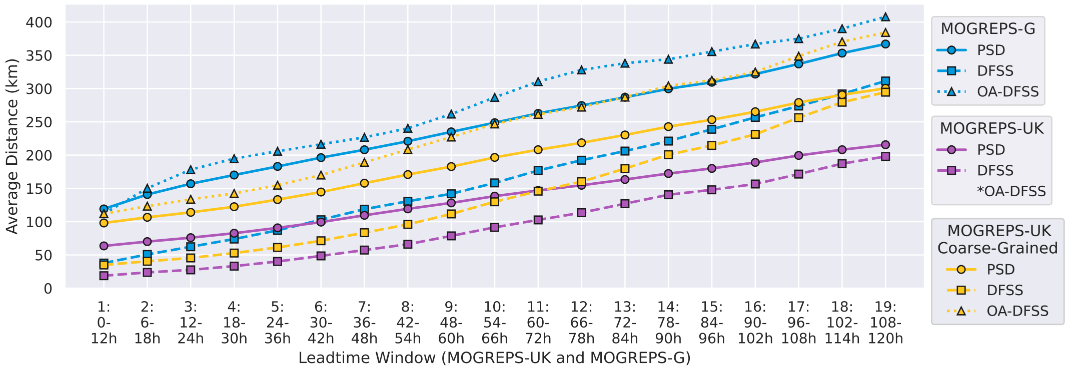

FIGURE S1 Average distances between all methods for 12 h leadtime windows for MOGREPS-G, MOGREPS-UK and MOGREPS-UK coarse-grained onto the MOGREPS-G grid. Coarse-graining was performed using a nearest-neighbour algorithm that masked extrapolated points. Overlap-Adjusted FSS Displacements were not able to be calculated for MOGREPS-UK on its native grid due to the increased time complexity of using this method. Averages were calculated over all forecasts initialised in July 2023.

In summary, it is clear there are large distance biases between the ensembles no matter the method used. These differences are important to consider when comparing cluster outputs between the two ensembles, as well as for future work that may wish to use these methods. Since all methods tested here show the same biases, the choice was made to continue with the PSD in the main article.

2 | CONVECTIVE CASE STUDY

For completeness, Fig. S2 shows the full case study period as discussed in Section 5 of the main text. Grey backgrounds are used to visually delineate different leadtimes.

3 | MULTI-SCALE CASE STUDY

Statistical analysis from the main text show that there is little difference in the composition and skill of clusters from convection-permitting (CPE) and global ensembles when applied over the UK domain. These findings indicate that the presence of convective-scale detail in the CPE has little effect on the construction of clusters, meaning that the clusters are mostly sensitive to the synoptic-scale variability common between the two ensembles. To show this sensitivity explicitly, this section presents a case study using a forecast containing hazardous weather across synoptic and convective scales.

The event presented in this section is from 15 July 2023 0000Z-0300Z. The forecast of this event was initialised 13 July 2023 0000Z, with leadtimes T+48-51 h. For ease of interpretation, this section will only consider a cluster window of 3 h, comprising of a single three-hourly accumulation timestep. During this period, heavy precipitation associated with an active south-westerly front impacted parts of Scotland and Ireland. At these leadtimes, there is some uncertainty in the position and orientation of this front as well as its associated impacts. In the wake of this front, convection pushed in from the southwest and brought more localised impacts to parts of Wales and southwest

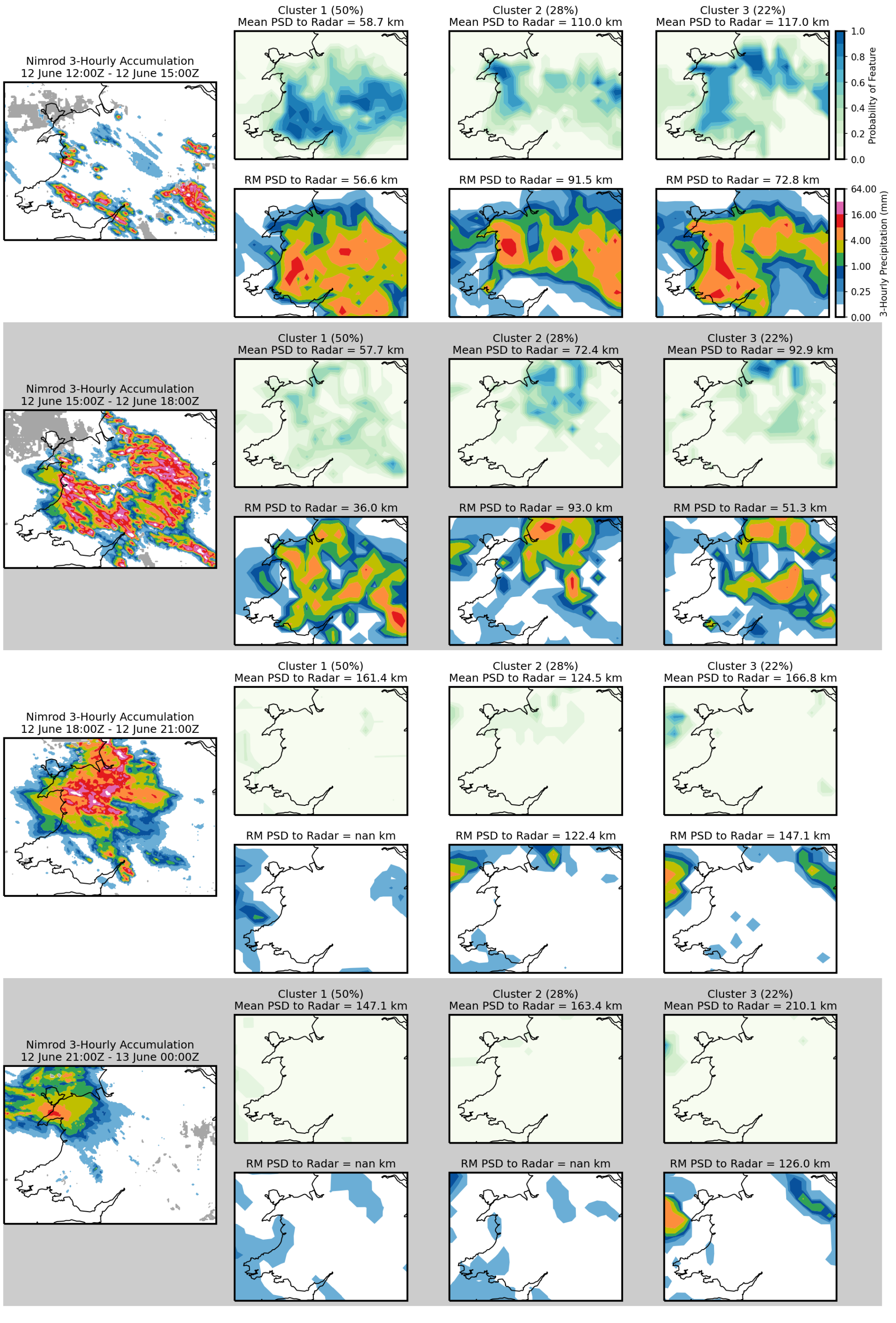

FIGURE S2 MOGREPS-G clusters for the full convective case study period discussed in Section 5 of the main text, and following the same format as Figs 10, 11, 12.

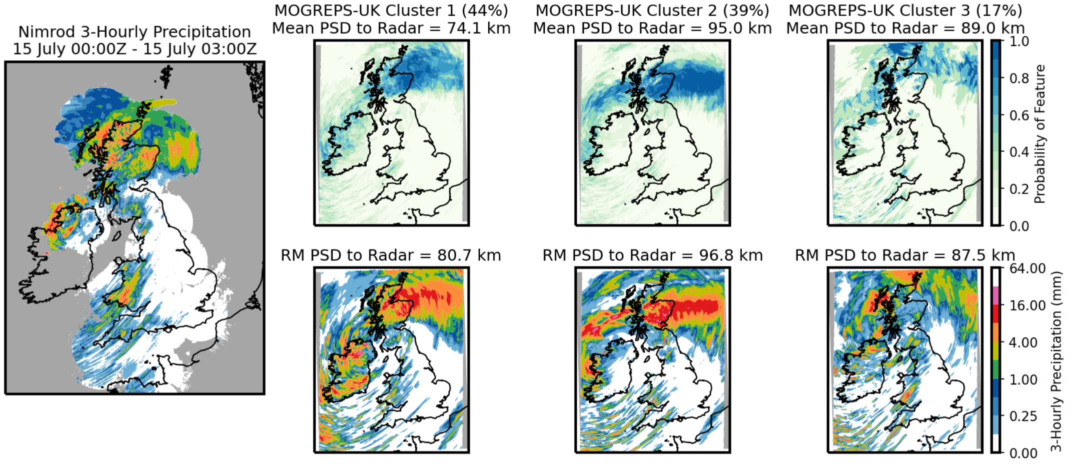

FIGURE S3 MOGREPS-UK clusters for a case demonstrating the sensitivity of clustering to different spatial scales. Follows the same format as Fig. S2 and other figures presented in Section 5 of the main text. Grey regions of the radar are masked due to insufficient returns in at least one hour of the three-hourly period (where insufficient returns is defined as less than 50 minutes of available data). This case was clustered using a 2 mm threshold and with a single 3 h timestep, initialised on 13 July 2023 with leadtime T+48-51 h, valid 15 July 2023 00:00Z - 03:00Z.

England. As with most forecasts of convection, there is uncertainty in the exact positioning of these convective-scale features, with some members showing large totals over Wales, and others keeping the convection from reaching land (not shown). Since there is uncertainty in the location of both the frontal and convective-scale features, this case provides an opportunity to demonstrate the scale sensitivity of the clustering procedure.

Figure S3 shows MOGREPS-UK clusters for this event using $k = 3$, which was subjectively identified as representing all possible scenarios without any scenario being repeated. We do not consider MOGREPS-G clusters for this case study, since we are interested in understanding the influence of convective-scale detail on clustering. For this reason, we also use a 2 mm absolute threshold to cluster the ensemble, since the threshold associated with the 90th centile was too large (4.5 mm) to provide sufficient coverage for convective-scale features. To capture all of the hazardous weather, clustering was run over the full UK domain.

Both the feature probability density and representative member plots in Fig. S3 show that the large-scale variability is well represented in these clusters. Frontal precipitation is noticeably less zonal over Scotland in cluster 1 than in cluster 2. These clusters also differ in their prediction of impacts across Ireland, with cluster 1 showing a greater likelihood than cluster 2. In contrast, cluster 3 shows the potential for some impacts across Northern Ireland, but with a much weaker frontal band of precipitation and located further north. Therefore, all three clusters show distinct scenarios in the positioning of large-scale features.

In contrast, the clusters have not represented the variability in the smaller-scale convection affecting the south-west of the UK. Each feature probability plot shows a similar chance of convection affecting similar regions of Wales, despite broader uncertainty existing within the ensemble. The representative members for these clusters demonstrate slightly more diversity, as may be anticipated from physical fields rather than probabilistic fields. However, these members do not contain the full range of solutions observed in the ensemble, since none of the representative members show a scenario that keeps the convection away from land. Therefore, the clustering workflow appears to be less sensitive to differences in the placement of convection.

In summary, this case has demonstrated that the small-scale convective detail contained in MOGREPS-UK has little effect on the clusters compared to the dominating influence of the variability in frontal positioning. This finding reinforces the statistical results found in the main text, and highlights the need to consider the scale of the hazardous weather when deciding on the clustering domain. This finding is also consistent with previous interpretations of the behaviour of spatial verification methods, which shows that they are most sensitive to the largest-scales contained in the domain (Roberts and Lean, 2008).

references

Faggian, N., Roux, B., Steinle, P. and Ebert, E. (2015) Fast calculation of the Fractions Skill Score. *Mausam*, **66**, 457–466.

Roberts, N. M. and Lean, H. W. (2008) Scale-Selective Verification of Rainfall Accumulations from High-Resolution Forecasts of Convective Events. *Monthly Weather Review*, **136**, 78–97. URL: https://journals.ametsoc.org/view/journals/mwre/136/1/2007mwr2123.1.xml. Publisher: American Meteorological Society Section: Monthly Weather Review.

Skok, G. and Roberts, N. (2018) Estimating the displacement in precipitation forecasts using the Fractions Skill Score. *Quarterly Journal of the Royal Meteorological Society*, **144**, 414–425. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.3212. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3212.