Evaluating different convective-scale ensembles over the UK: preliminary results

Carlo Cafaro

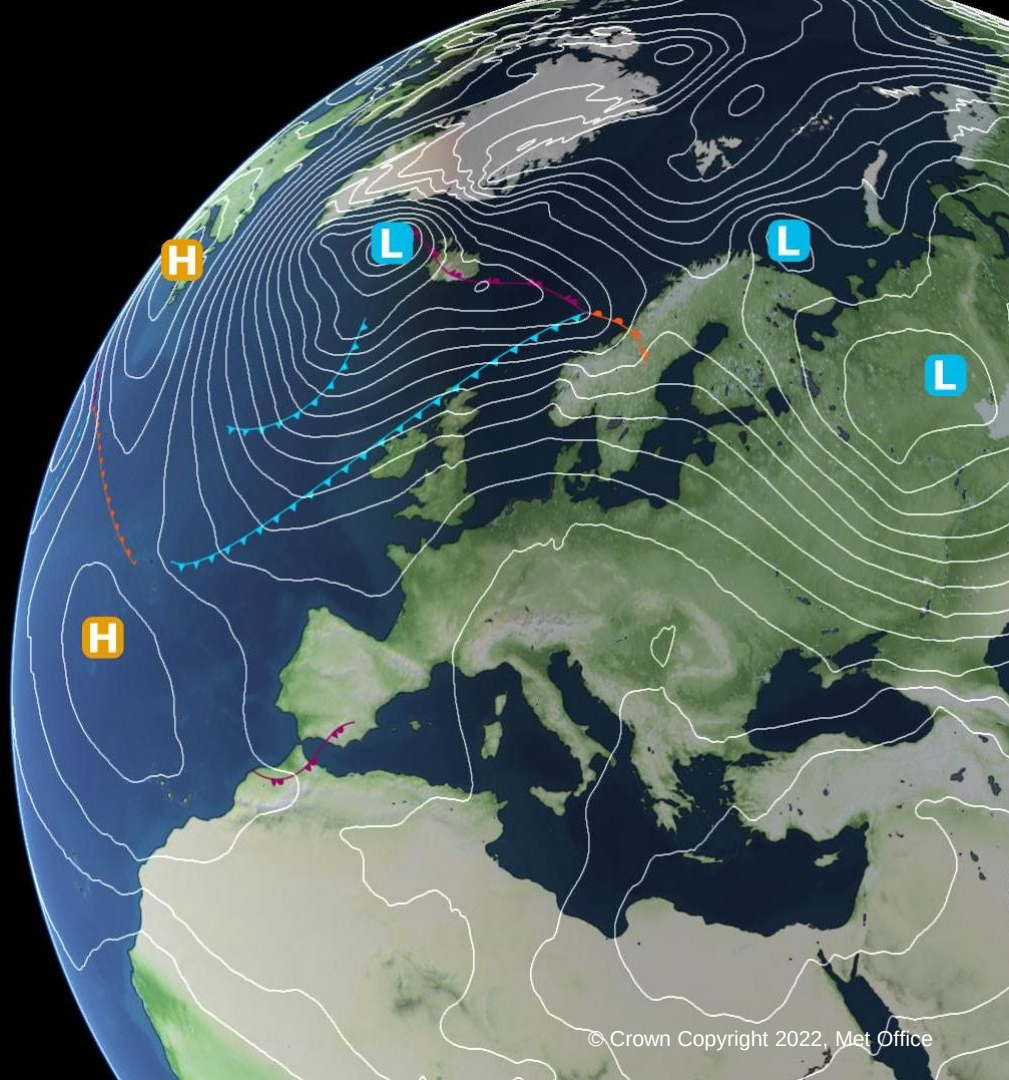
Research to Operations, Met Office

Thanks to *Anne Mccabe, Aurore Porson, Stuart Webster, Nigel Roberts, David Flack*

44th EWGLAM - 29th SRNWP Workshop

26-29 September 2022

Royal Library of Belgium, Brussels

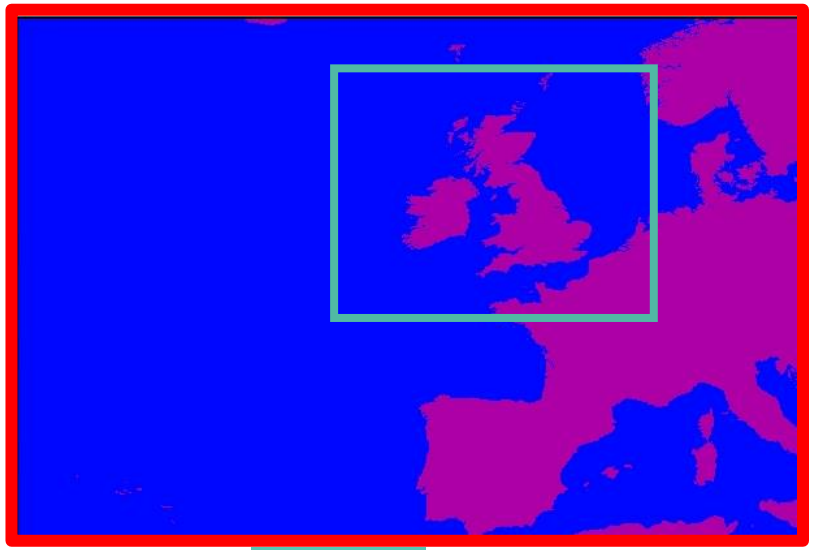


www.metoffice.gov.uk

Motivations

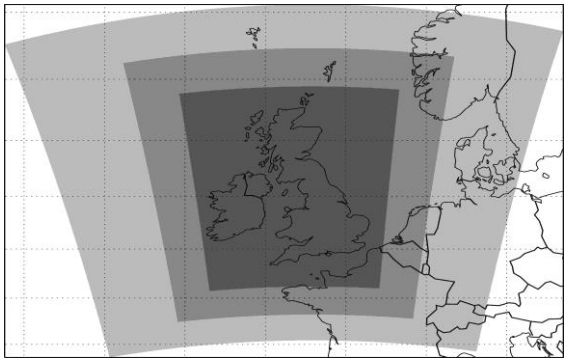
- The UK regional ensemble (MOGREPS-UK), has been operational since 2012 and since the last update (*Porson et al., 2020*) **has been running as hourly time-lagged ensemble** (‘hourly cycling’).
- An usual complaint by the operational forecasters is that **MOGREPS-UK lacks spread** and follows the deterministic forecast too closely. Preliminary study confirm the forecasters are right (*Mccabe et al., 2020, internal report*).
- How to **improve the spread** ? This work is part of larger project with the aim to tackle the lack of ensemble spread (e.g. last year talk by A. Mccabe)
  Here **we test different new configurations** of the UK regional ensemble, to **explore the sensitivity of the ensemble spread** to either the domain size, science configuration and the impact of the parent global ensemble (‘downscaling’).

Experiments domains



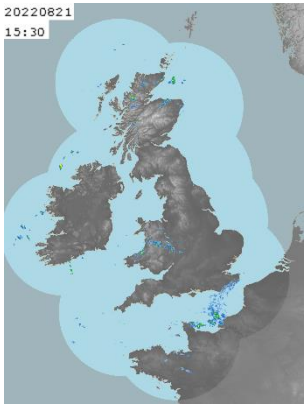
"UK small"

"UK big"



MOGREPS-UK

Hagelin et al., 2017



Verification (radar)

Experiments

Experiments	Domain	Science configuration	LBC & IC
Hourly cycling	UK small	RAL2	MOGREPS-UK
Downscaling – RAL2	UK small	RAL2	MOGREPS-G
Downscaling – RAL3	UK small	RAL3	MOGREPS-G
Downscaling - big	UK big	RAL2	MOGREPS-G

Simulations have been run on a non-rotated grid with fixed resolution ~4km, 4 times per day (0000, 0600…), for 48h with 18 members

For more info about RAL3 please have a look at the talk by Anke Finnenkoetter on 28th Sept in the Upper Air Physics session

Verification methodology

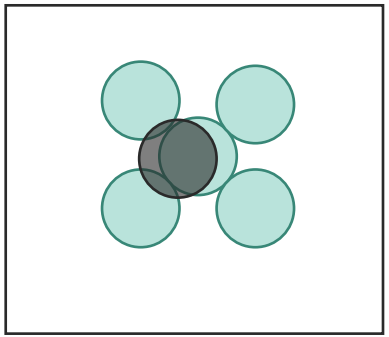
- **Spatial spread/skill relationship** using the Fractions Skill Score (FSS; *Roberts and Lean, 2008*), for **precipitations** forecast using percentile thresholds.

- **Error FSS (eFSS)** is calculated to measure the **skill**, for each member-obs pair and then averaging. **Dispersion FSS (dFSS)** is calculated to measure the spatial agreement (or the **spread**) of the members, for each member-member pair and then averaged (*Dey et al., 2014*)

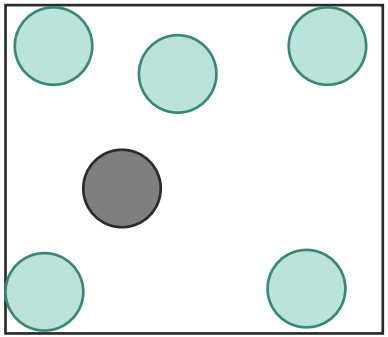
- **eFSS** and **dFSS** both range in [0,1]. Ideally we want eFSS=dFSS=1 (high skill, low [high] spread [agreement]), or at least eFSS=dFSS.

Spatial spread/skill

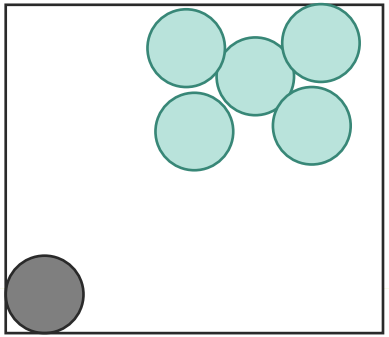
observation **ensemble member**



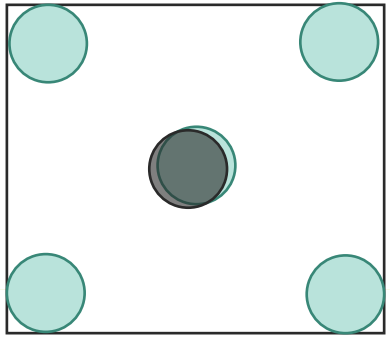
high eFSS (high skill)
high dFSS (low spread)



lower eFSS (higher skill)
lower dFSS (higher spread)



low eFSS (low skill)
high dFSS (low spread)



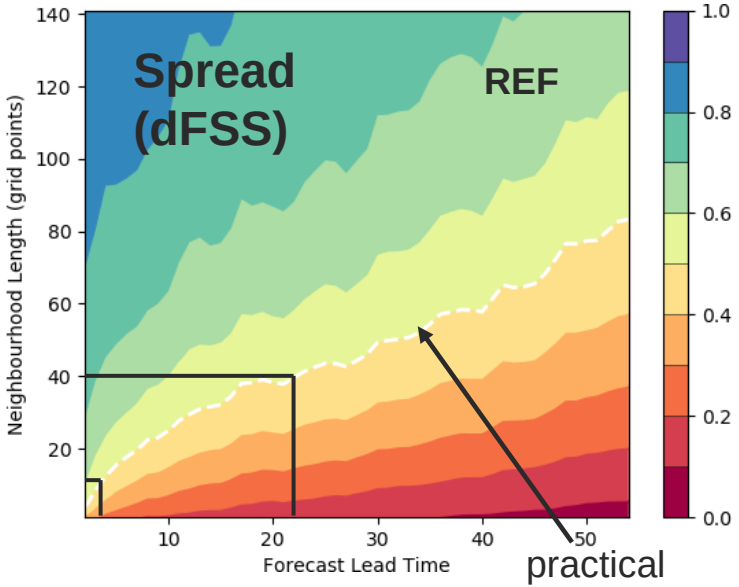
high eFSS (high skill)
low dFSS (high spread)

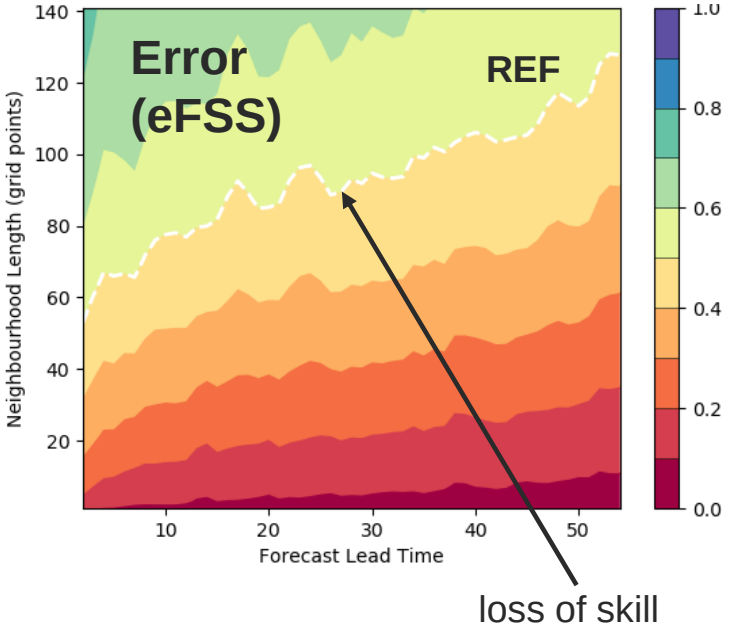
Adapted from Anne Mccabe

Met Office

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)





Courtesy of Anne Mccabe

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

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

Example of eFSS/dFSS metrics

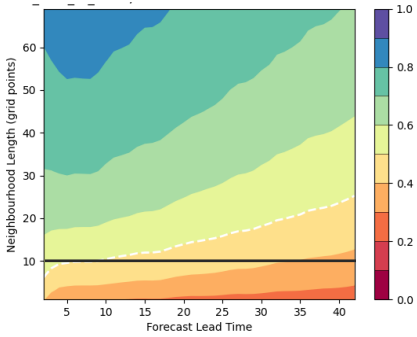
Downscaling – RAL2

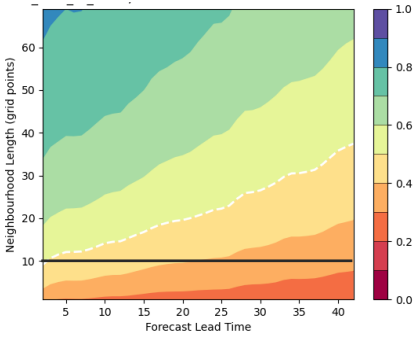
Hourly cycling

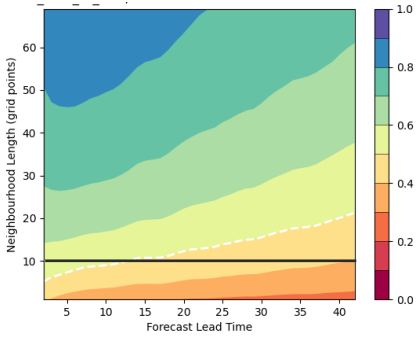
Downscaling – RAL3

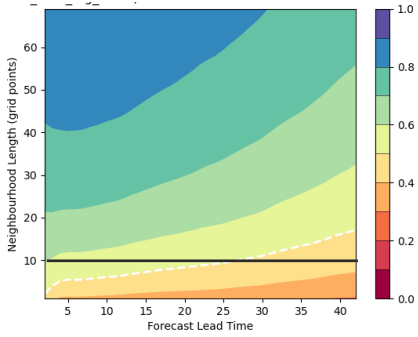
Downscaling – big

dFSS

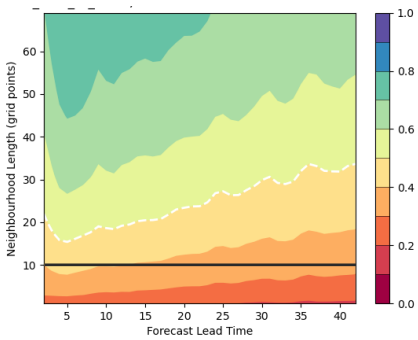


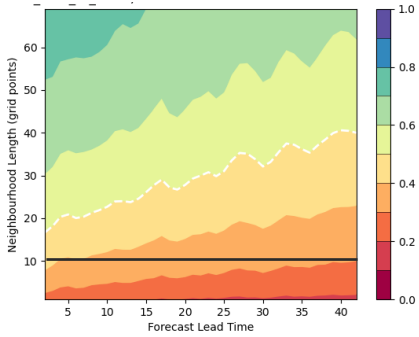


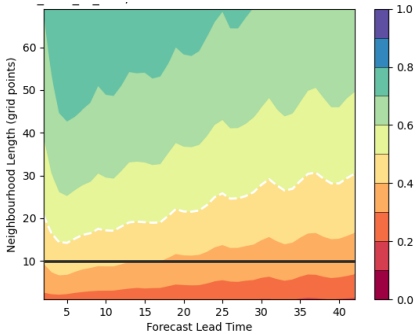


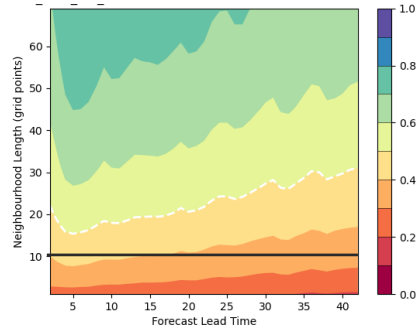


eFSS





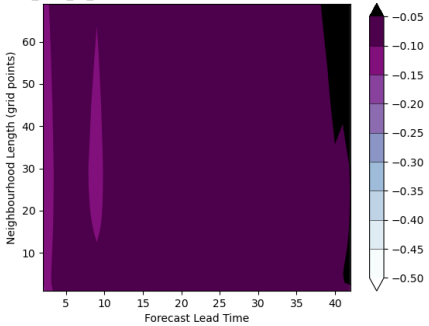




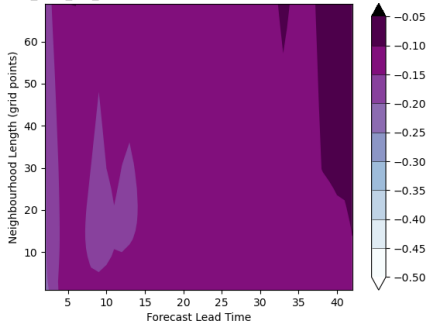
Dataset: 9th – 15th June, 24th June – 4th July, 20th July – 21st July, 15th -17th August 2022 **95th centile**

eFSS minus dFSS results

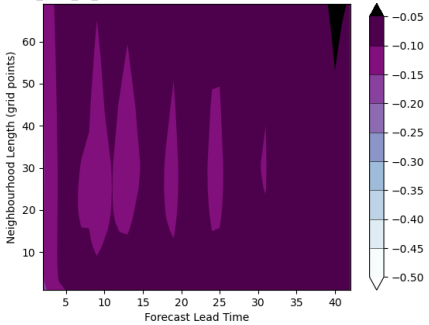
Downscaling – RAL2



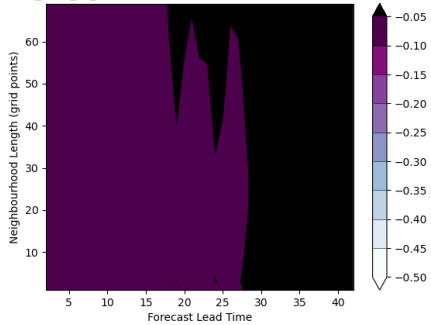
Downscaling – big



Downscaling – RAL3



Hourly cycling

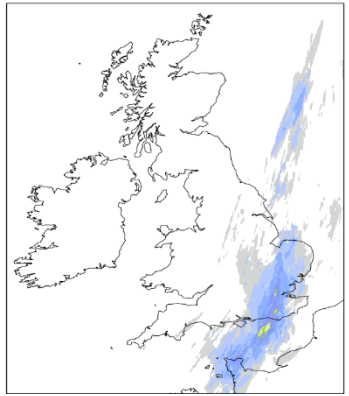


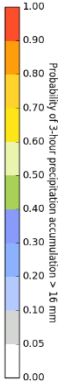
- Negative values mean dFSS>eFSS -> under-spread
- Hourly cycling has better spread/skill relationship
- Generally the difference reduces with lead time

Dataset: 9th – 15th June, 24th June – 4th July, 20th July – 21st July, 15th -17th August **95th centile**

Case study maps

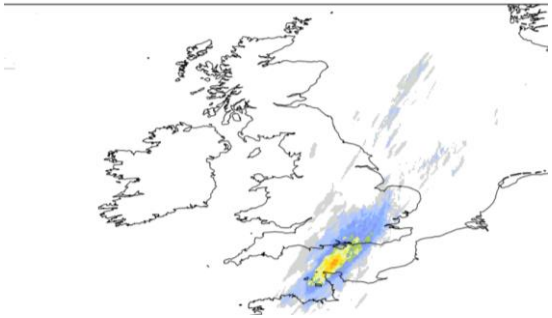
Hourly MOGREPS-UK
2022/08/25 0300Z to 2022/08/25 0600Z, T+6.0 to T+9.0, from 2022/08/24 2100Z



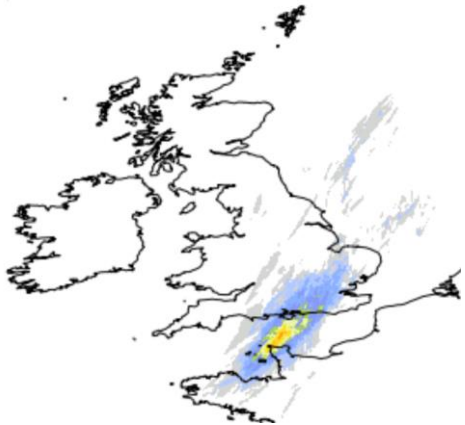


MOGREPS-UK

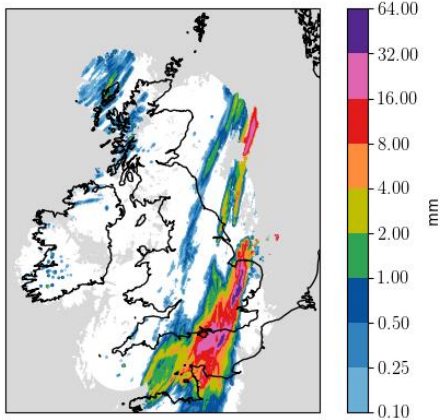
Downscaling – RAL2

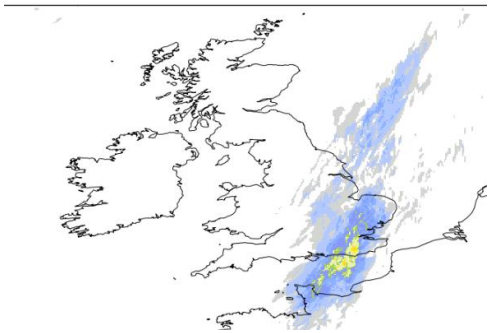


Downscaling - big



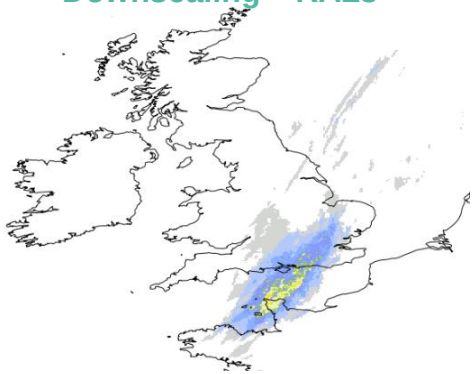
Radar accumulation 0600Z





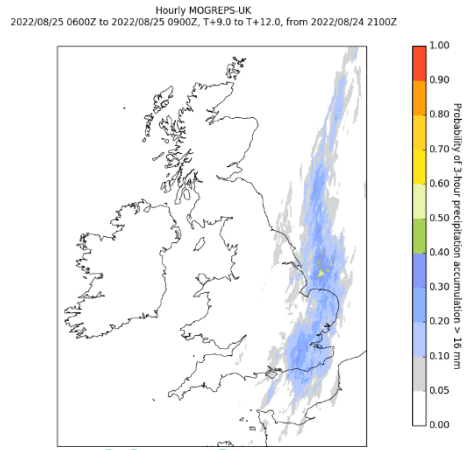
Hourly cycling

Downscaling – RAL3



Case 25th August 2022, init 24th August 18Z **Prob 3h acc >16mm**

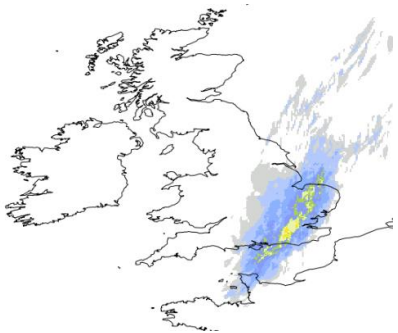
Case study maps



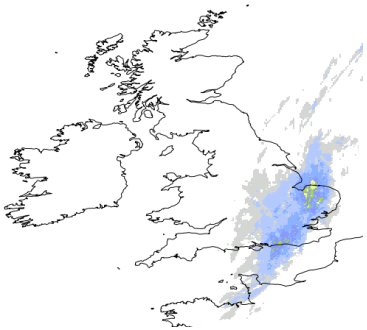
MOGREPS-UK

Hourly cycling

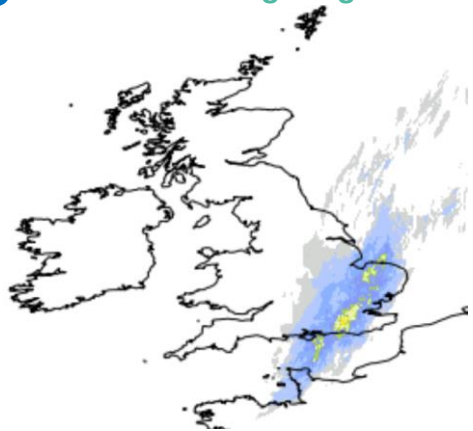
Downscaling – RAL2



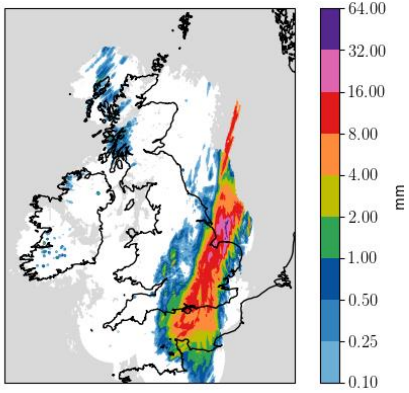
Downscaling – RAL3



Downscaling - big

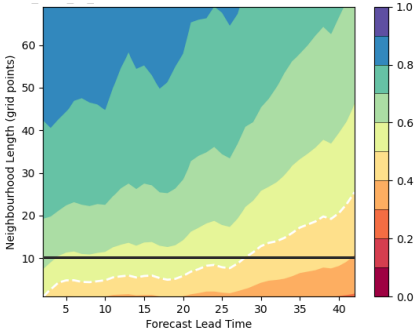
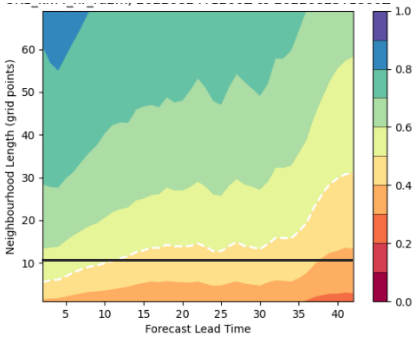
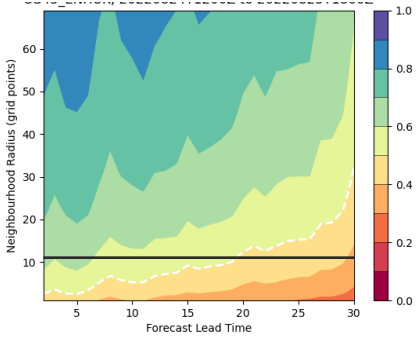
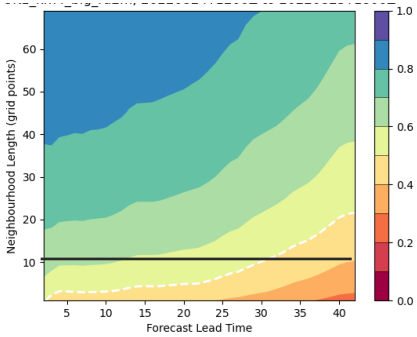
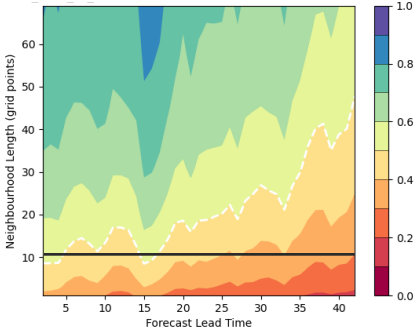
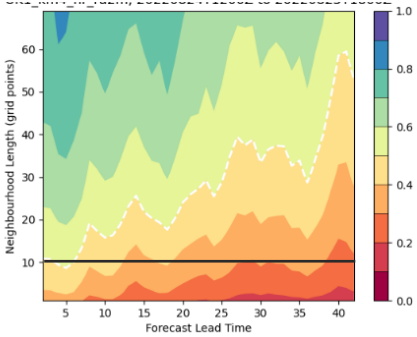
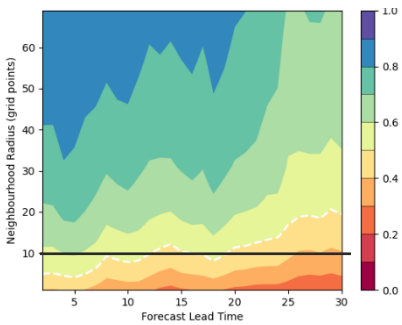
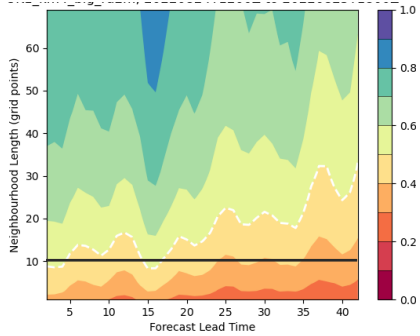


Radar accumulation 0900Z



Case 25th August 2022, init 24th August 18Z **Prob 3h acc >16mm**

eFSS/dFSS results

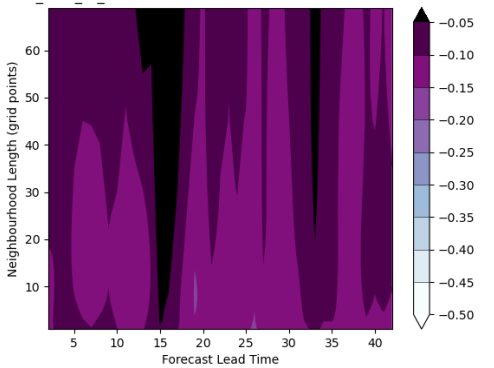
	Downscaling – RAL2	Hourly cycling	MOGREPS-UK	Downscaling – big
dFSS				
eFSS				

Dataset: 24th August 2022 12Z

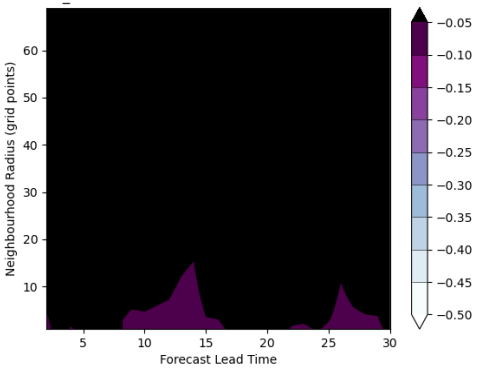
95th percentile

eFSS minus dFSS results

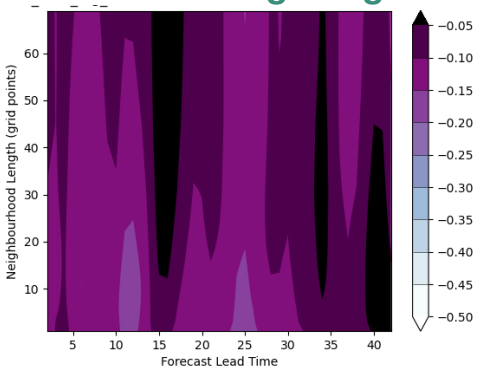
Downscaling – RAL2



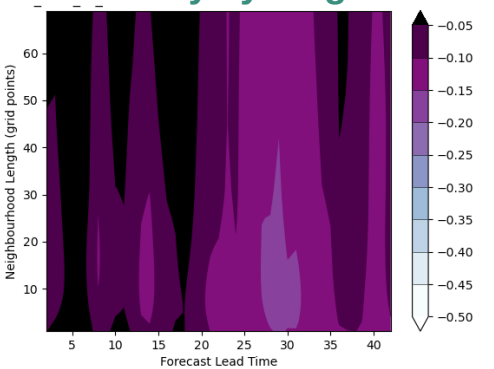
MOGREPS-UK



Downscaling – big



Hourly cycling



- MOGREPS-UK has the best spread/skill relationsnhip

Dataset: 24th August 2022 12Z **95th percentile**

Summary

- All the experiments show **similar eFSS/dFSS results**, with the downscaling ones with slightly higher eFSS but lower dFSS, so **hourly cycling has a better eFSS-dFSS balance** for the **95th percentile** (99th percentile had some opposite results, not shown). RAL3 did not seem to have a significant impact with respect to RAL2.
- **Big domain** seemed to have more impact, with **higher eFSS** but **lower dFSS** than the small domain.

- Skill (eFSS) seem to be lost more quickly in the hourly cycling than in the downscaling exp, and predictability (dFSS) too. **Spin-up** quite evident in the eFSS for the downscaling exp, but not in the hourly cycling.e

- Which experiments are better ? Hard to tell as limited number of cases so far and therefore these are only preliminary results. Other complementary ensemble metrics need to be computed for a longer period.

Future work

- Running the experiments in real-time for the current period.

- Extend the possibility to run the **same experiments for past case studies**. This will also enable to stratify the verification results for **different weather regimes**, to differentiate cases with strong/weak large-scale forcing.

- Calculate the contribution of **IC, LBC, RP scheme and BL perturbations** for the different experiments (cfr Anne Mccabe talk last year)

- Calculate **other ensemble verification metrics** to evaluate probabilistic forecasts as well (Brier Score, FSS applied to probabilities, etc.) to help identify which ensemble would be better for generating precipitation probabilistic forecasts.

- Running **downscaling with ECMWF** BCs & ICs

Thank you for listening

Any question ?

E-mail:
carlo.cafaro@metoffice.gov.uk

One Chance Left · GreenFutures (exeter.ac.uk)

- Extra slides