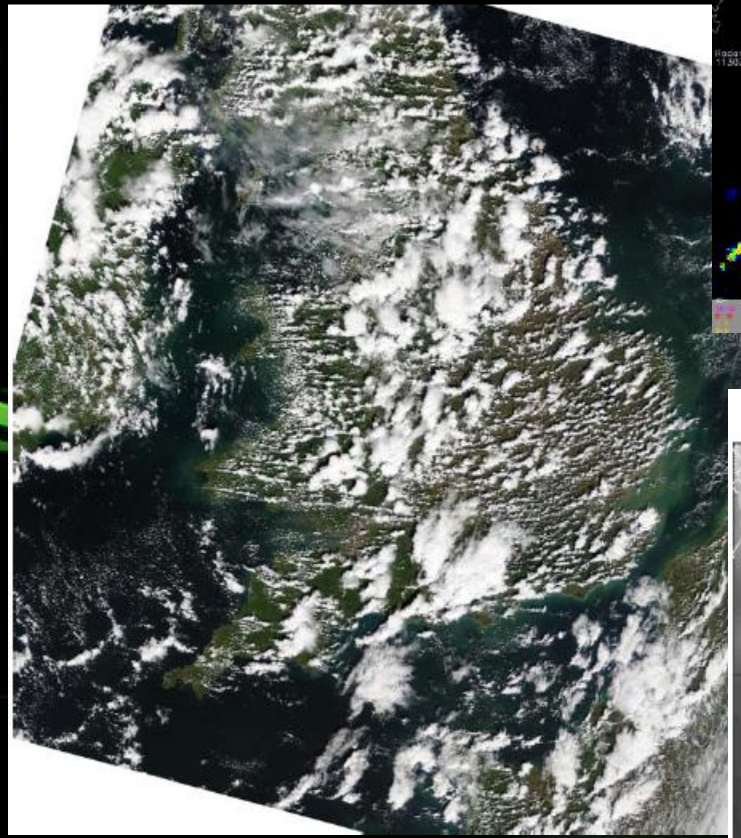
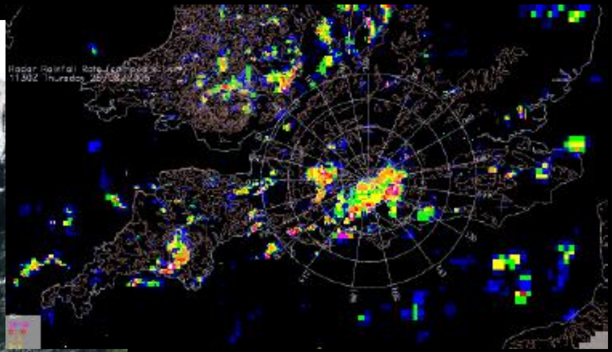
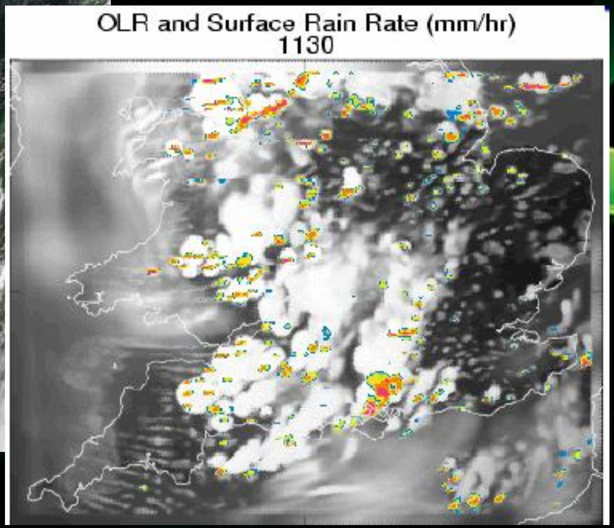






Applying a single-observation-forecast neighbourhood framework to the verification of km-scale NWP

Marion Mittermaier

Manager: Novel verification methods and diagnostics, Weather Science

Background

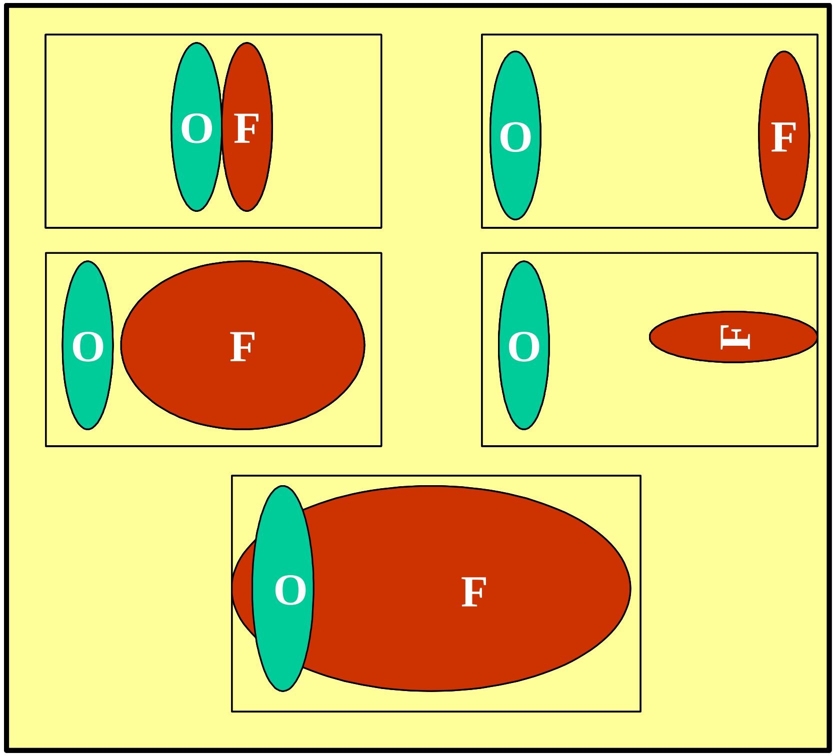
- Traditional metrics can be misleading → trust in objective results undermined, especially for testing model changes.

- Representativeness of observations and model grid values → implications for highly localised events.

- Lack of predictability and rapid error growth at km-scale → impact on perceived skill.

- Difference between grid scale (Δx) and model resolution ($y * \Delta x$, typically $y >= 4$) → now even more reason that km-scale model forecasts must be treated differently (probabilistically) for product generation and verification.

The double penalty



Closeness not rewarded

Detail is penalised unless exactly correct

- higher resolution is more detailed!

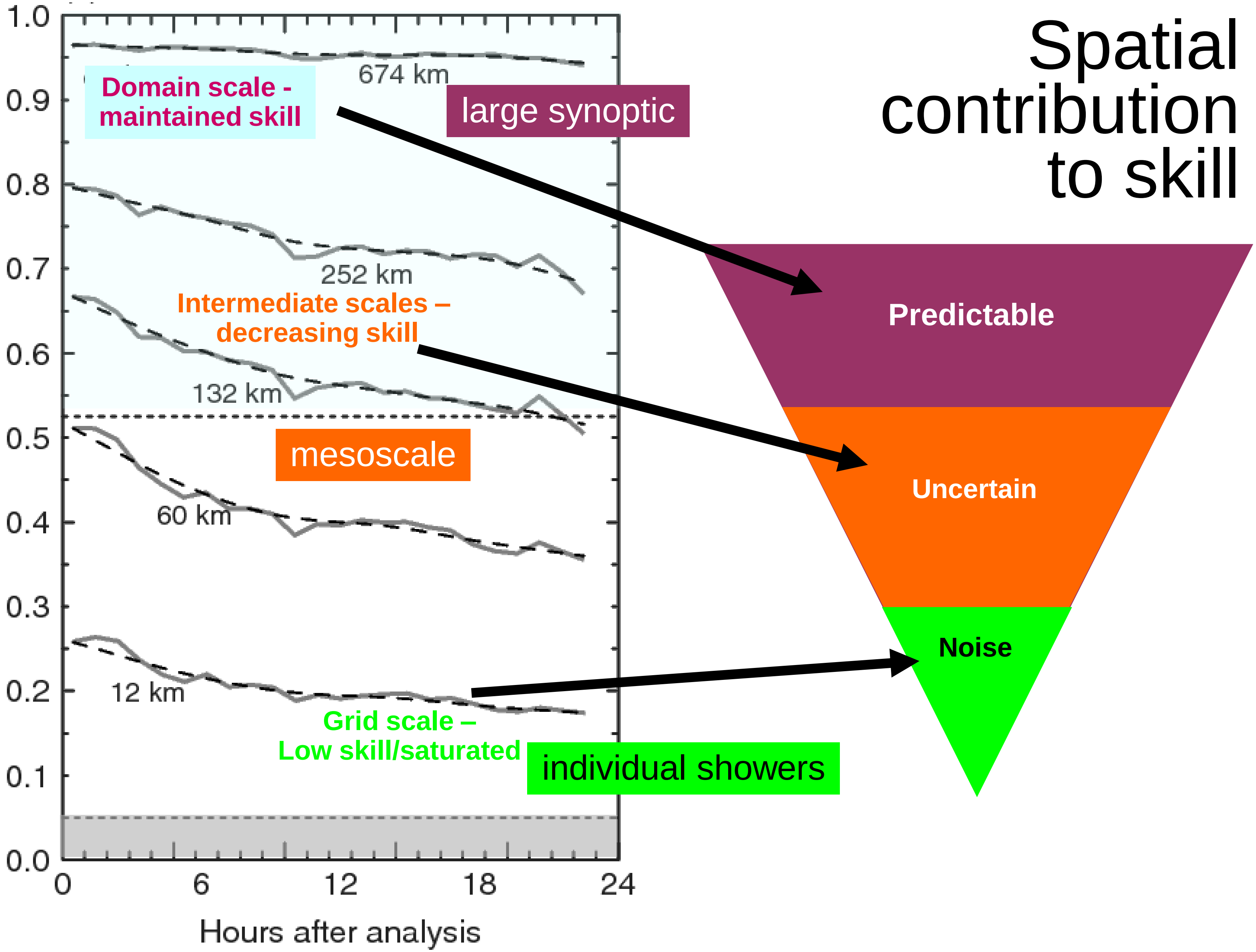
CSI = 0 for first 4;

CSI > 0 for the 5th

$$CSI = \frac{hits}{hits + false\ alarms + misses}$$

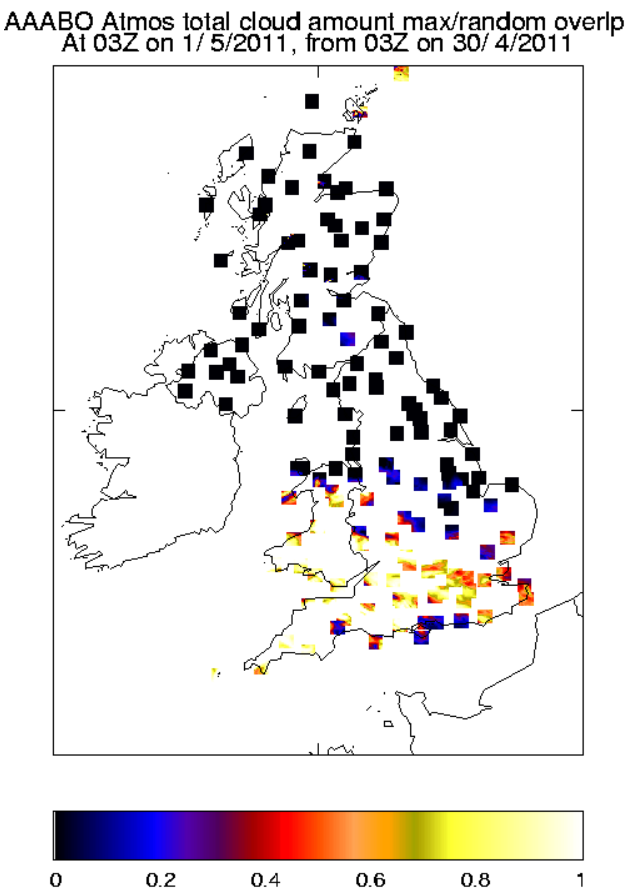
From Roberts 2008

Spatial contribution to skill

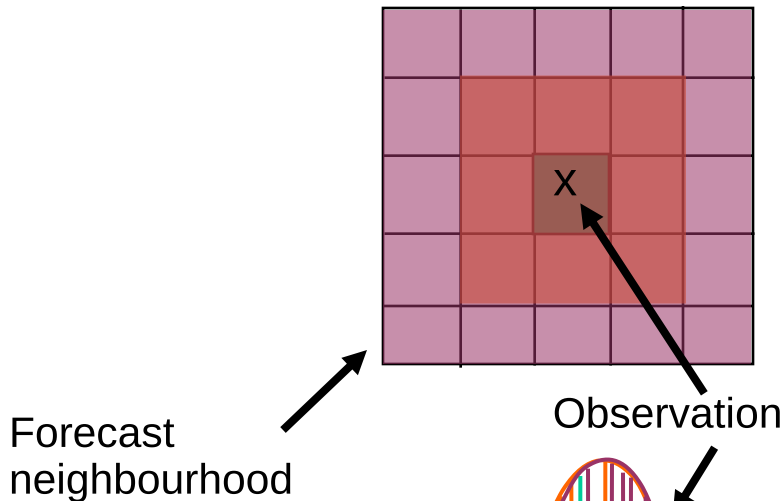


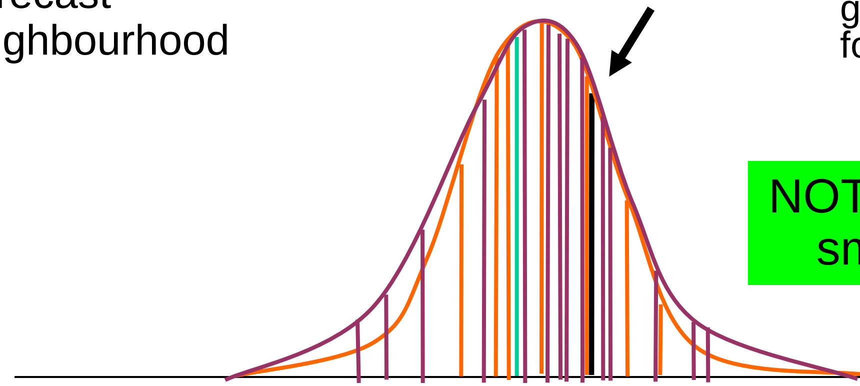
Spatial sampling

17 x 17



- Make use of spatial verification methods which compare **single observations** to a **forecast neighbourhood around the observation location.** → **SO-NF**





- Represents a **fundamental departure from our current verification system strategy** where the emphasis is on extracting the nearest GP or bilinear interpolation to get matched forecast-ob pair.

NOT upscaling/ smoothing!

Only ~130 1.5 km grid points in >500 000 domain used to assess entire forecast!

Note the variability in the neighbourhoods.

Framework outline

- Use standard synoptic observations and a range of neighbourhood sizes
- Use 24h persisted observations as reference
- The method needs to be able to compare:
  - ✔✔ Deterministic vs deterministic (different resolutions, and test vs control of the same resolution)
  - ✔ Deterministic vs EPS
  - EPS vs EPS
- Test whether differences are statistically significant (Wilcoxon) ["s" denotes significant at 5%]
- Grid scale calculated for reference → NOT main focus.

@ grid scale

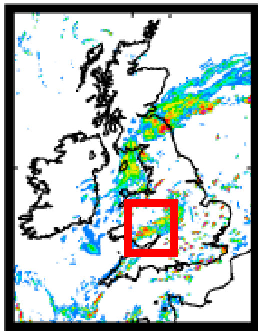
Variable	Old	New
Temp	RMSESS →	MAE
Vector wind (wind speed)	RMSVESS →	MAE
Cloud cover	ETS →	PC
CBH	ETS →	PC
Visibility	ETS →	PC
1h precip	ETS →	PC

RMS(V)ESS = Root Mean Square (Vector) Error Skill Score
ETS = Equitable Threat Score
BSS = Brier Skill Score
RPSS = Ranked Probability Skill Score
CRPSS = Continuous Ranked Probability Skill Score
MAE = Mean Absolute Error
PC = Proportion Correct

Mittermaier 2012, accepted subject to revisions WAF

Approach

2.2 km MOGREPS-UK ensemble



- Deterministic forecast with/ without neighbourhood

or

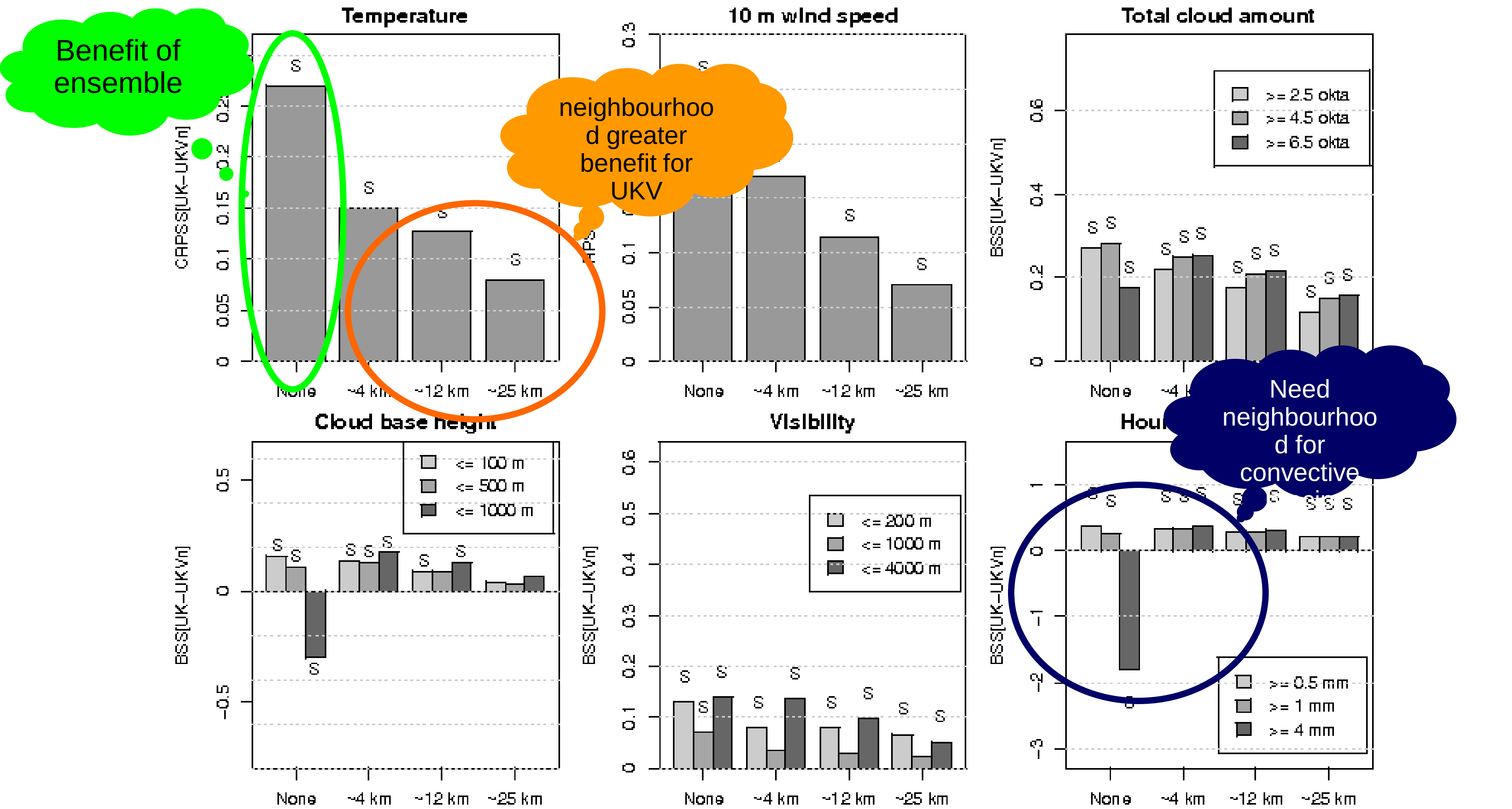
- Ensemble members with/without neighbourhoods

Comparisons:
1 GP with 12 single ensemble GPs
or
9 GP with 12 * 9 ensemble GPs → enhanced sampling

MOGREPS-UK @ 2.2 km
UKV @ 1.5 km

Deterministic vs EPS

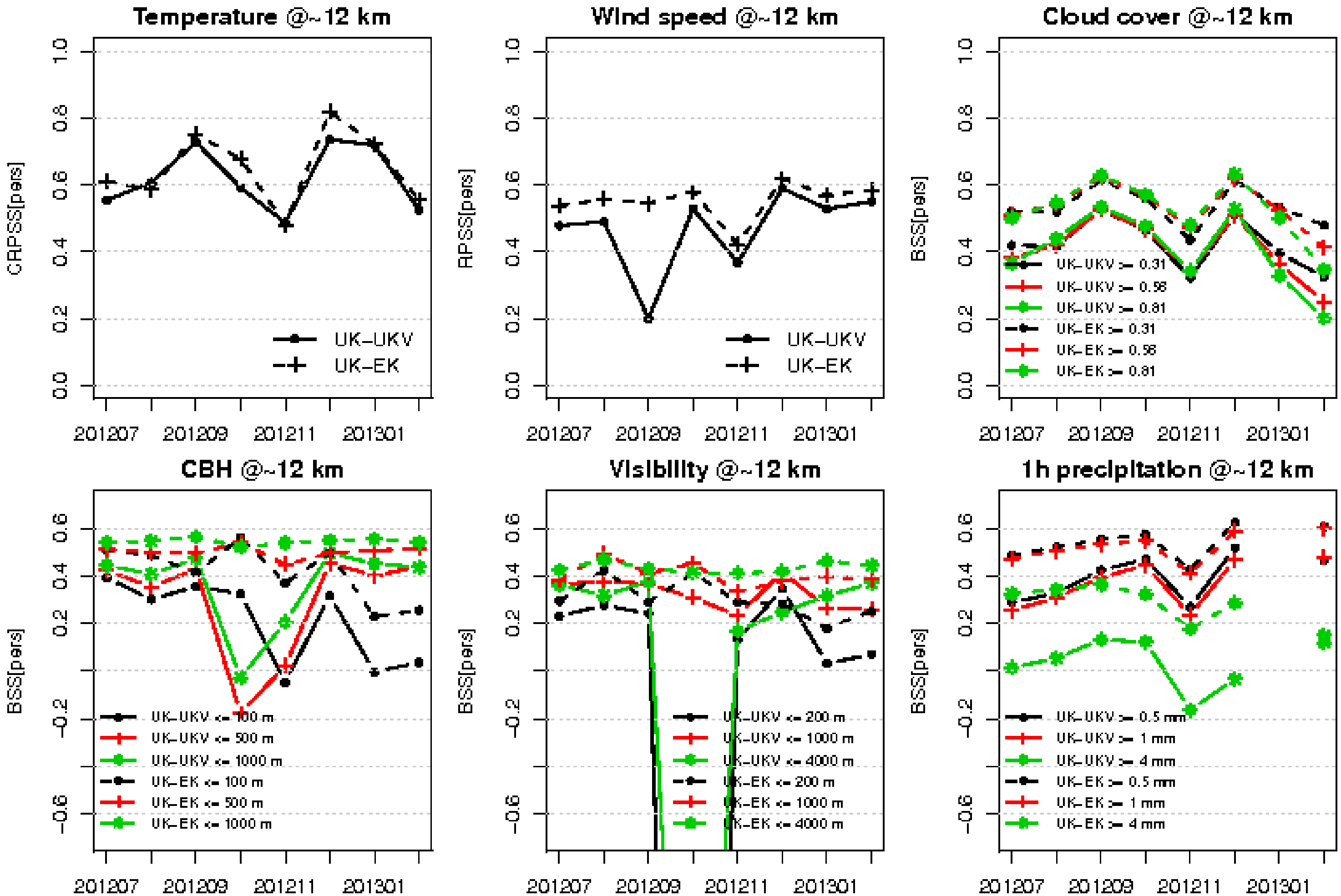
+ve = MOGREPS-UK ensemble better
“none” = 12 nearest GP values MOGREPS-UK vs 1 nearest GP UKV



1st 5 weeks of 03Z MOGREPS-UK

MOGREPS-UK @ 2.2 km
UKV @ 1.5 km

Time series – skill against persistence



Conclusions

- Method aims to provide objective reflection of inherent skill within a forecast neighbourhood in the vicinity of an observing site in a quasi-probabilistic way.
- Method can not fabricate "skill" where there is none. Model deficiencies are clearly highlighted.
- Method appears robust for all three scenarios tested → key requirement for Met Office Unified Model R & D.
- Results point the way for post-processing km-scale NWP output to maximise skill of forecast products.

Questions?