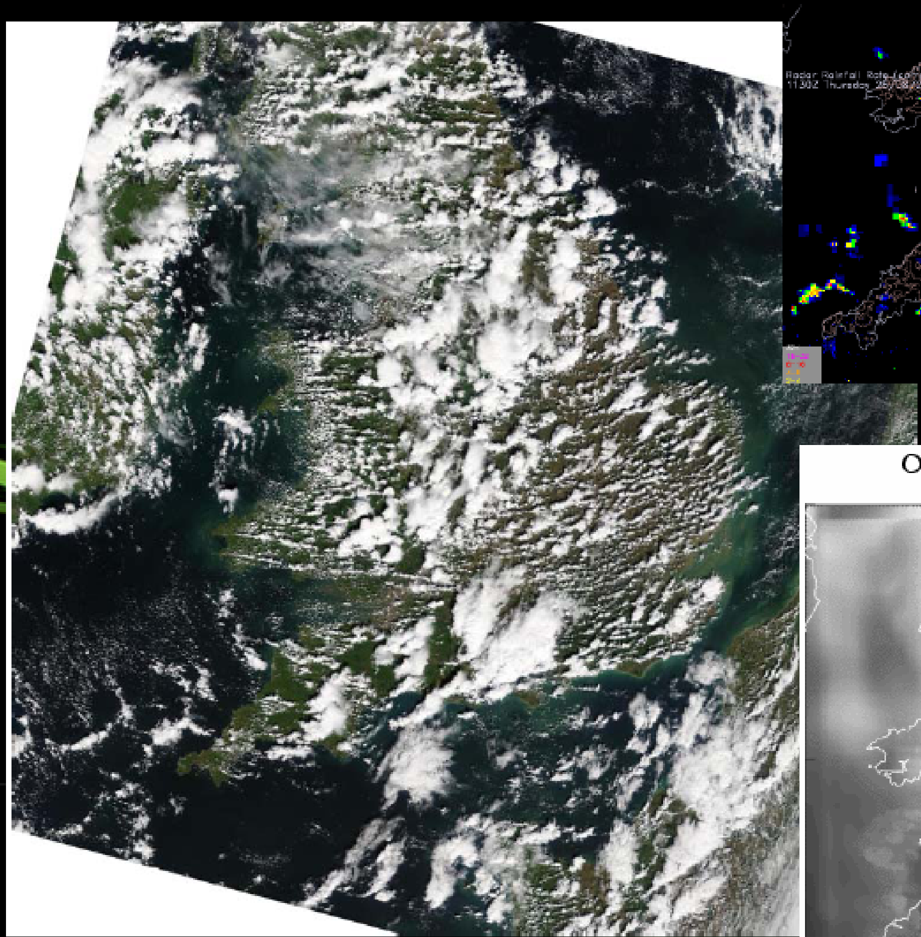
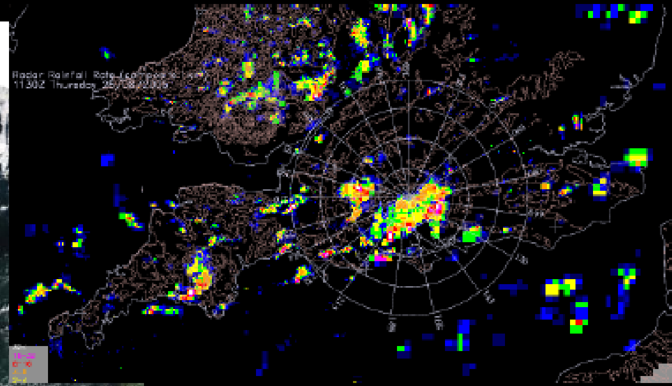
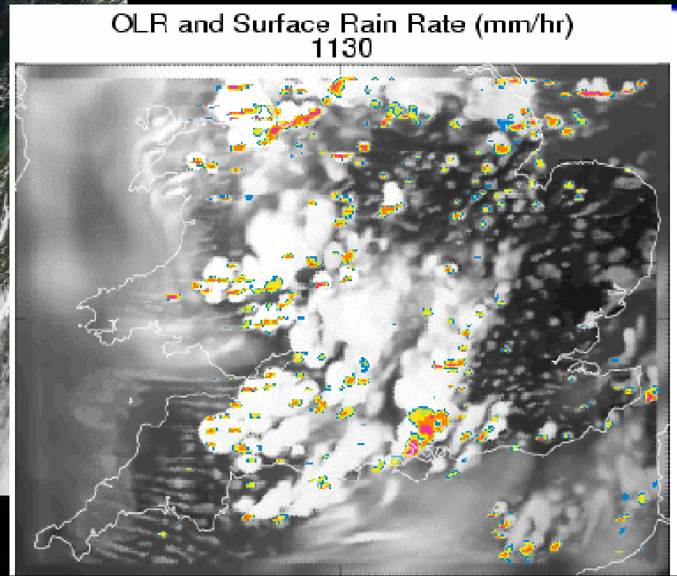






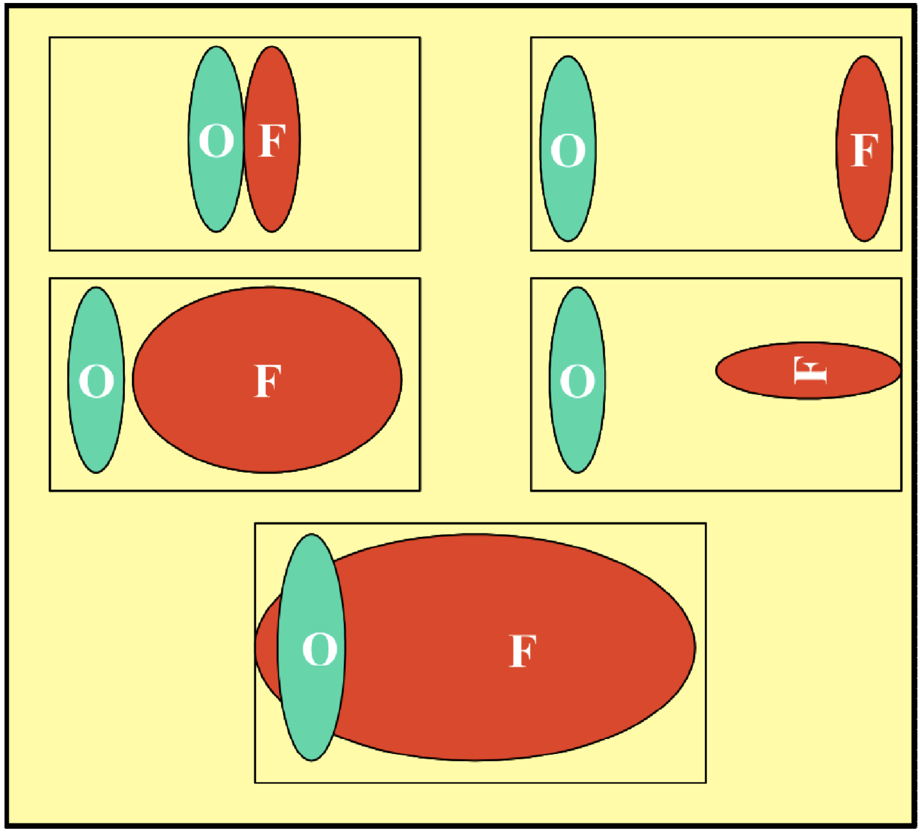
Demonstrating a strategy for verifying km-scale NWP forecasts at observing sites

Marion Mittermaier

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!

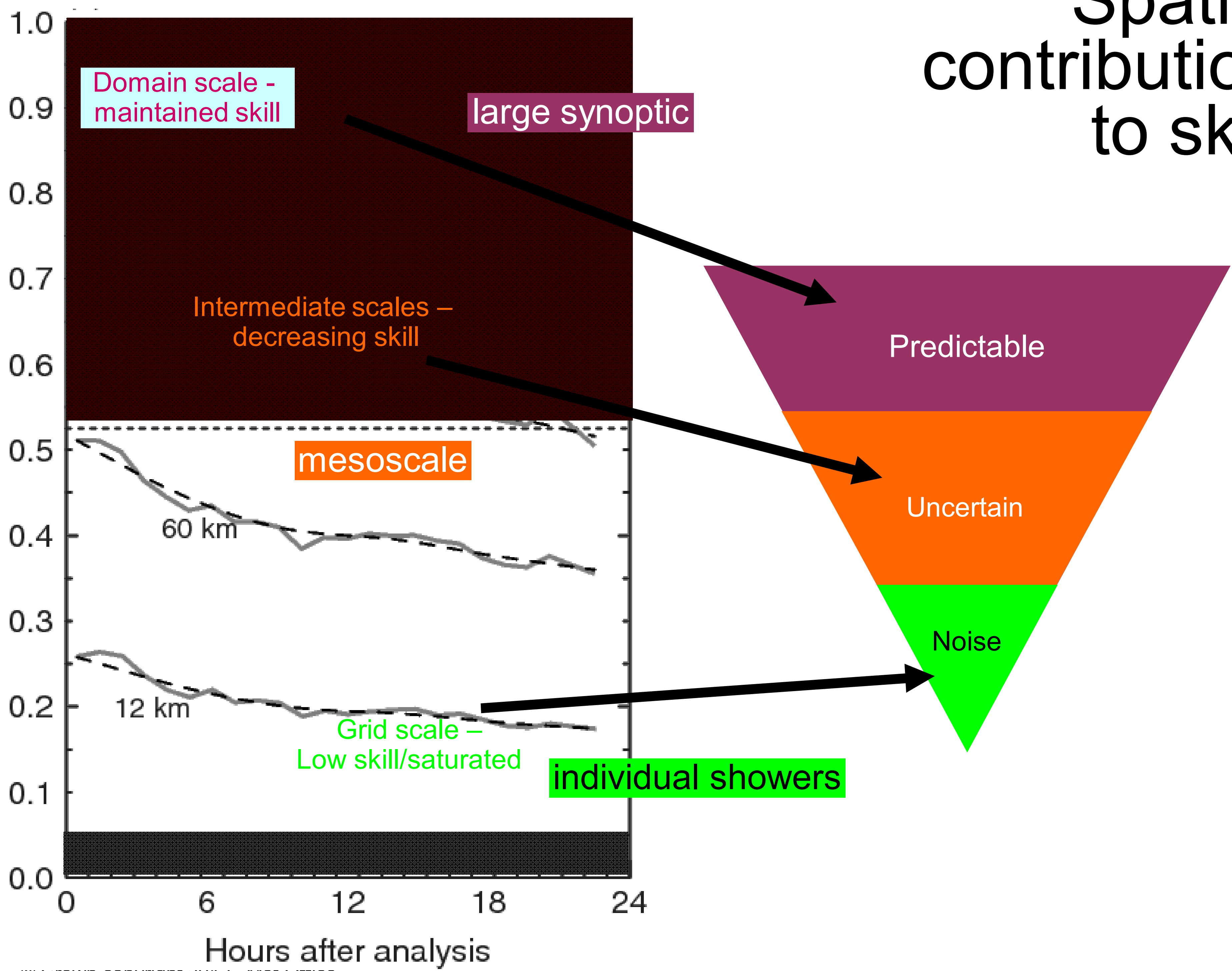
CSI = 0 for first 4;

CSI > 0 for the 5th

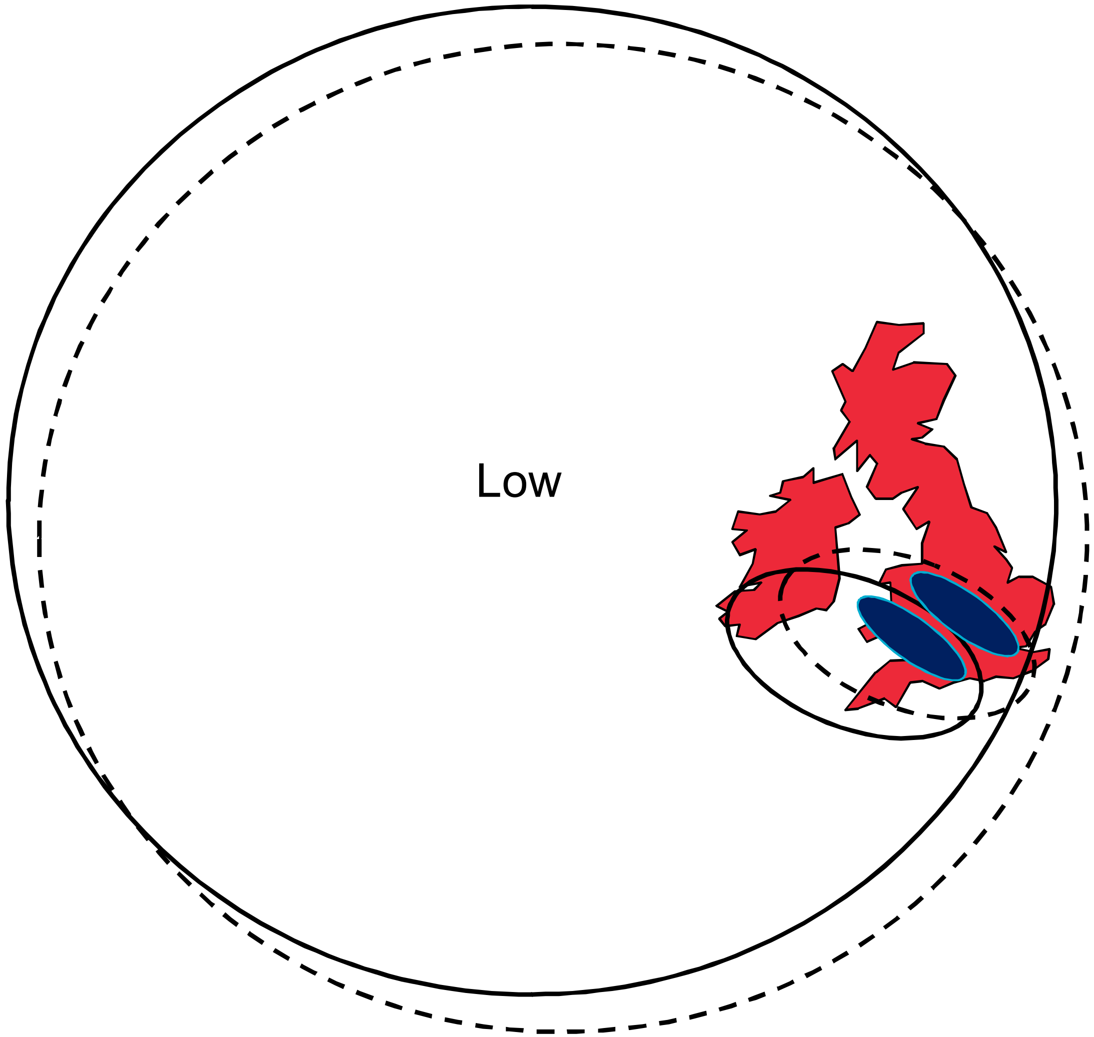
$$CSI = \frac{hits}{hits + false\ alarms + misses}$$

Spatial contribution to skill

From Roberts 2008



Small uncertainty at large scales = large uncertainty at small scales



Justifies the use of a downscaling ensemble (MOGREPS-UK)

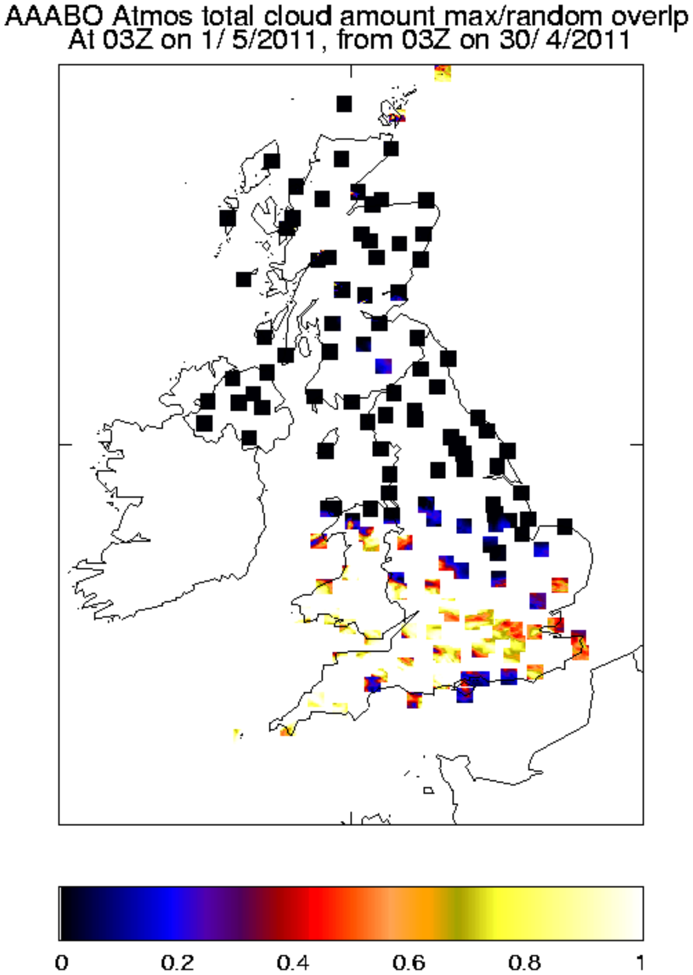
Link to larger scale:
Russell *et al*. 2008
Hanley *et al.* 2011, 2012

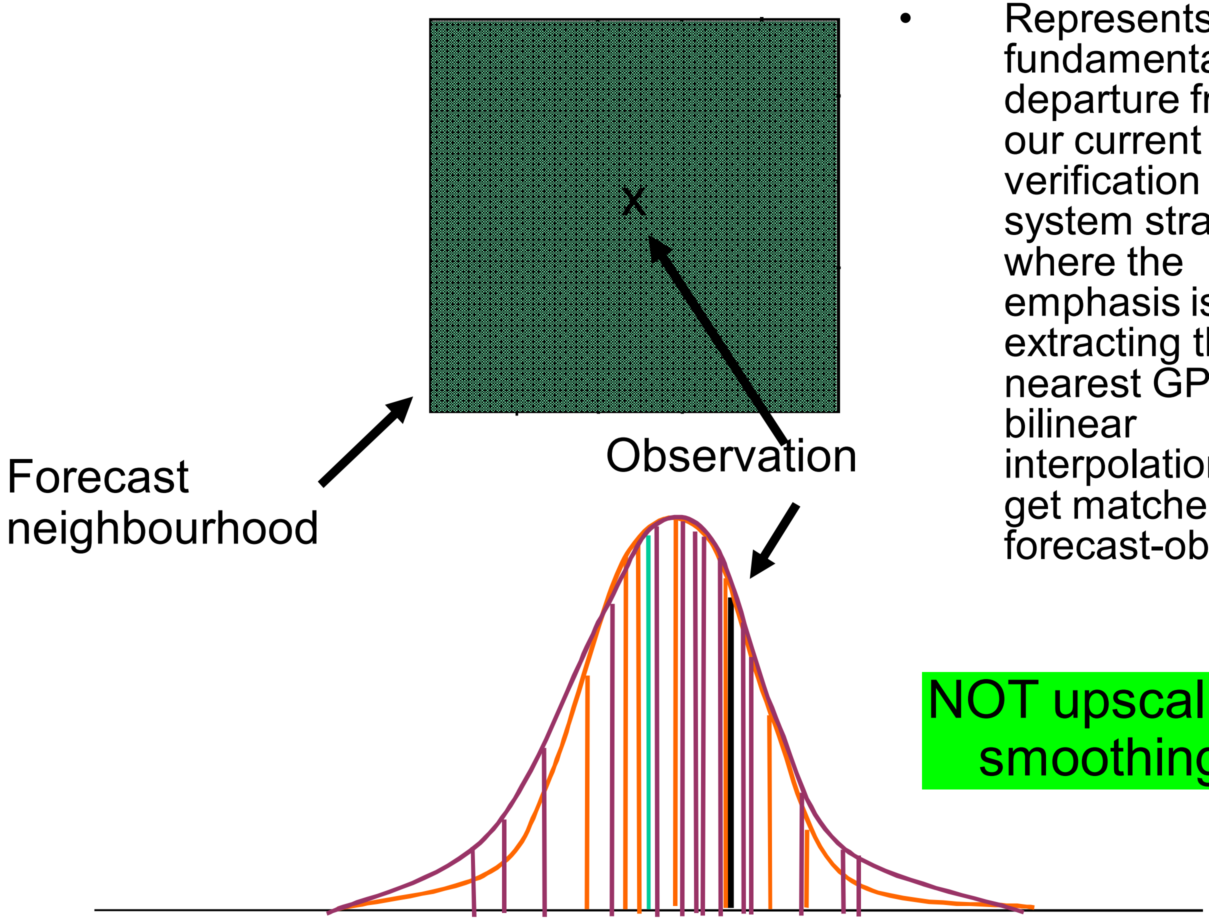
5% error at 1000 km = 100% error at 50 km

Spatial sampling

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

17 x 17





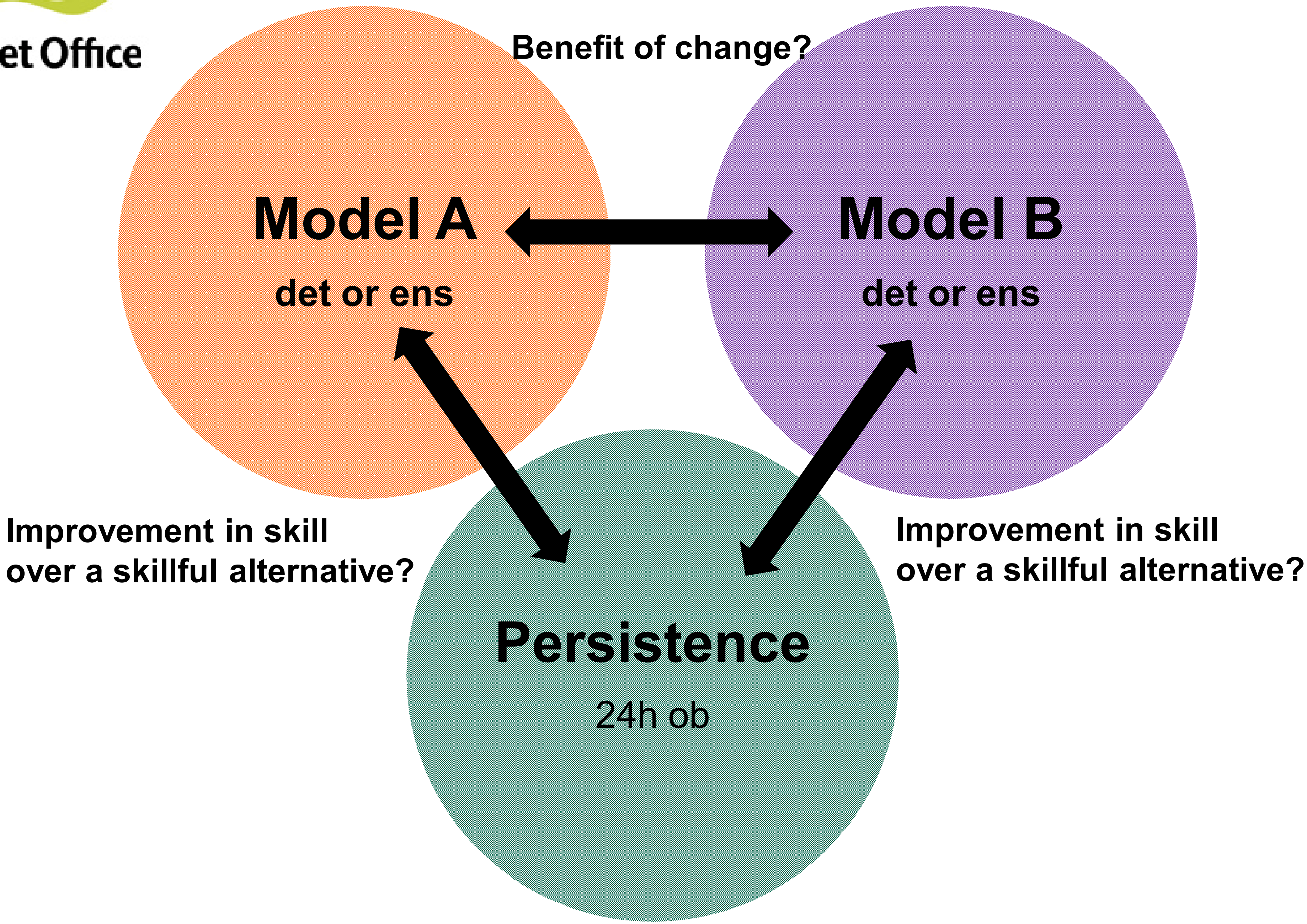
- Represents a fundamental departure from our current verification system strategy where the emphasis is on extracting the nearest GP or bilinear interpolation to get matched forecast-ob pair.

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.

Relative framework



Framework outline

@ grid scale

- 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.

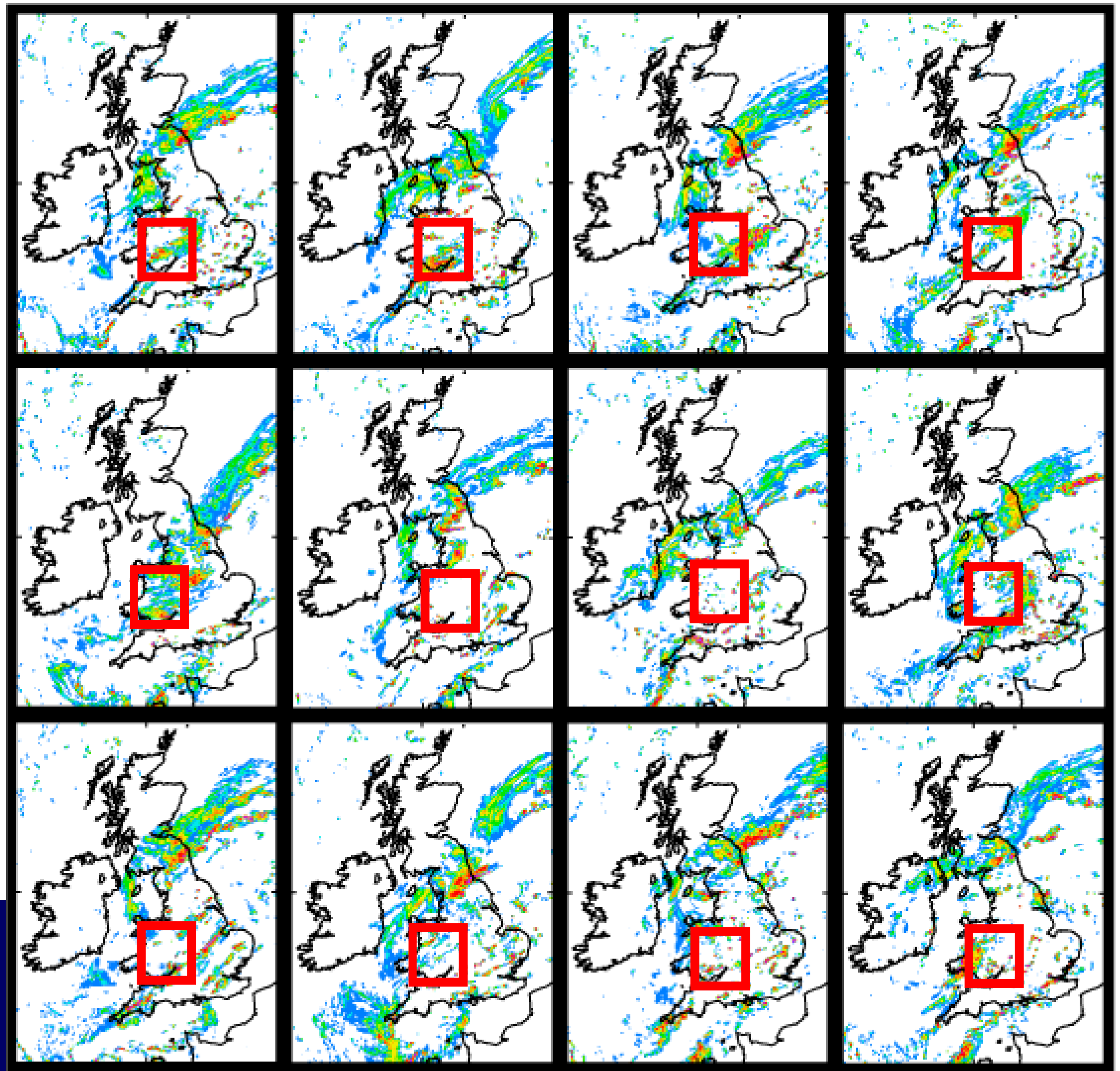
Variable	Old		New
Temp	RMSESS	→	MAE
Vector wind (wind speed)	RMSVESS	→	MAE
Cloud cover	ETS	→	PC
CBH	ETS	→	PC
Visibility	ETS	→	PC
1h precip	ETS	→	PC

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 2014, 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

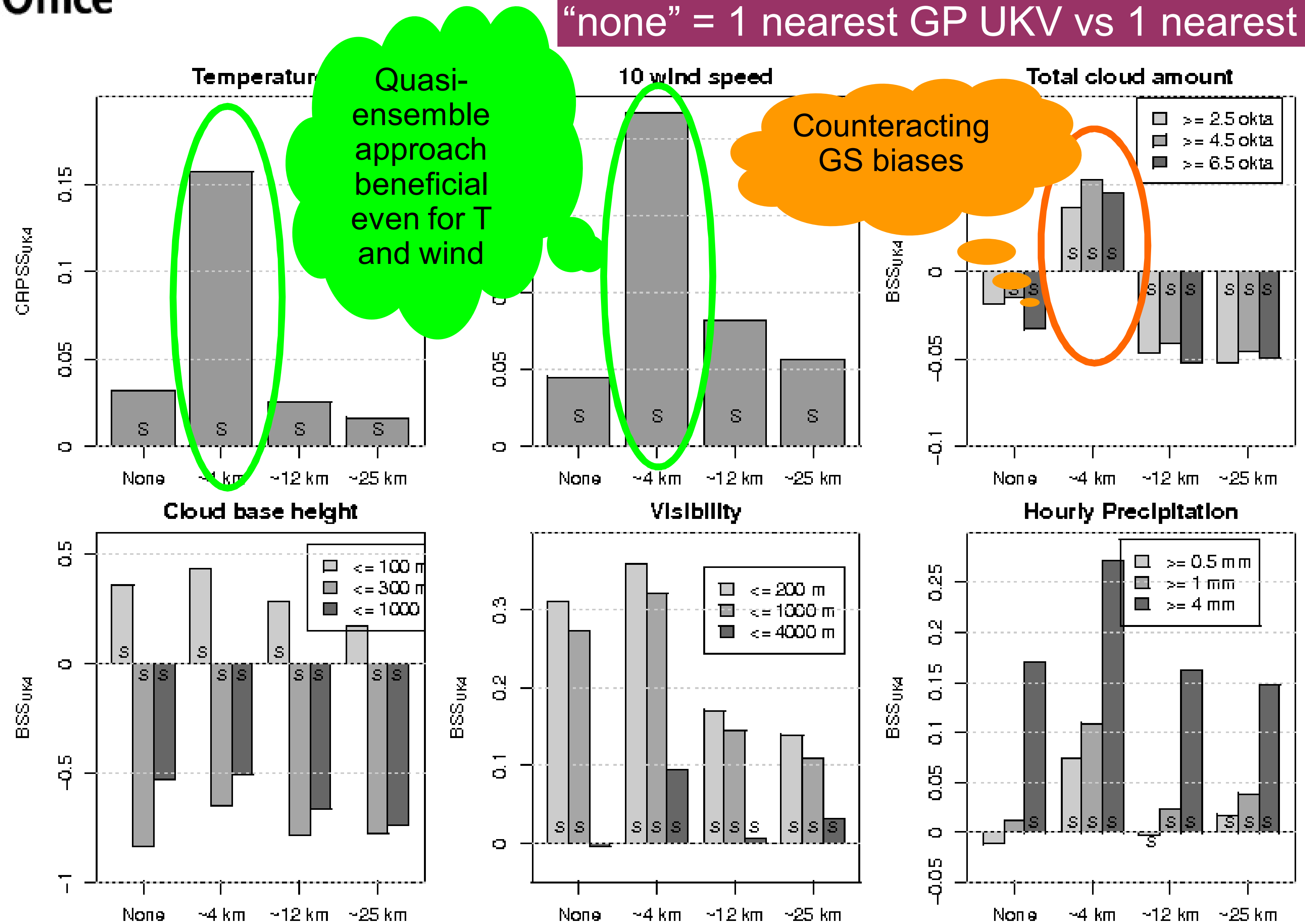
Three scenarios …

UK4 @ 4 km
UKV @ 1.5 km

Deterministic vs deterministic (different resolution)

Met Office

+ve = UKV test better than UK4
“none” = 1 nearest GP UKV vs 1 nearest GP UK4



Cumulative 36h benefit, December 2011 03Z

UKV @ 1.5 km

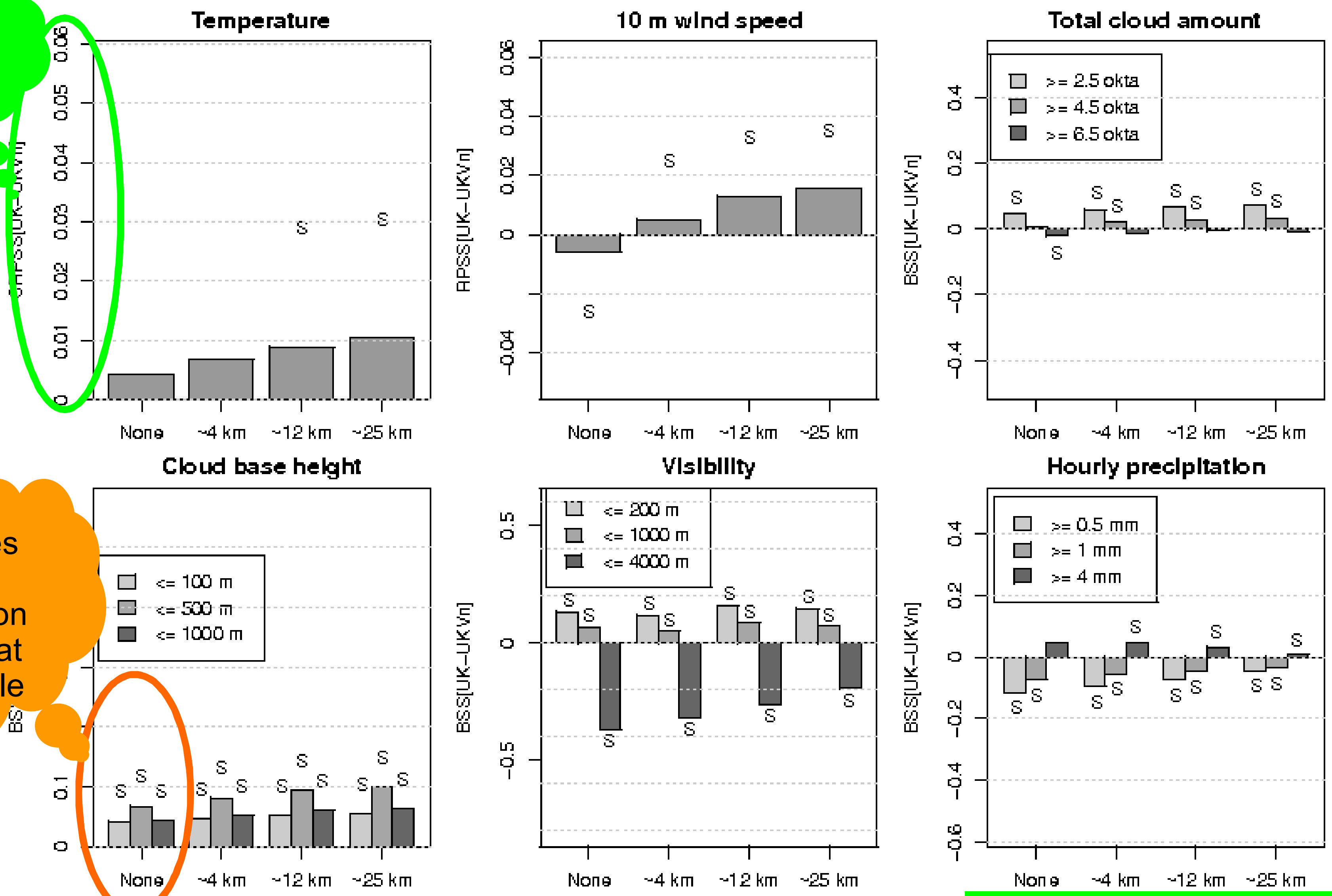
Deterministic test vs control (model trialling)

Met Office

+ve = UKV test better than control
“none” = 1 nearest GP UKV vs 1 nearest GP UKV

Detect small changes

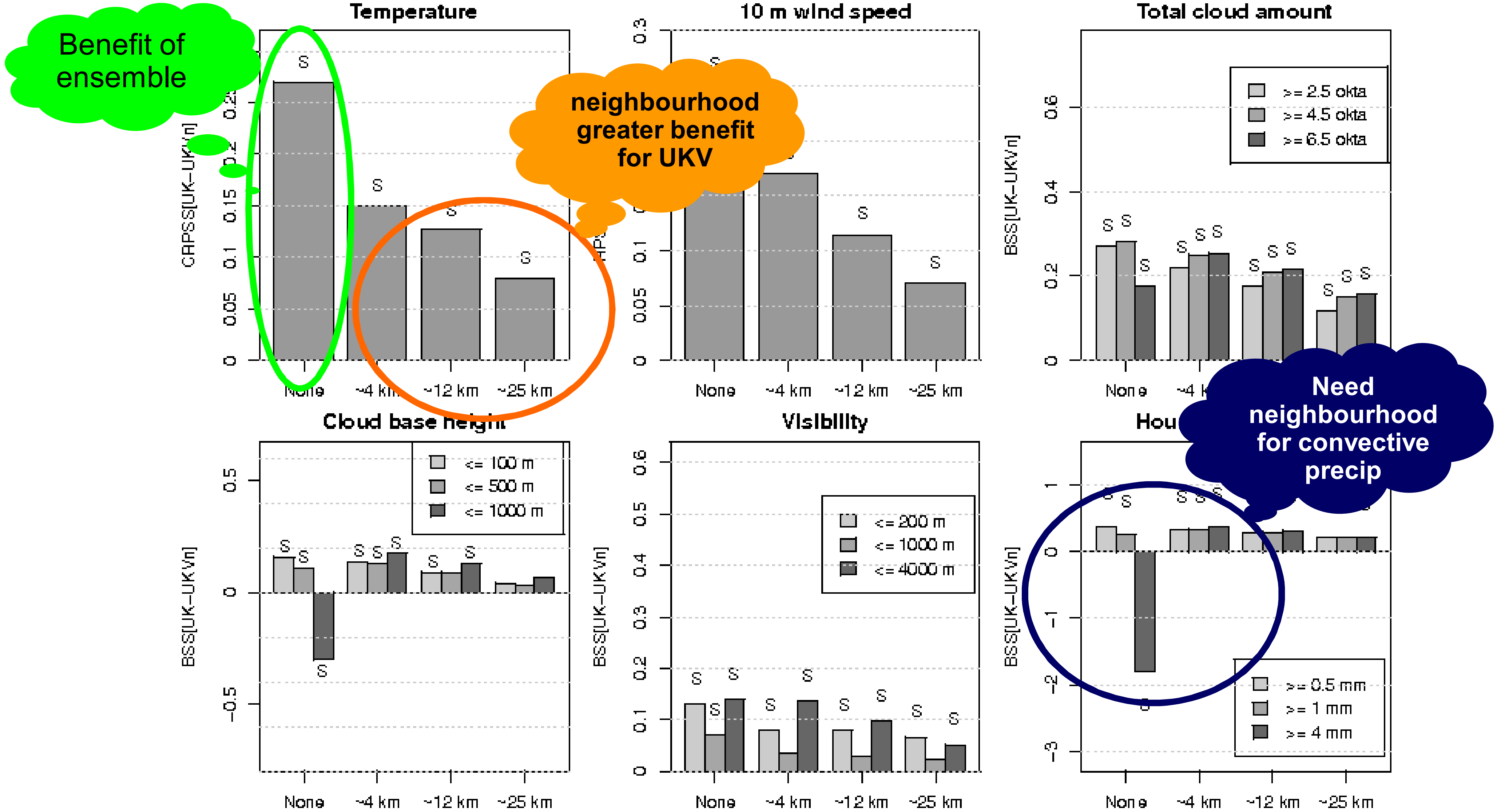
Reduces to proportion correct at grid-scale



PS31 September 2012 03Z

Deterministic vs EPS

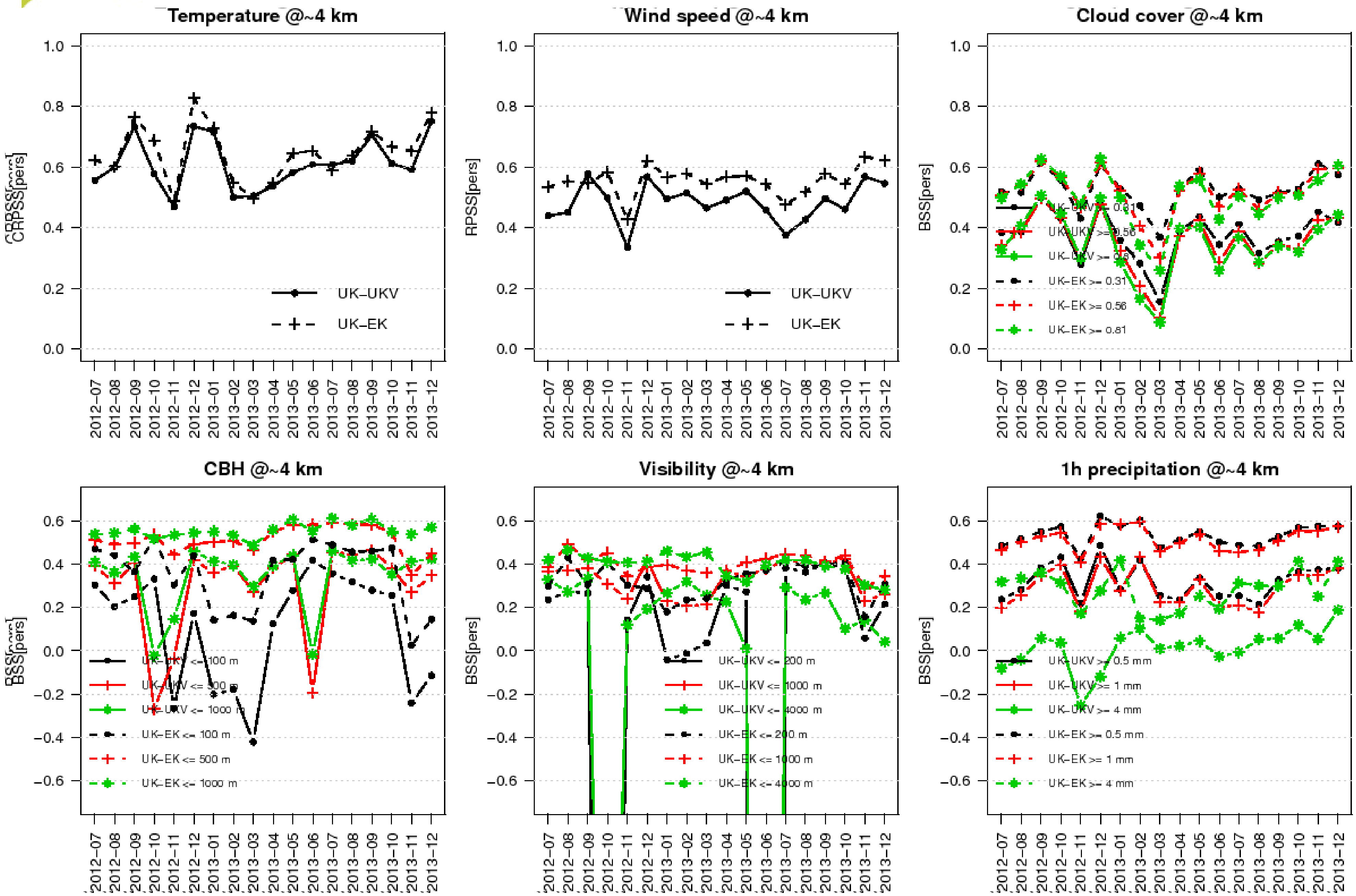
+ve = MOGREPS-UK ensemble better
"none" = 12 nearest GP values MOGREPS-UK vs 1 nearest GP UKV



1st 5 weeks of 03Z MOGREPS-UK

MOGREPS-UK @ 2.2 km
UKV @ 1.5 km

Skill against persistence

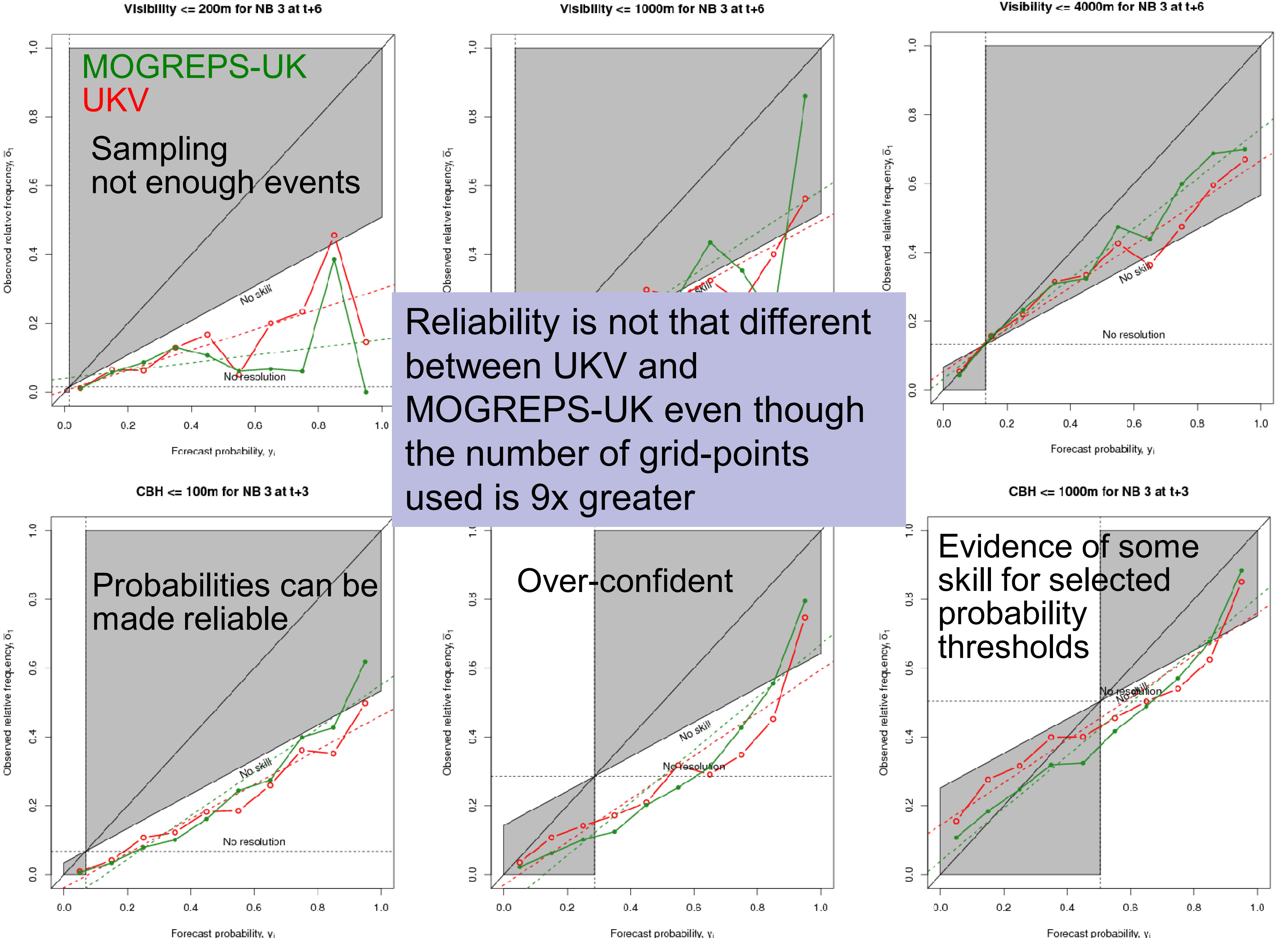


Met Office

A few words on reliability ….

Visibility and cloud base height (CBH) for 3 months JFM 2013

NB3 ~ 25 km



Reliability is not that different between UKV and MOGREPS-UK even though the number of grid-points used is 9x greater

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.

Conclusions (cont.)

Met Office

- New verification framework illustrates **benefit of km-scale ensemble** over deterministic.
- **Bigger neighbourhoods will improve forecast skill** (for the most part) but the UKV needs (and benefits more from) neighbourhood processing, i.e. better "harvesting" of information content.

Questions?

Mittermaier MP, 2014: A strategy for verifying near-convection-resolving forecasts at observing sites. *Wea. Forecasting*. **29(2),** 185-204.