Identifying skilful spatial scales using the Fractions Skill Score

Marion Mittermaier and Nigel Roberts

Outline

1. Brief overview of the method
2. Synthetic cases
   - Geometric cases
   - Perturbed cases
3. Domain size influences
4. Conclusions

The method

Verification approach

We want to know:

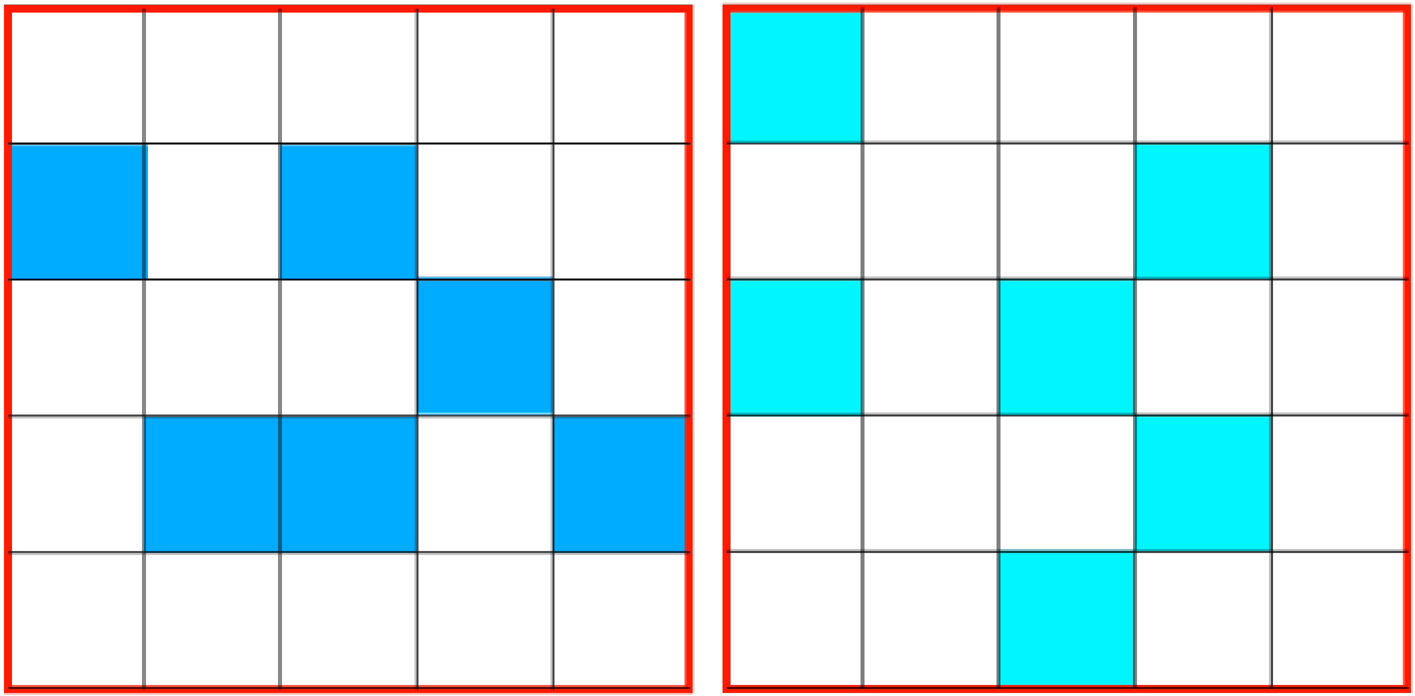
3. How forecast skill varies with neighbourhood size.
4. The smallest neighbourhood size that can be used to give sufficiently accurate forecasts.
5. Does higher resolution provide more accurate forecasts on scales of interest (e.g. river catchments)

Compare forecast fractions with fractions from radar over different sized neighbourhoods (squares for convenience)

Use rainfall accumulations to apply temporal smoothing

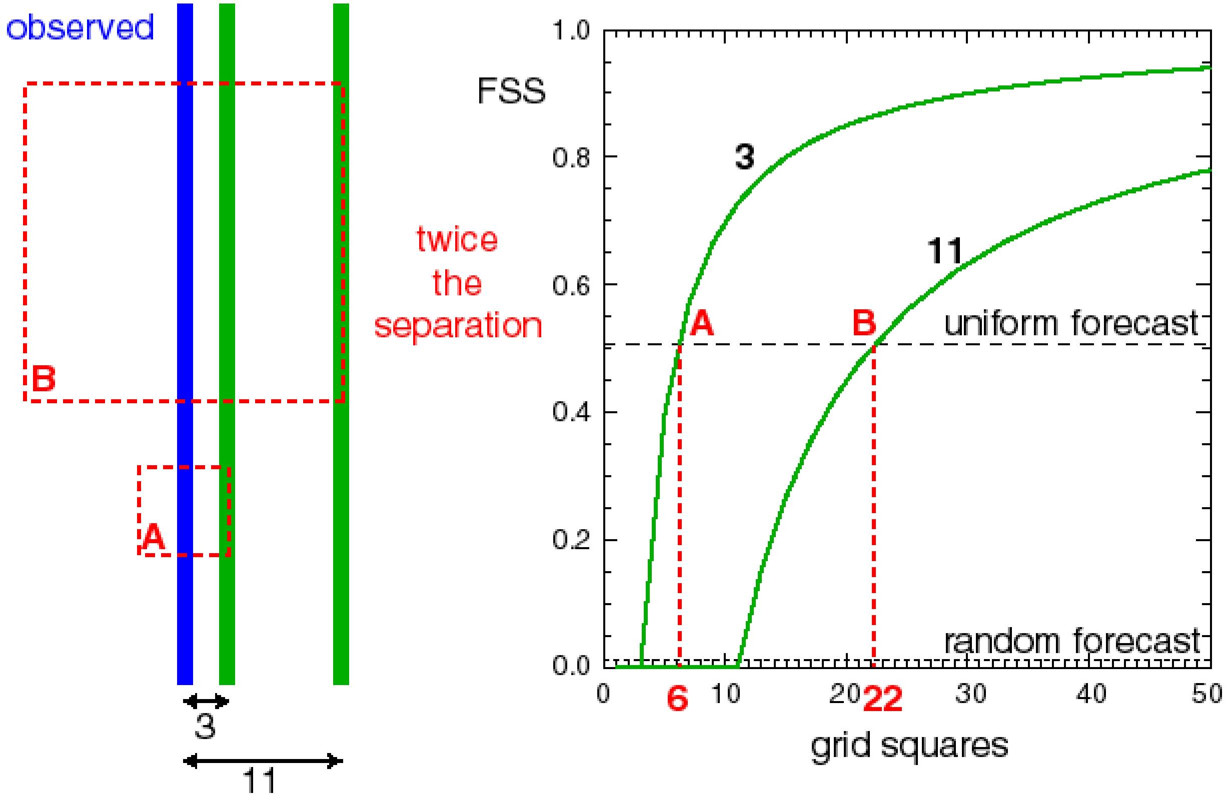
Schematic comparison of fractions

observed   forecast



Fraction = 6/25 = 0.24   Fraction = 6/25 = 0.24

Idealised example



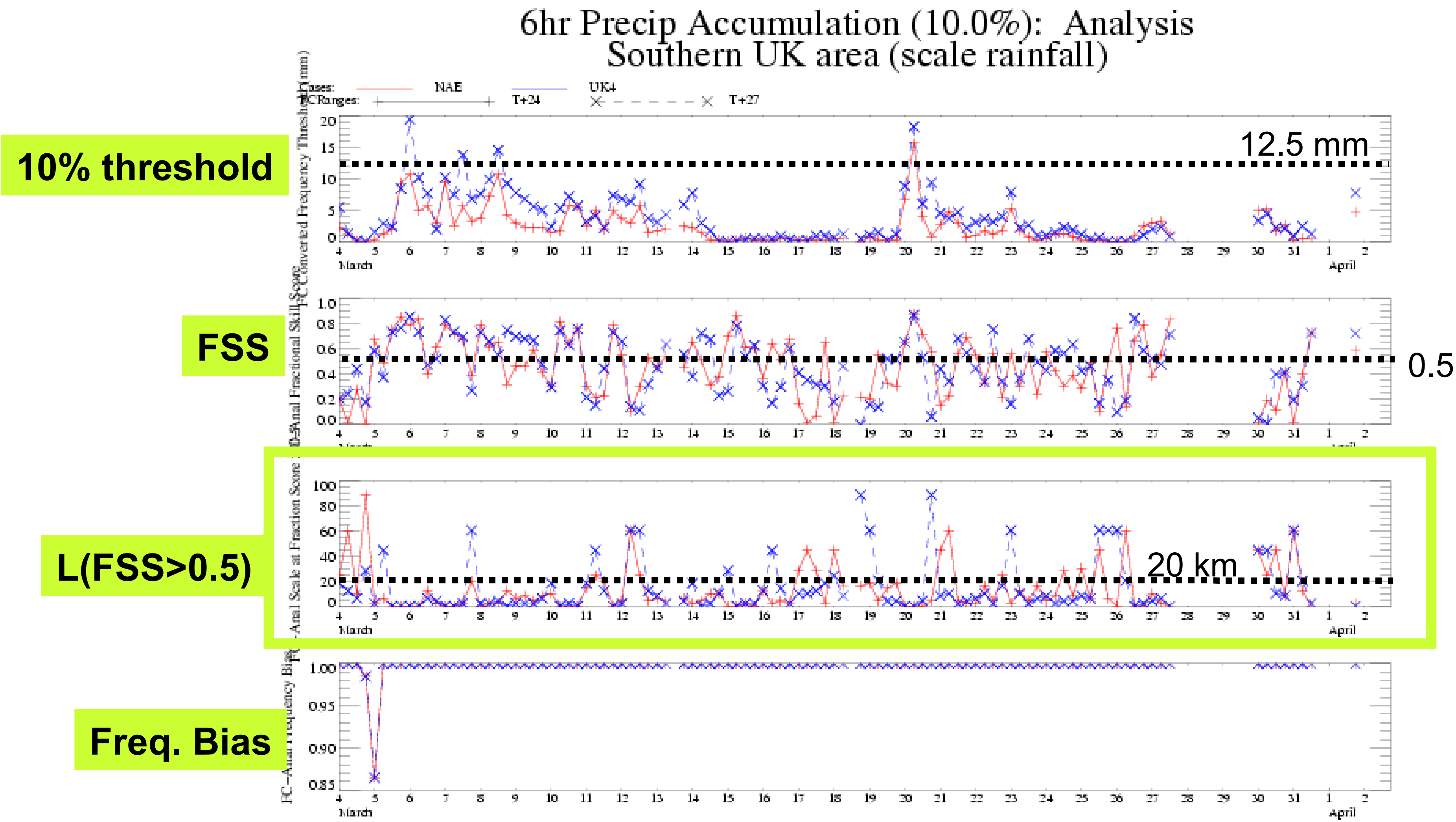
In summary

This verification method provides a way of answering **some** important questions about forecasts from ‘storm-resolving’ NWP models.

- How does forecast skill vary with spatial scale?
- At what scales are higher resolution forecasts more skilful (if any)?
- At what scales are forecasts sufficiently accurate?...........

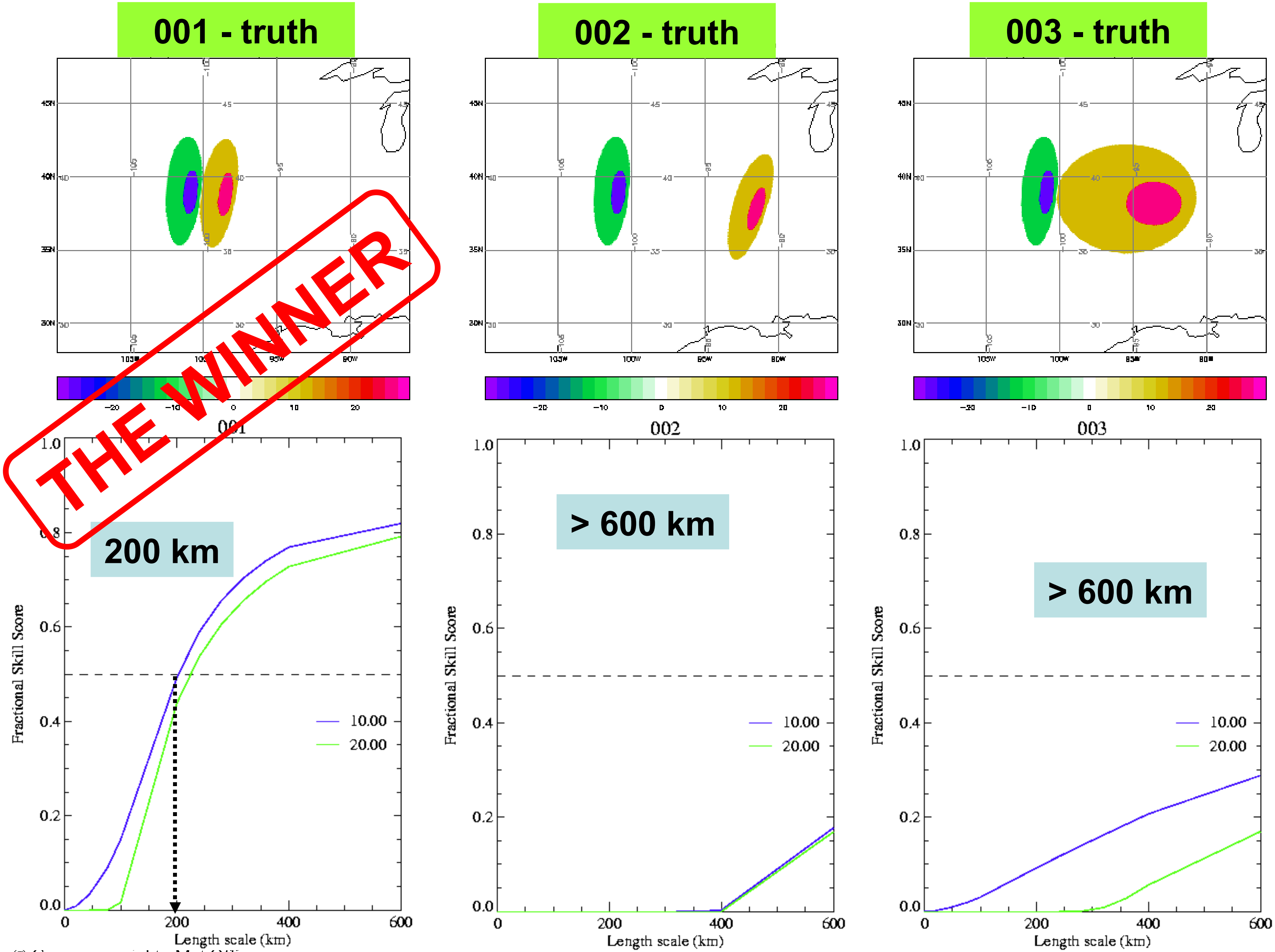
(There are other questions that need different approaches)

How we are using it

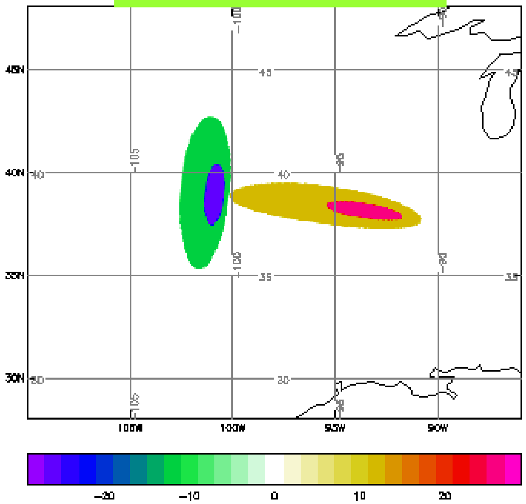


See poster by Mittermaier and Thompson

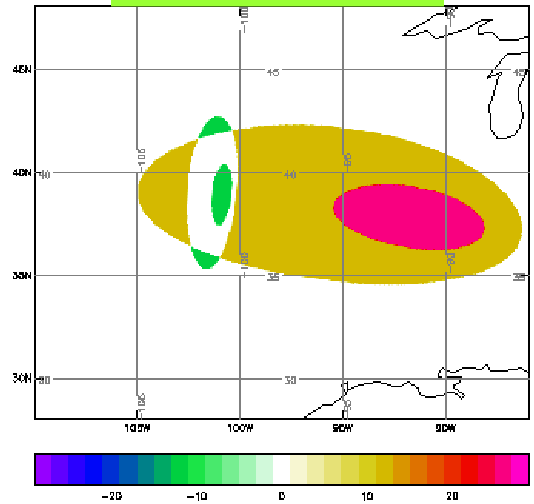
Synthetic cases

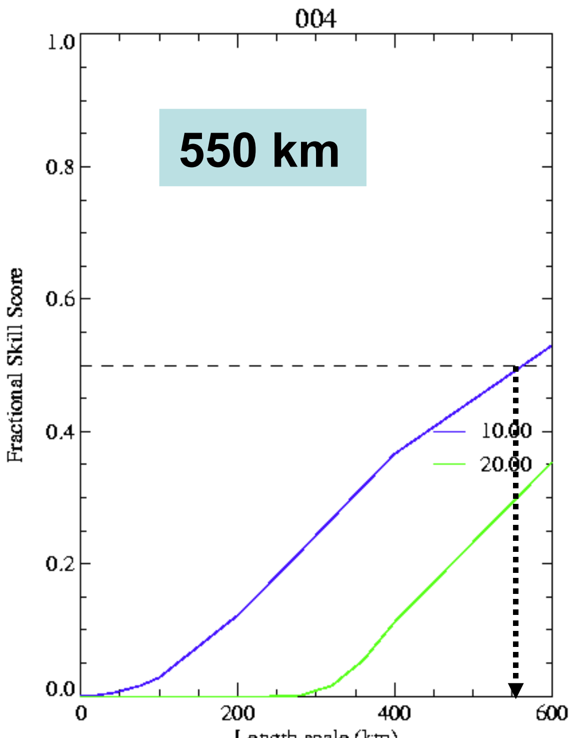

001 - truth
002 - truth
003 - truth
THE WINNER
001
002
003
200 km
> 600 km
> 600 km
Fractional Skill Score
Length scale (km)
10.00
20.00
0.0
0.2
0.4
0.6
0.8
1.0
0
200
400
600

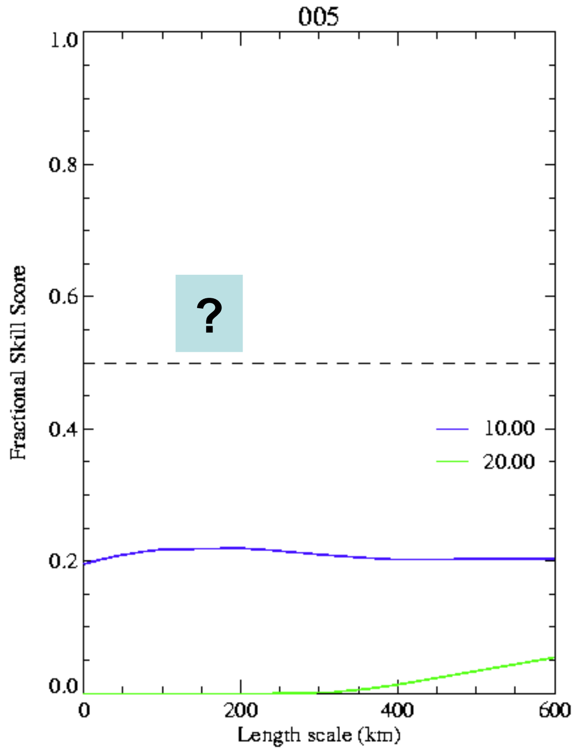
004 - truth



005 - truth







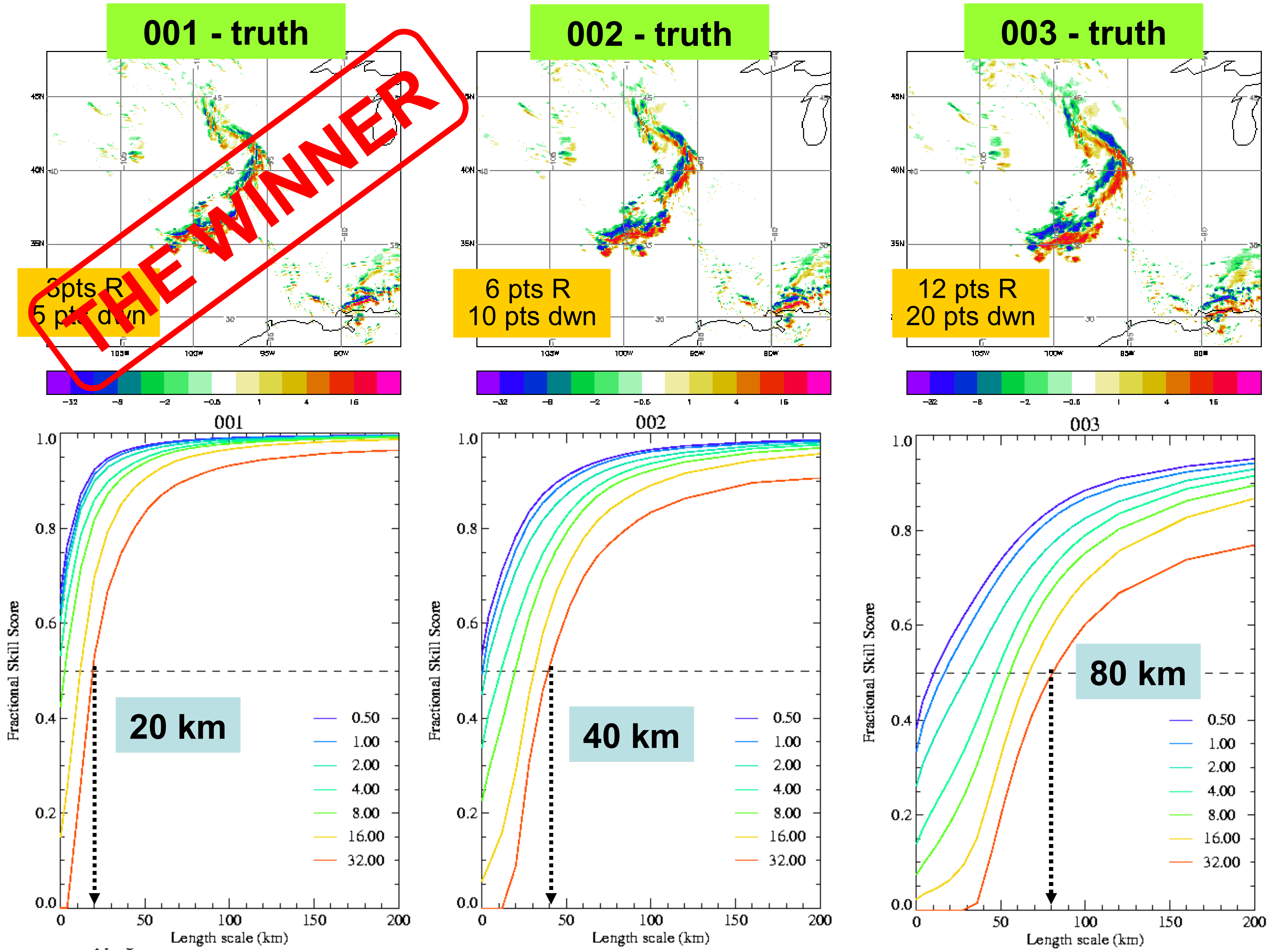
Summary : L(FSS > 0.5)

Case	> 10 mm
001	200 km
002	> 600 km
003	> 600 km
004	550 km
005	?

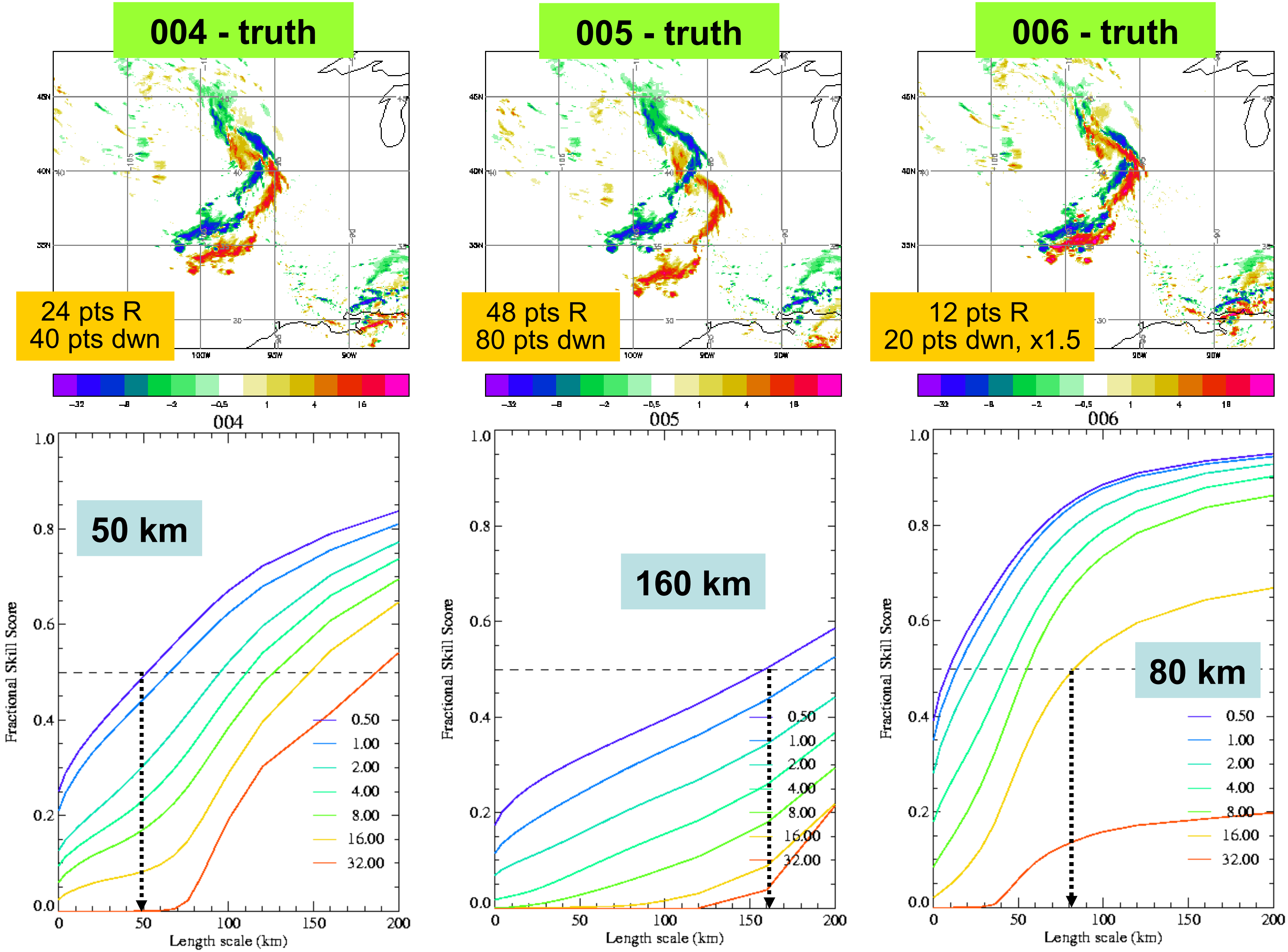
Conclusions: geometric cases

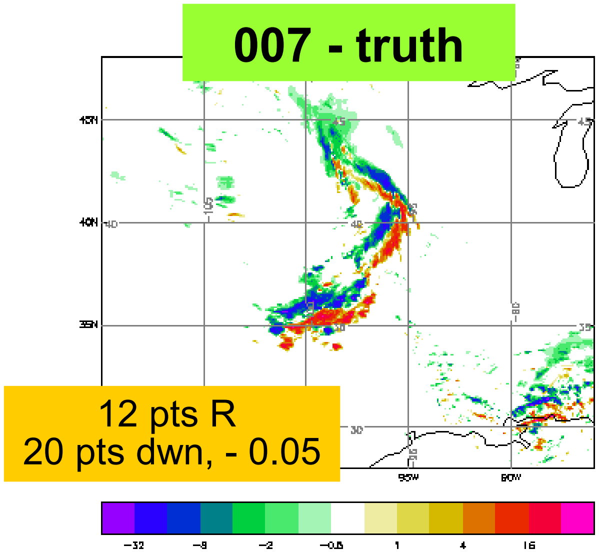
- Case 005 has the largest FSS at the grid scale but in terms of L(FSS > 0.5) never reaches a skillful scale.
- Case 001 is the most skillful in terms of scale but has no skill at all at the grid scale, based on the FSS.

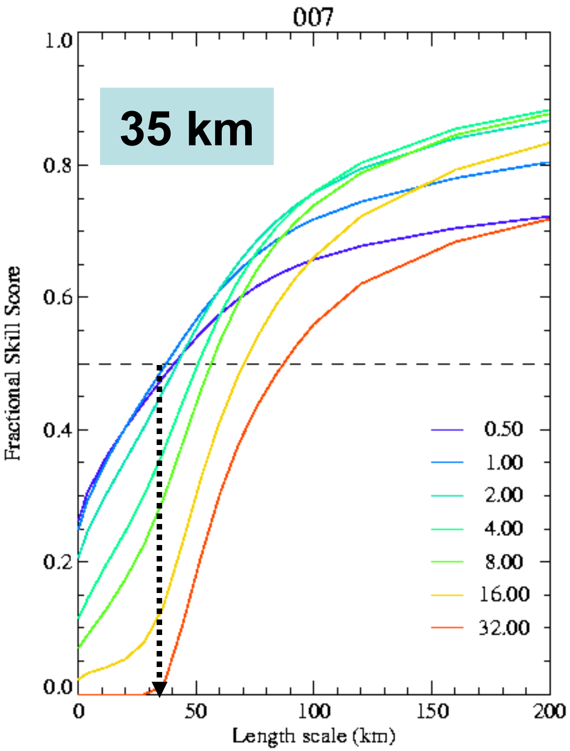
Perturbed forecasts



001 - truth
002 - truth
003 - truth
THE WINNER
3pts R
5 pts dwn
6 pts R
10 pts dwn
12 pts R
20 pts dwn
001
002
003
Fractional Skill Score
Length scale (km)
20 km
40 km
80 km
0.50
1.00
2.00
4.00
8.00
16.00
32.00


004 - truth
24 pts R
40 pts dwn
005 - truth
48 pts R
80 pts dwn
006 - truth
12 pts R
20 pts dwn, x1.5
004
50 km
005
160 km
006
80 km
Fractional Skill Score
Length scale (km)
0.50
1.00
2.00
4.00
8.00
16.00
32.00





Summary : L(FSS > 0.5)

Case	> 0.5 mm	> 32 mm
001	4 km	20 km
002	4 km	40 km
003	10 km	80 km
004	50 km	190 km
005	160 km	> 200 km
006	10 km	?
007	35 km	90 km

Conclusions: perturbed cases

- Small shifts/timing errors are initially unharmful to the FSS and scale. Gross timing errors result in large degradations in skillful scales and the overall magnitude of the FSS.

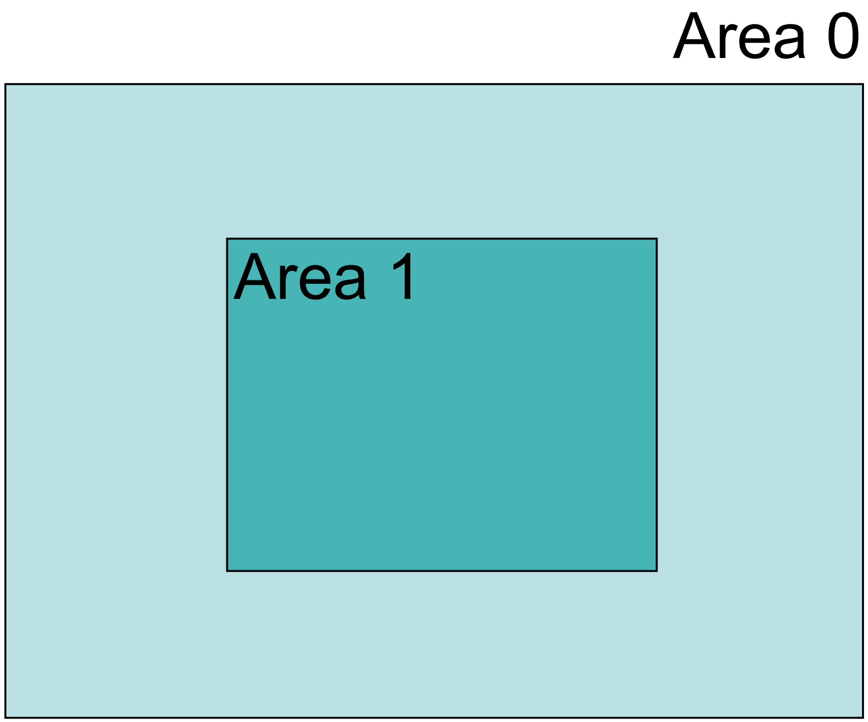
- Changes in the actual magnitudes (006 and 007) show greater impact. Subtle differences (due to near-threshold level misses) seem to be potentially more serious in terms of FSS magnitude.

Domain size influences

t+24h 1-hr forecast accumulation

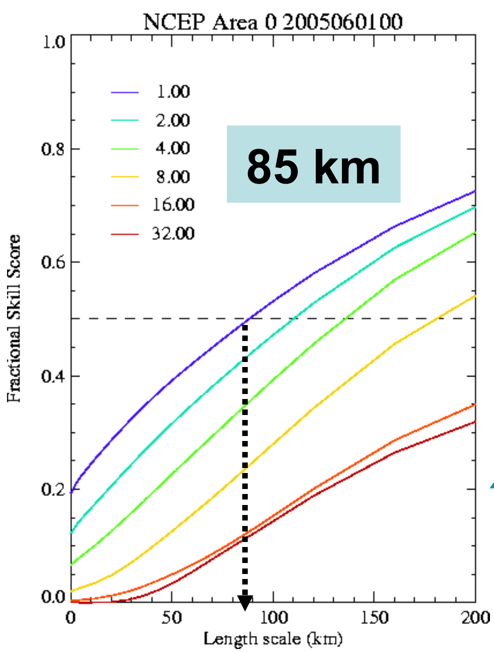
Impact of domain size

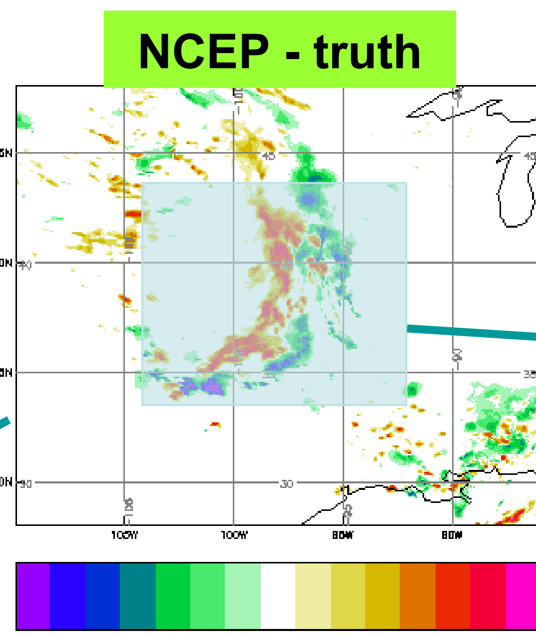
- With such a large domain the **wet-area ratio is quite small**, even for extensive precipitation areas.
- *<u>How sensitive are the results to domain size?</u>*
- Perform tests where around ~30% of the domain is used

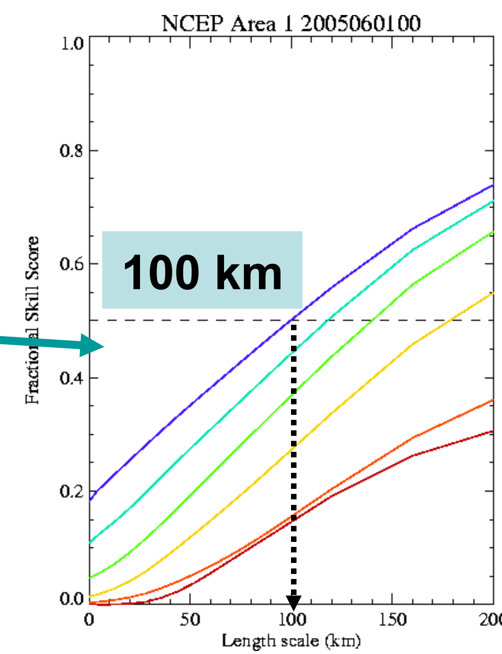


Impact of domain size 2

- Domain size affects the magnitude of the FSS and the spatial scales.
- Smaller domains are faster to compute.
- Sub-regions take into account that neither precipitation nor skill is uniform over a large domain.







General conclusions

- The absolute value of FSS is less potentially useful than the scale where an acceptable level of skill is reached.
- The skillful scales in the fake cases ties in well with the 1-d idealized example results.
- Frequency thresholds are potentially useful but not when the domains are very large. [The zeros dominate the cdf.]
- Consider a “roving” (fixed size) verification domain that focuses on an area of interest.
- Other optimizing ideas are being explored.

Questions?

Mittermaier M.P. and N. Roberts, 2009: Intercomparison of Spatial Forecast Verification Methods: Identifying skilful spatial scales using the Fractions Skill Score, accepted by *Wea. Forecasting* subject to revisions, 24 March 2009.