Verification of operational precipitation forecasts over the UK

Clive Wilson

30th EWGLAM/15th SRNWP meetings – Madrid 6-9 Oct 2008

Contents

- Aims
- Standard operational verification
- Radar composites and gridded gauges
- Fuzzy scores
- Conclusions

Aims

- Operational verification of 40km,12km and 4km models uses synoptic stations and nearest gridpoint
- 12km and 4km also routinely verified against radar analyses using common 12km grid
- Can we find common signals and explain differences ?
- Compare radar and climatological gauge analyses to estimate uncertainty
- (Use idealised tests to understand
  - Domain size; Fractional pattern; sub-sampling)
- Is the higher resolution model better?

Primary Operational Forecast Systems

UK 4km

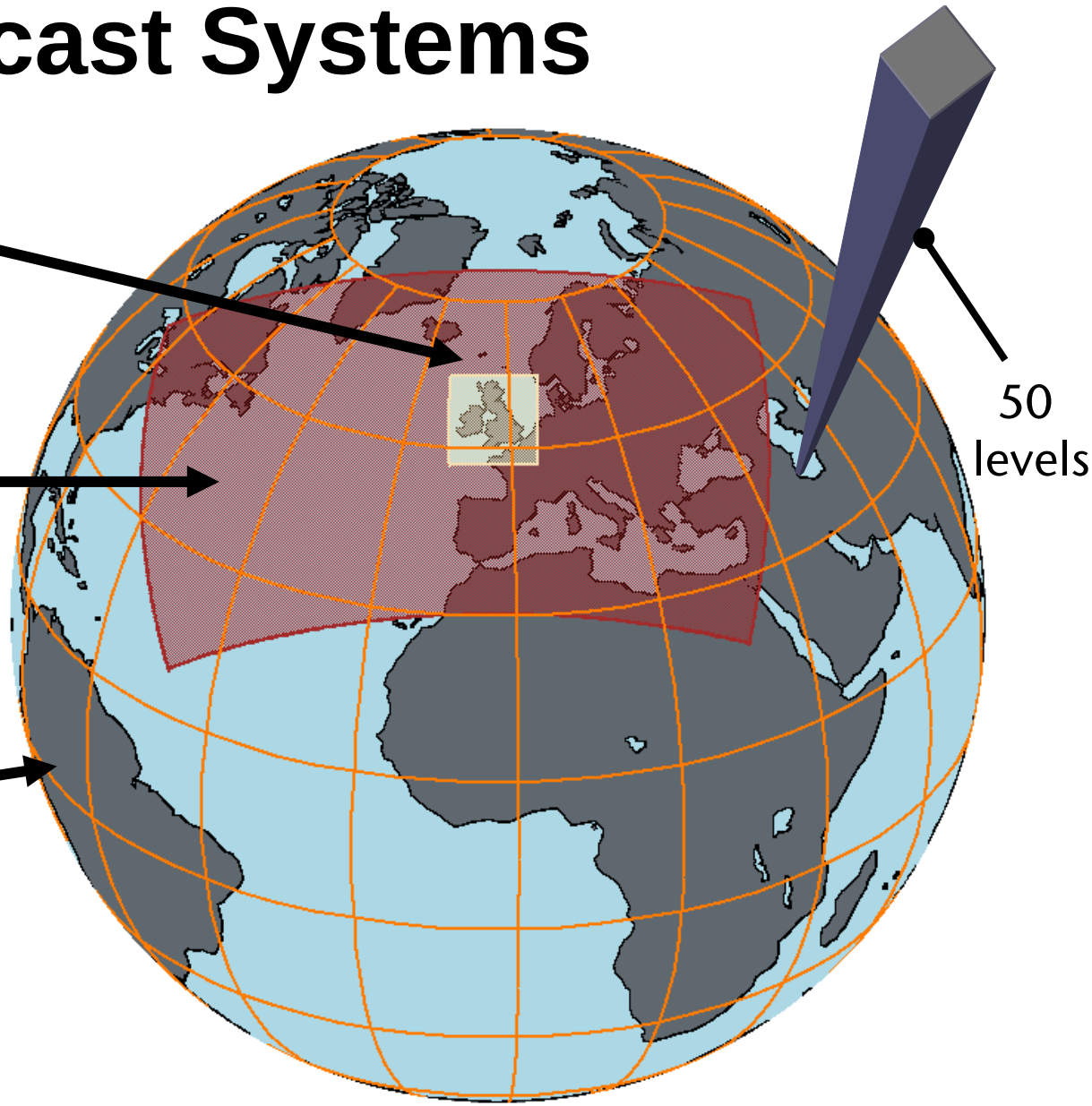
- 36 hour forecast
- 70 levels
- 4 times per day

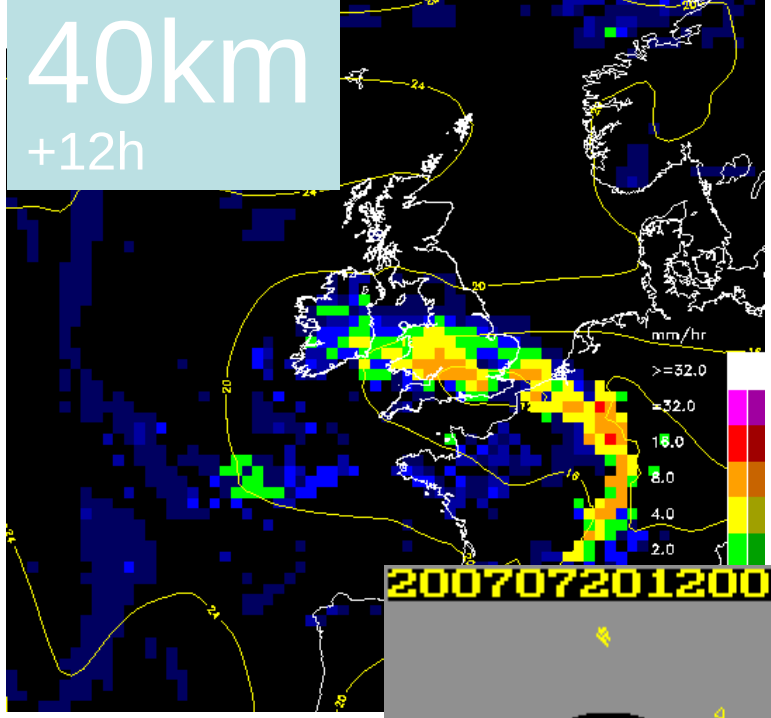
Regional 12km

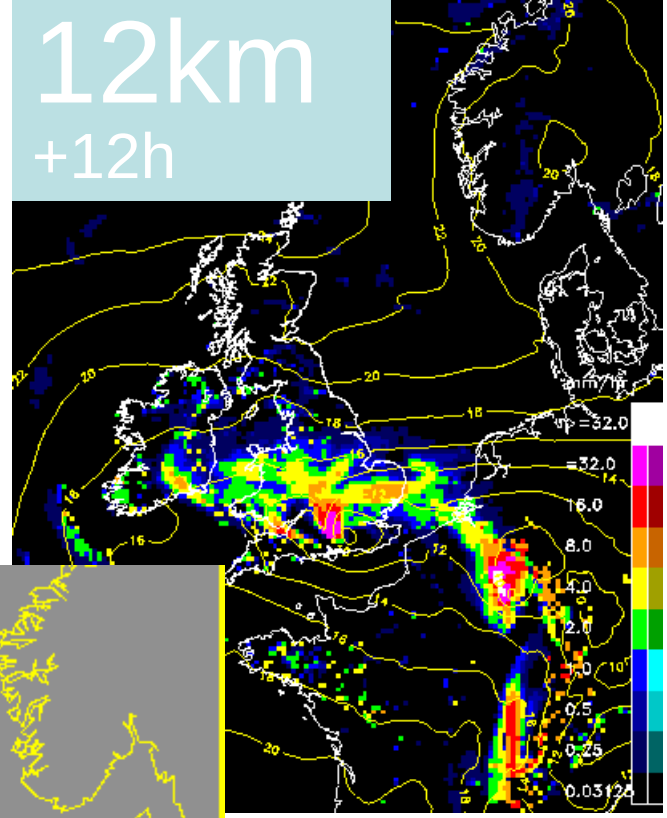
- 60 hour forecast
- 38 levels
- 4 times per day
- +EPS 24km, 24member

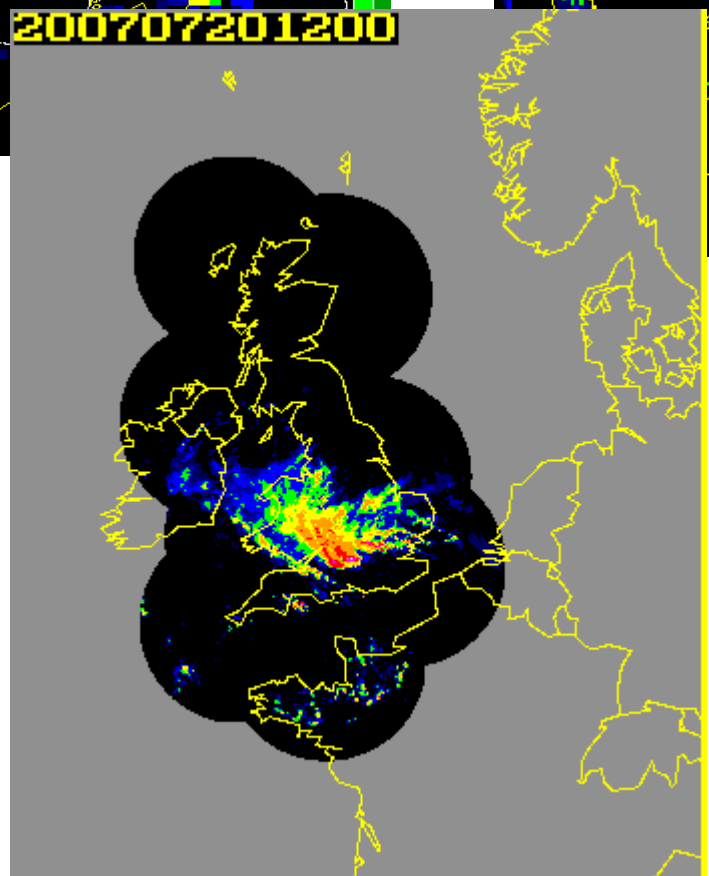
Global 40km

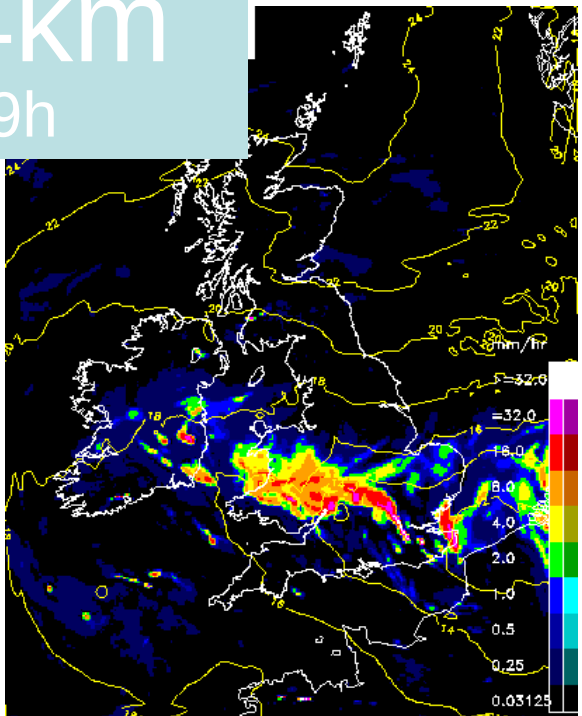
- 60 hour forecast twice/day
- 144 hour forecast twice/day
- +EPS 24member, 90km







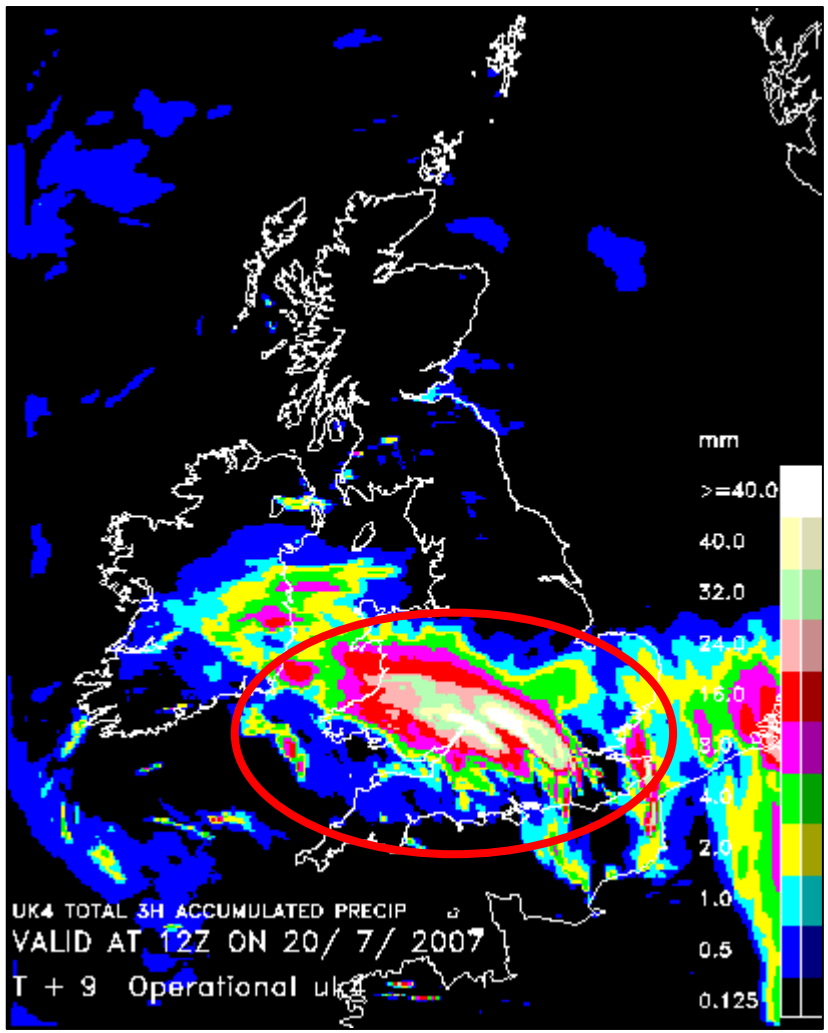




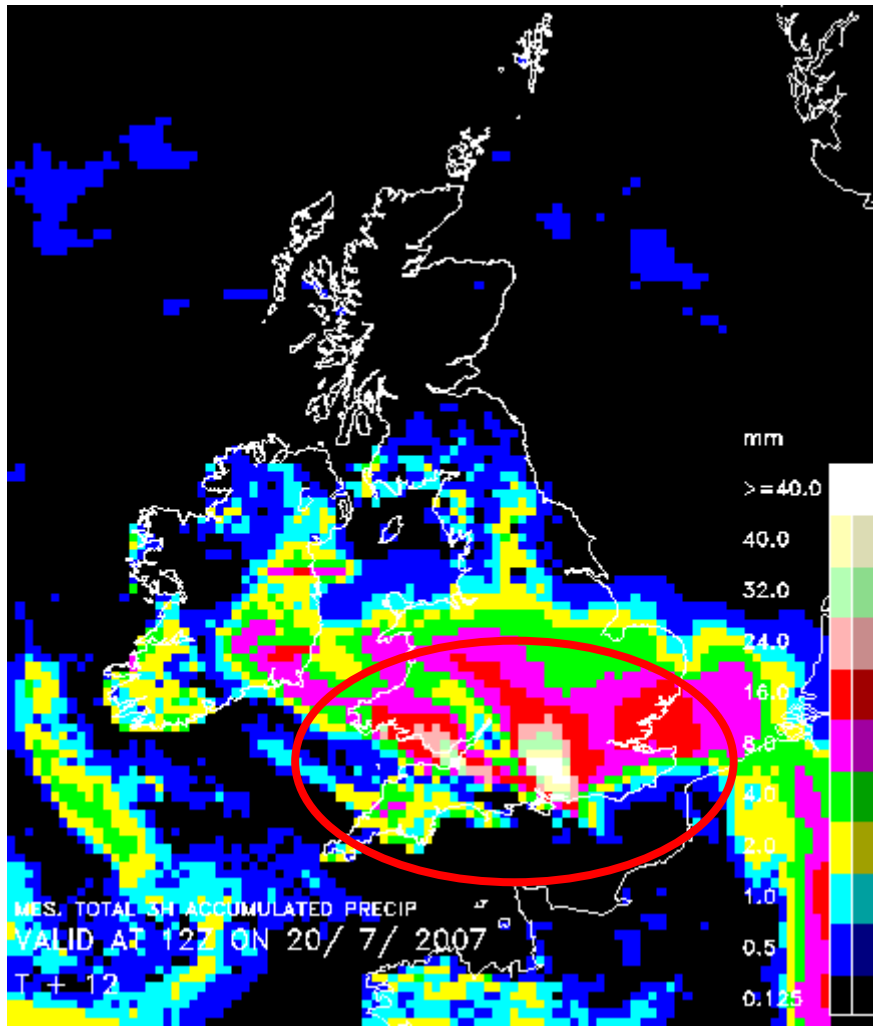
20 July 2007 Floods

3h accumulations

4km (6-9h)



12km (9-12h)

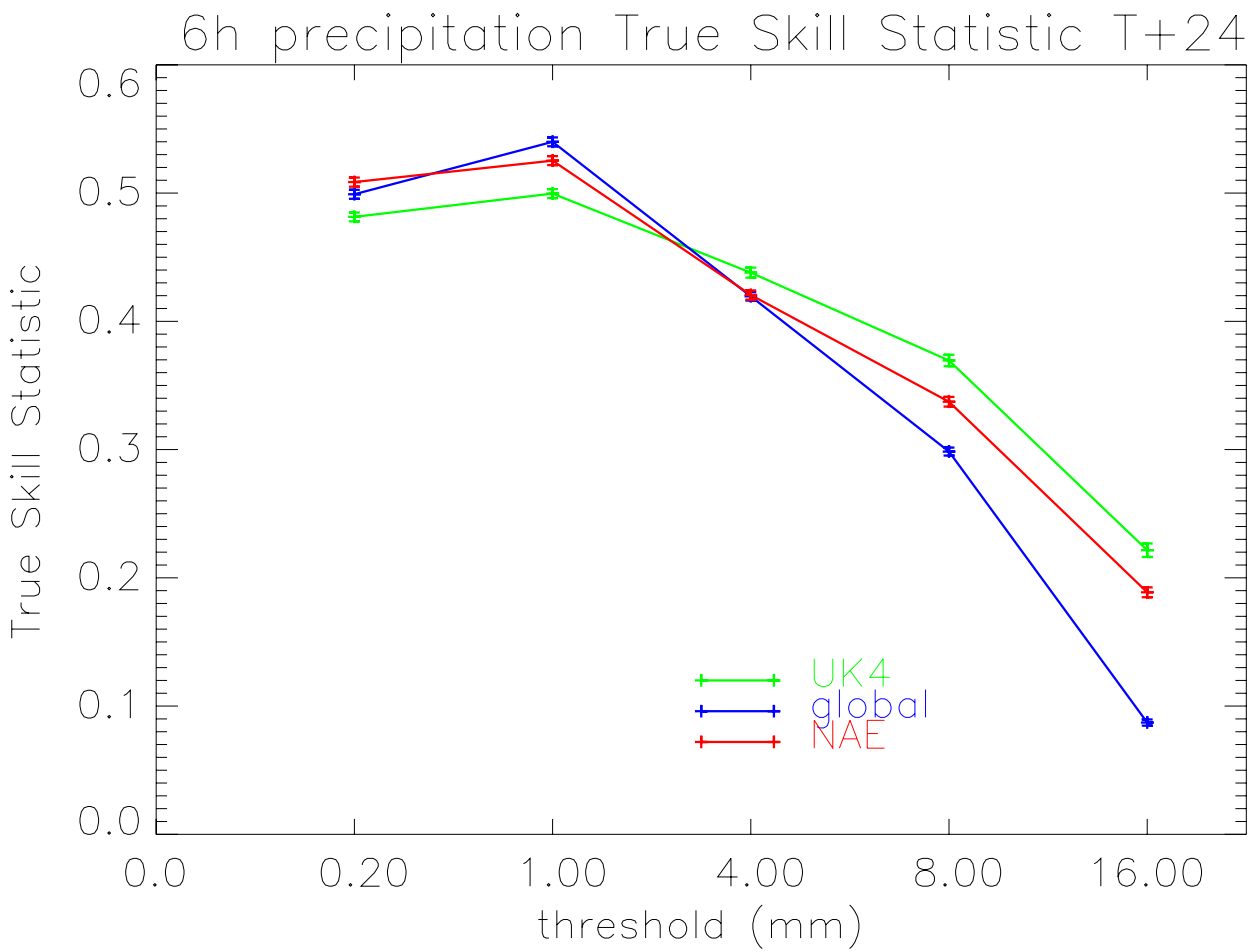


Standard Verification

Against synoptic station gauges – Inconsistent signals !

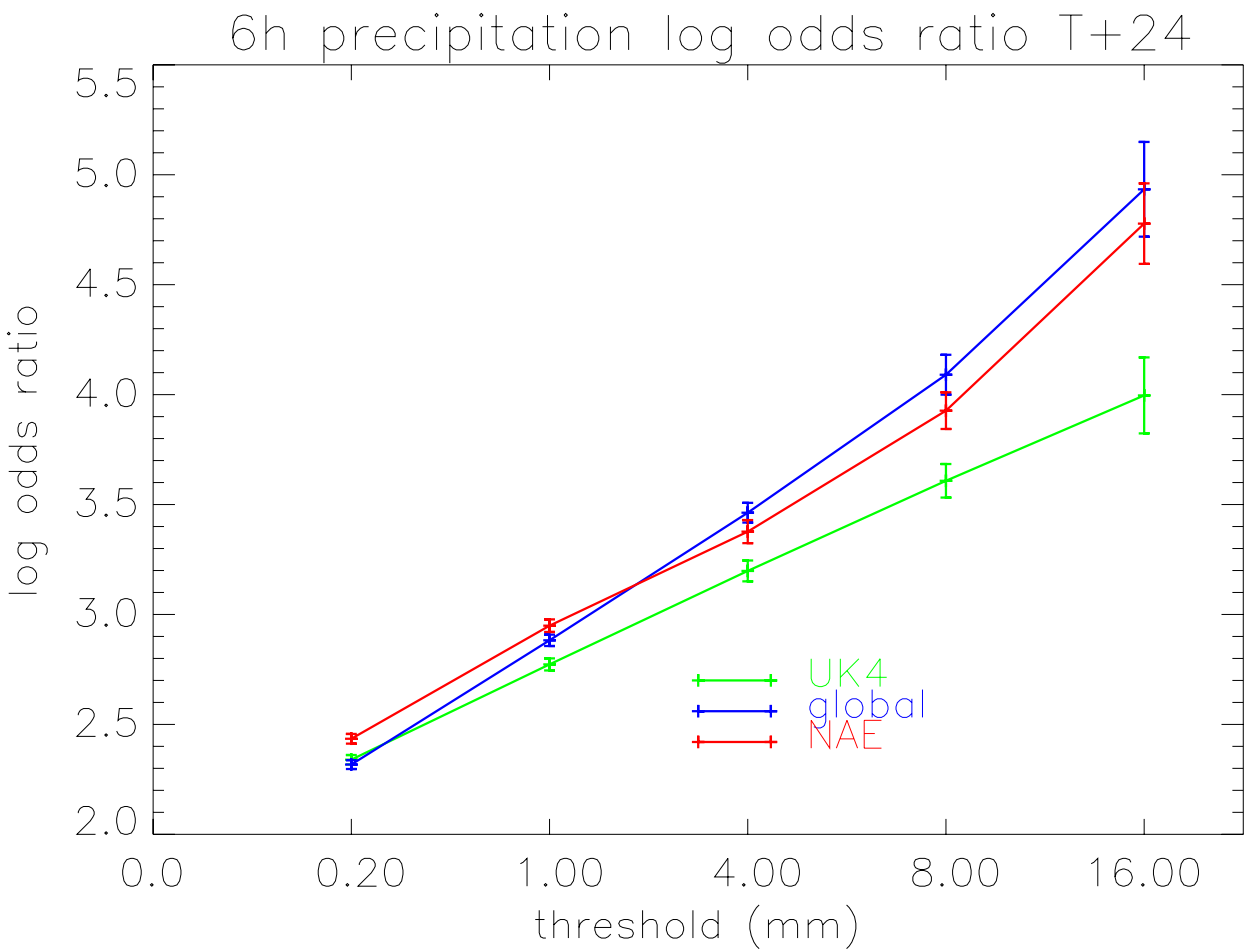
Standard UK area verification against surface stations- Nearest grid point – Dec06-Sep08

- True SS=Hit rate – False alarm rate
- UK4 better at higher thresholds



Standard UK area verification against surface stations- Nearest grid point – Dec06-Sep08

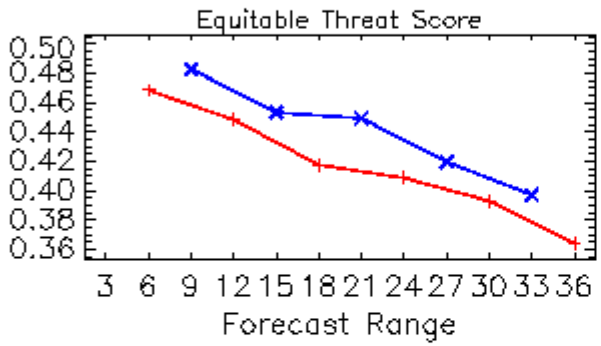
- Odds ratio increases as rarer events
- 40km model better at higher thresholds ?

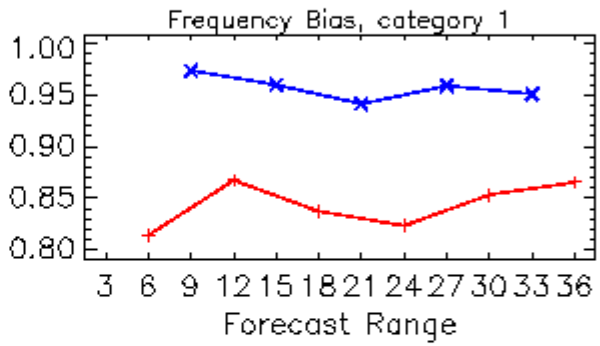


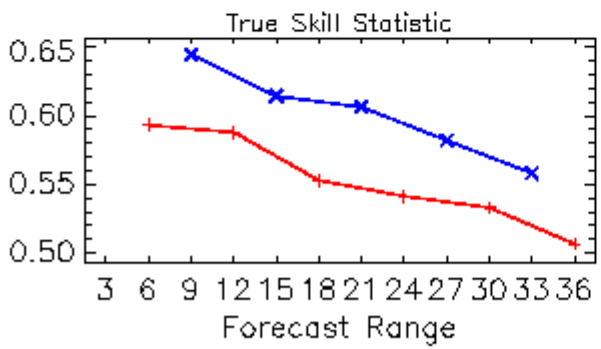
Operational verification v synops (Dec 2006)

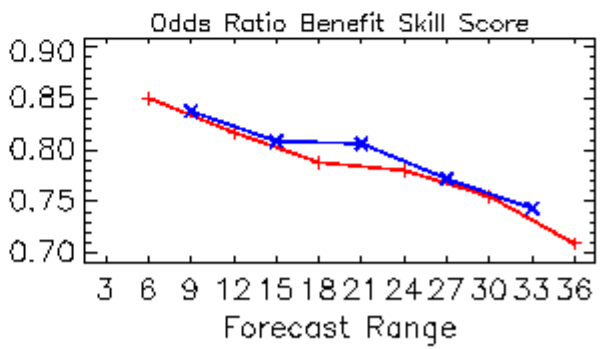
Combined times: 6hr Precip Accm (>= 1.0mm) (Corrected obs)
Combined dates from 01/12/2006 to 31/12/2006: Combined stations: Surface Obs

Cases: Operational NAE Operational UK4



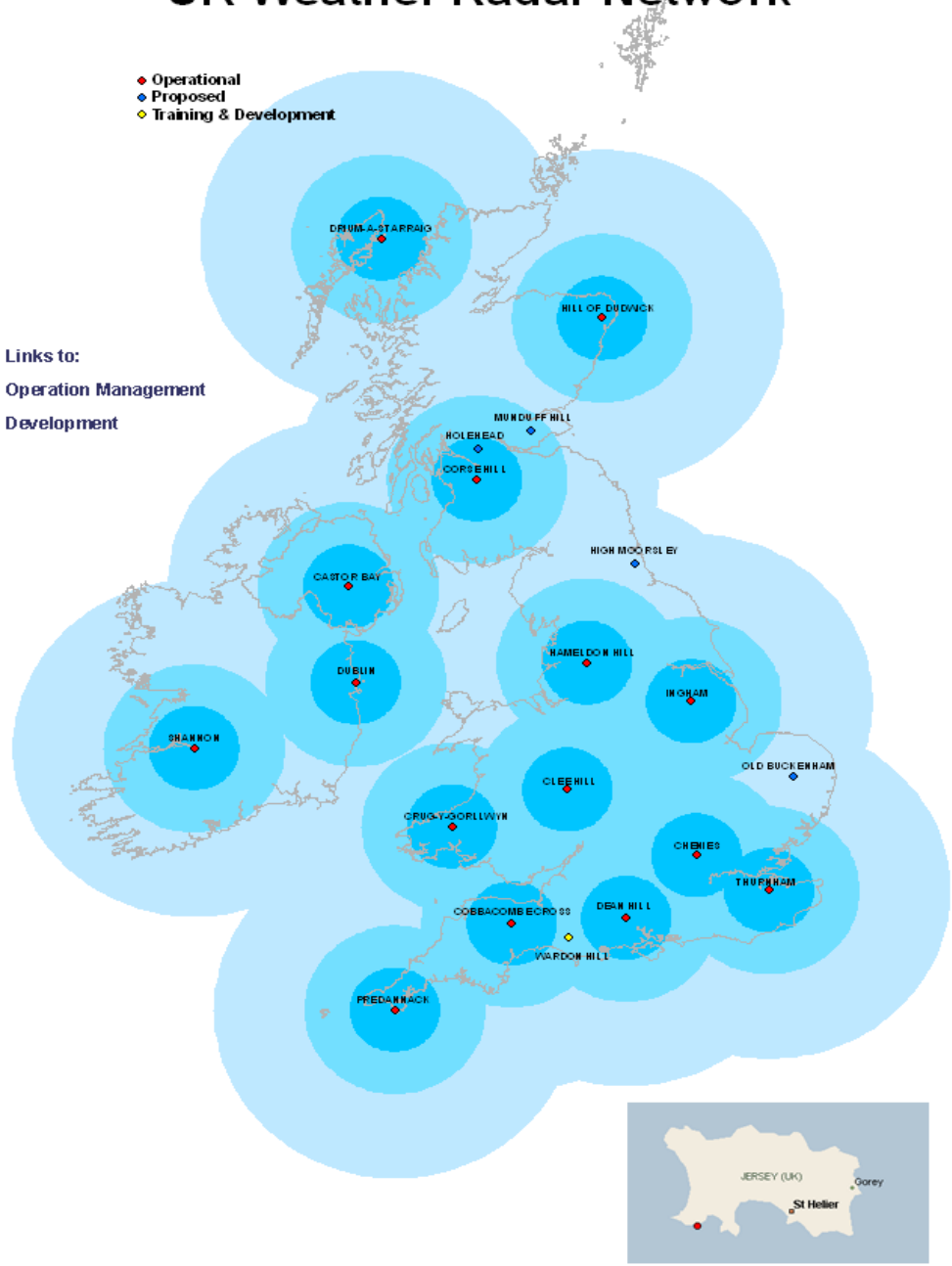






Radar composites & Gridded gauges

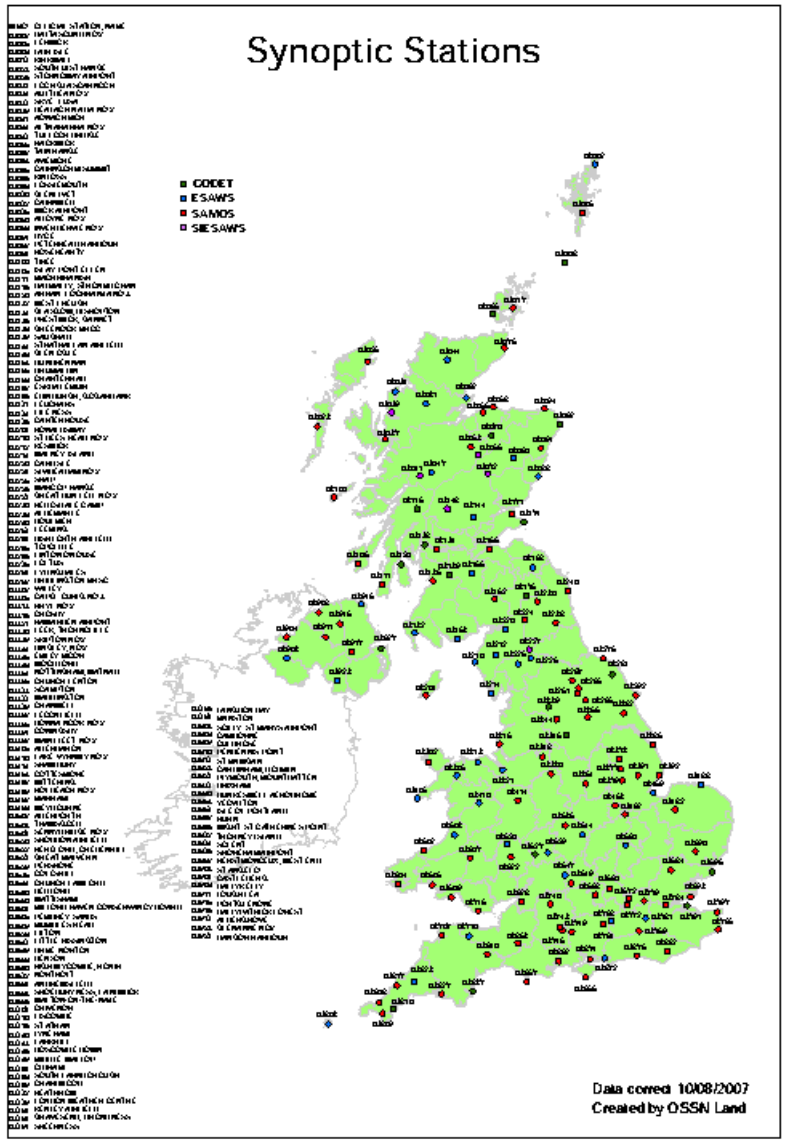
UK Weather Radar Network



Links to:

Operation Management

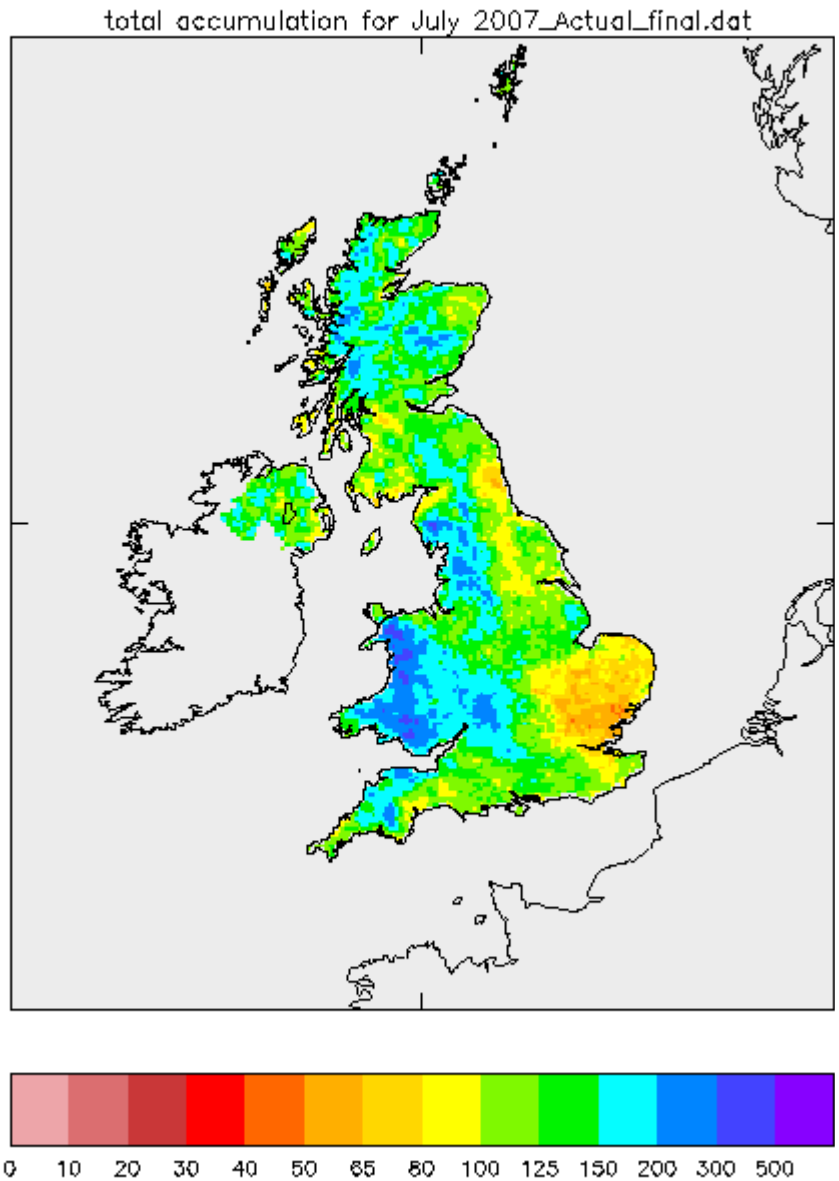
Development

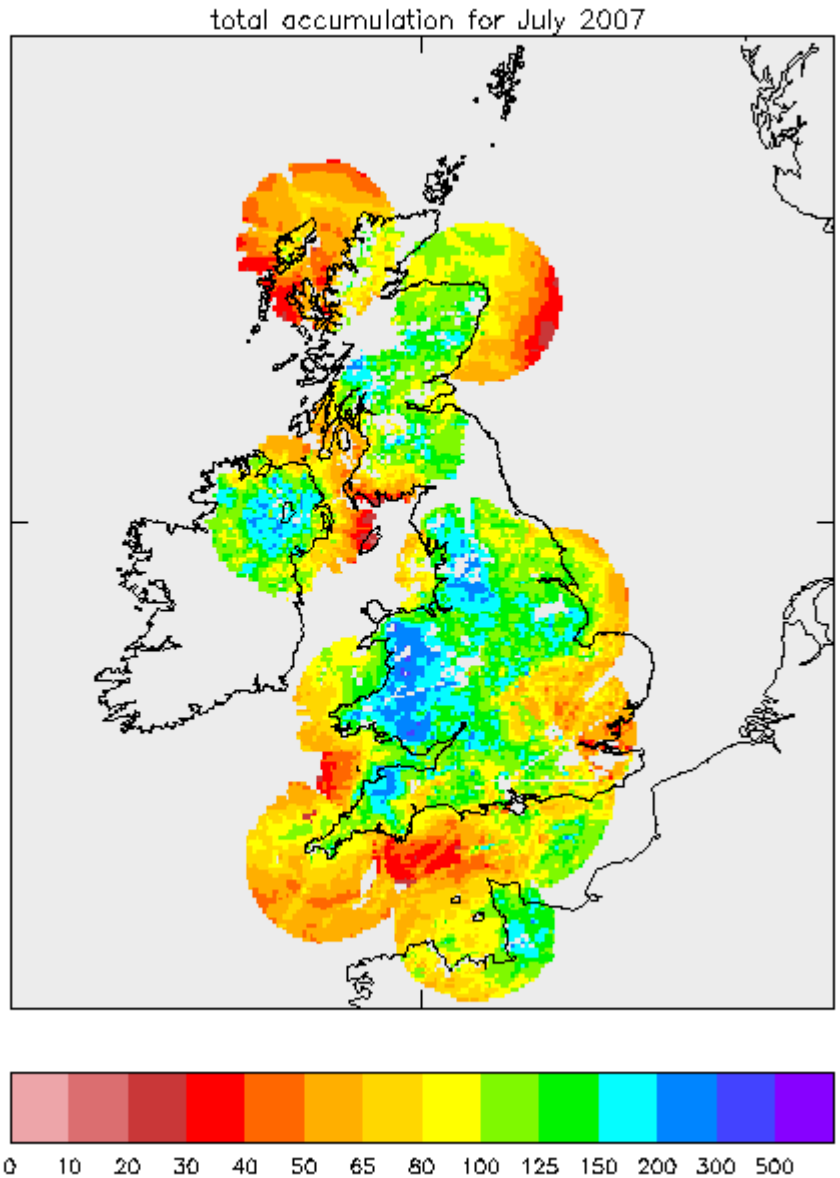


Uncertainties

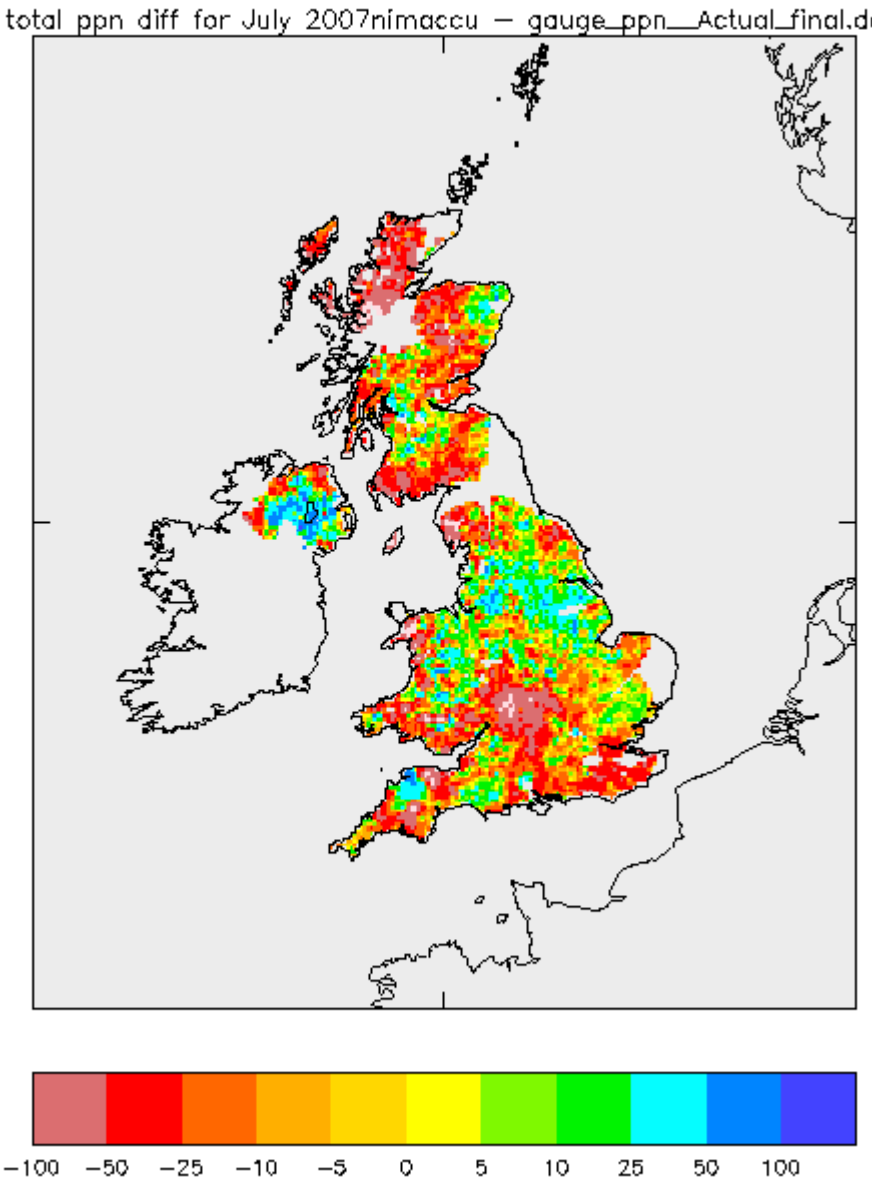
- Synoptic network sparse ~25- 50km
  - Nearest grid point model forecast
- Radar estimates ~-50%/+100% error
  - Averaged to 3x grid length = 15km
- Climatological precipitation stations
  - ~4400
  - 7 x7 km typical spacing
  - Gridded analysis -5x5km (Perry & Hollis,2005,Int J Climatol)
    - Monthly
    - daily

Gridded Gauges July 2007 Radar

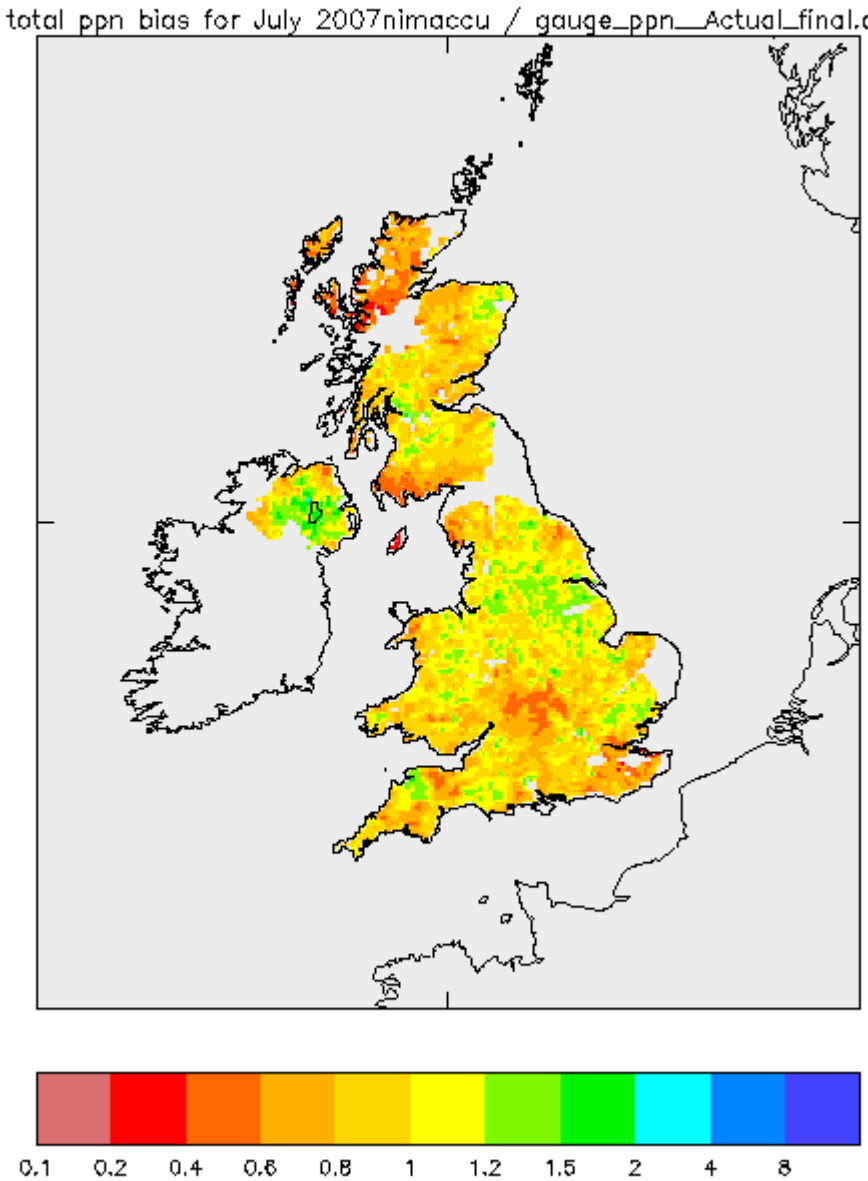




Radar – gauge (mm) July 2007

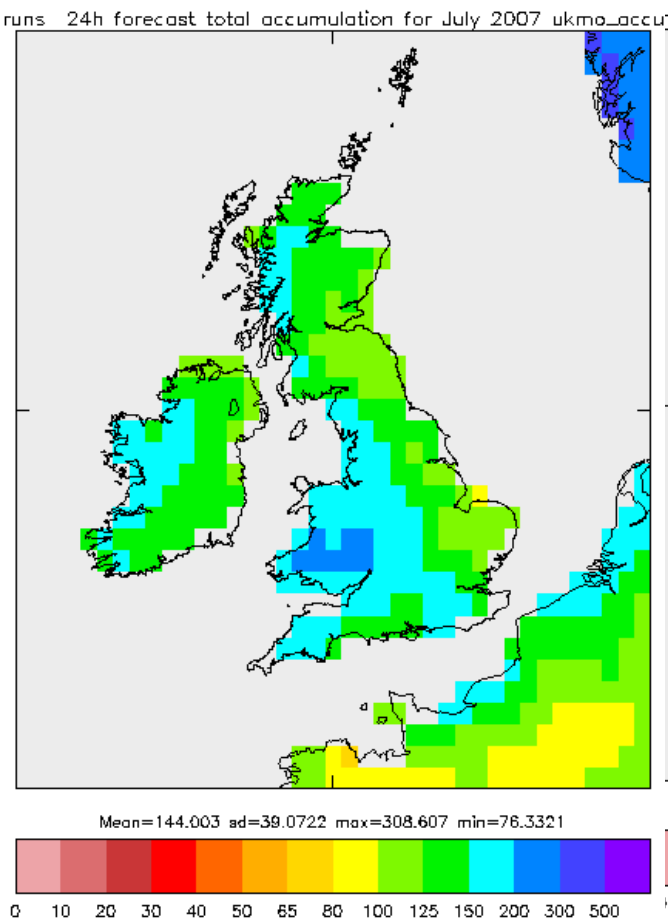


Radar/gauge

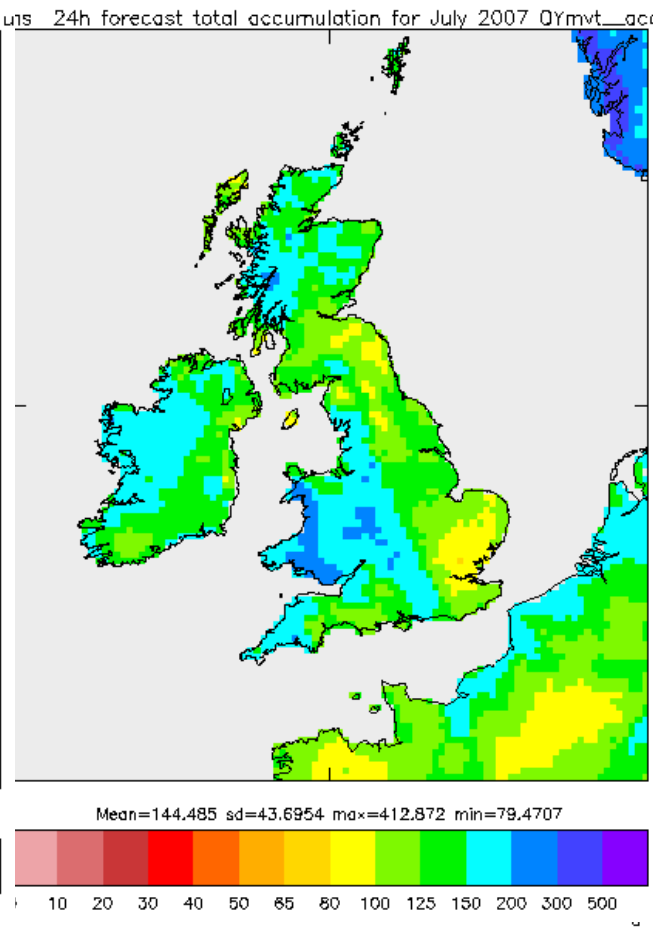


Forecast monthly totals - July

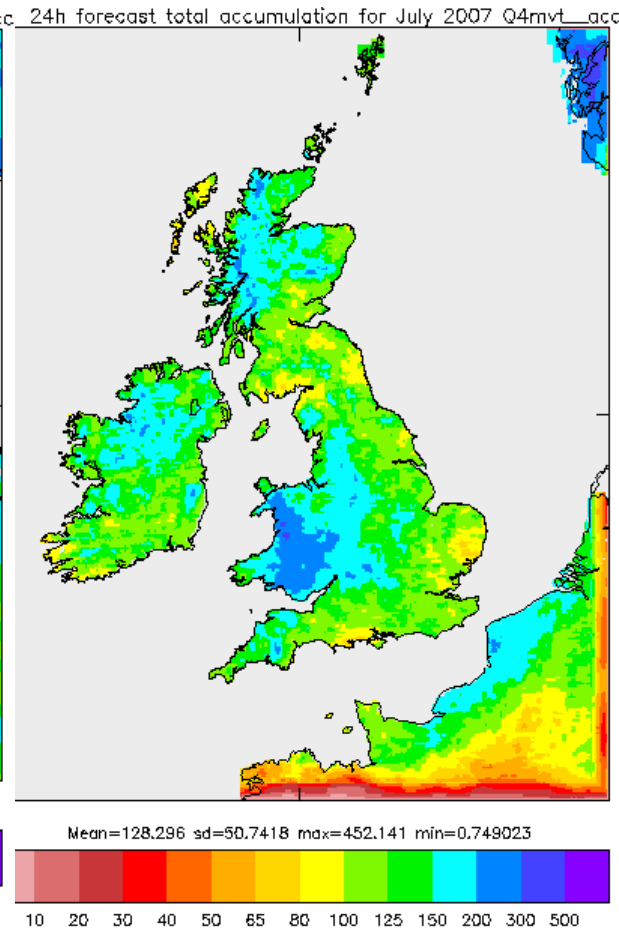
40 km



12km

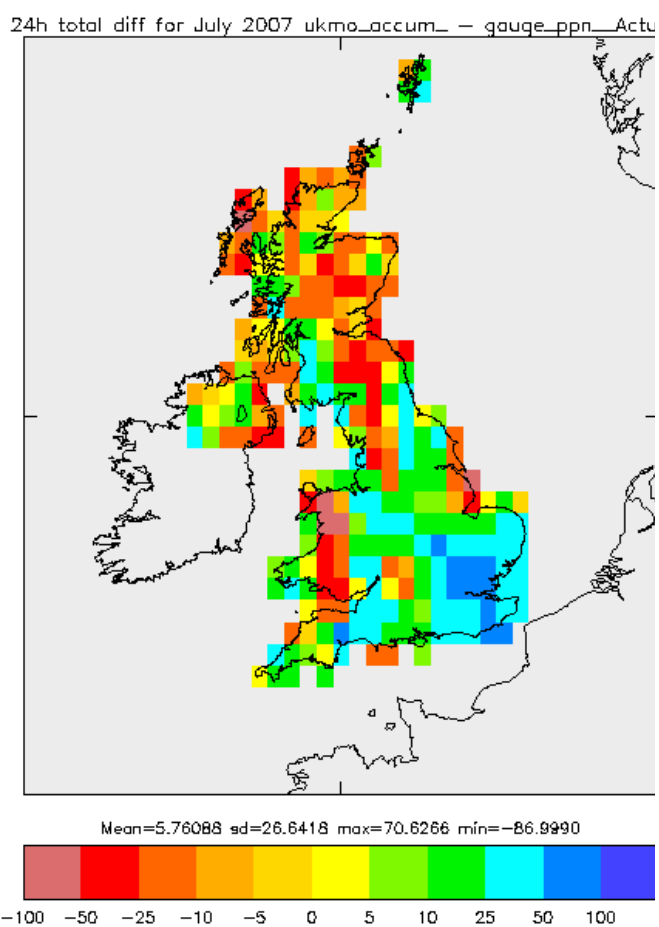


4km

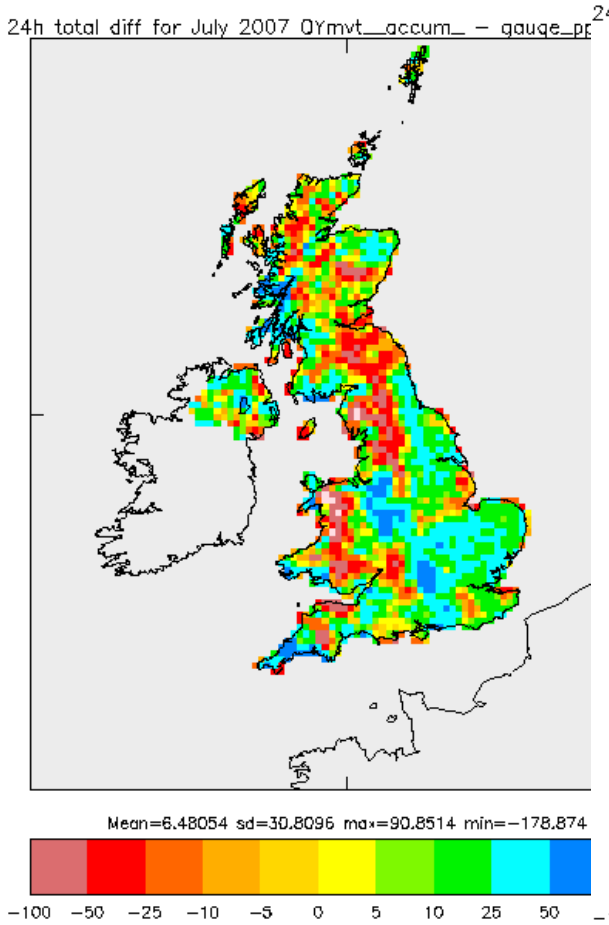


Forecast monthly totals – July Errors v gauges

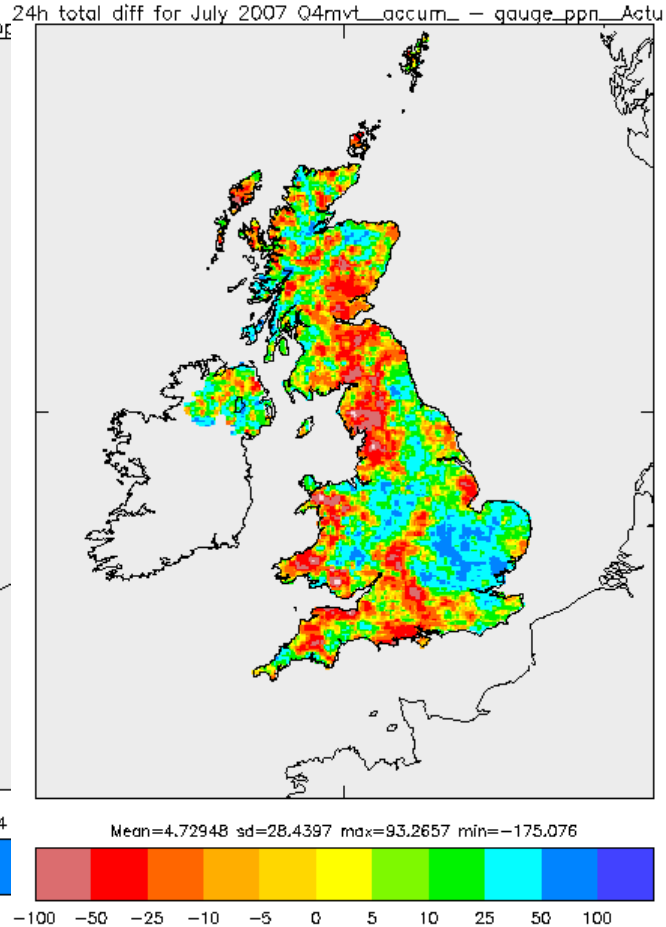
40 km



12km

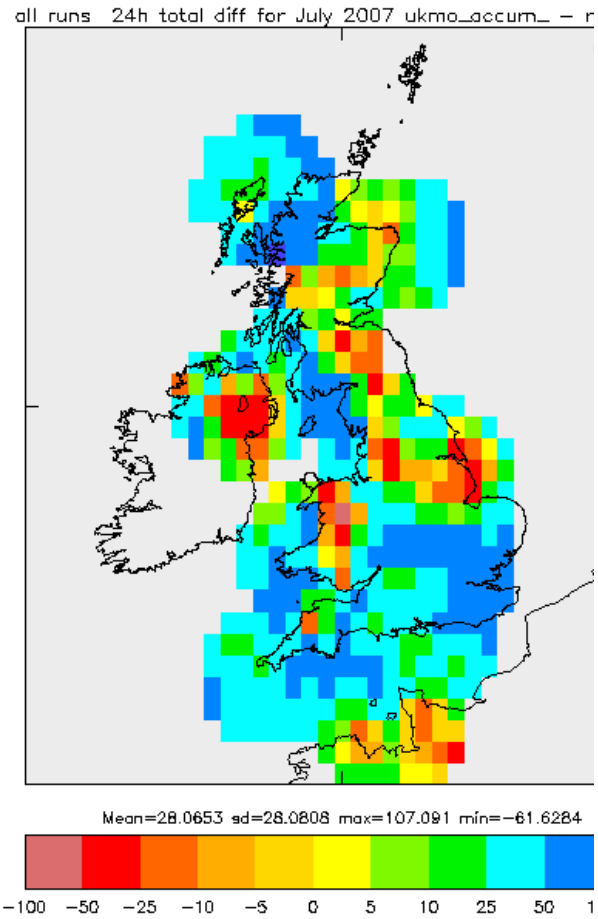


4km

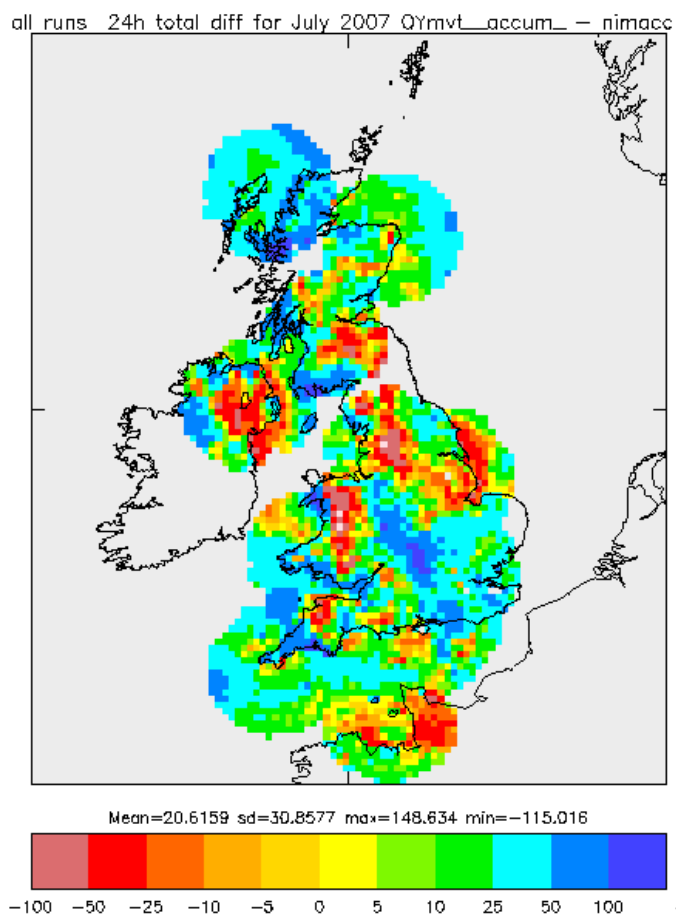


Forecast monthly totals – July Errors v radar

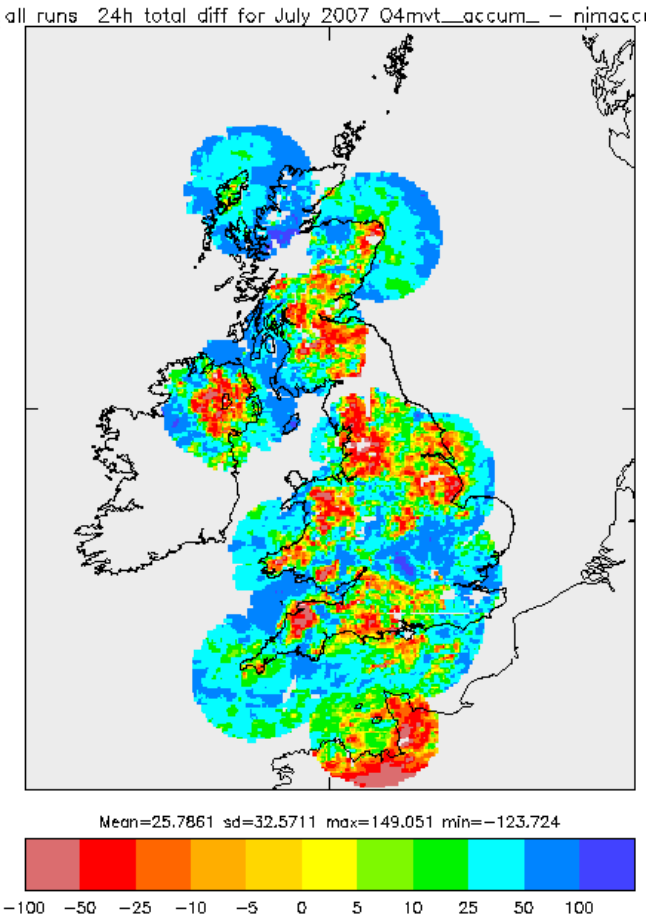
40 km

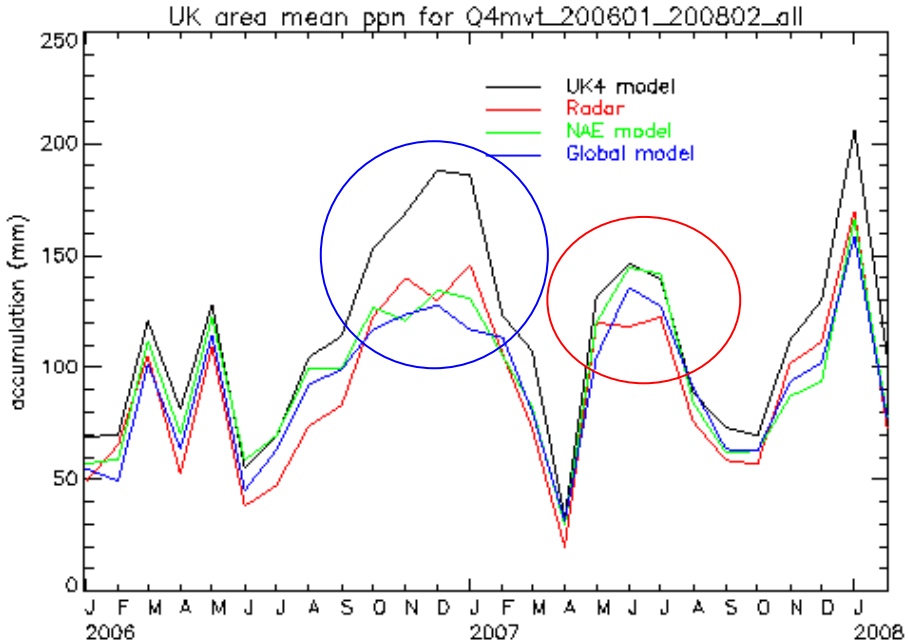


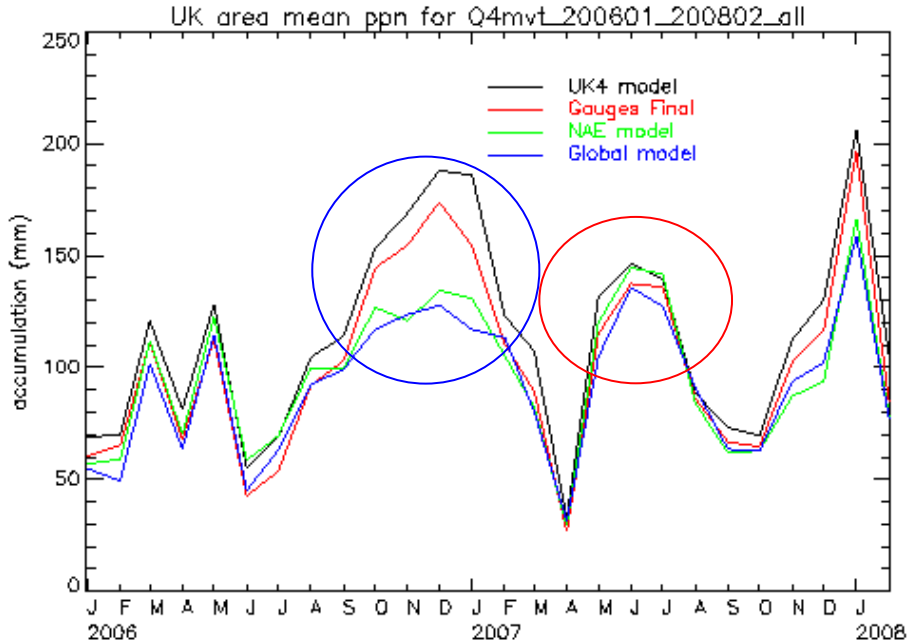
12km

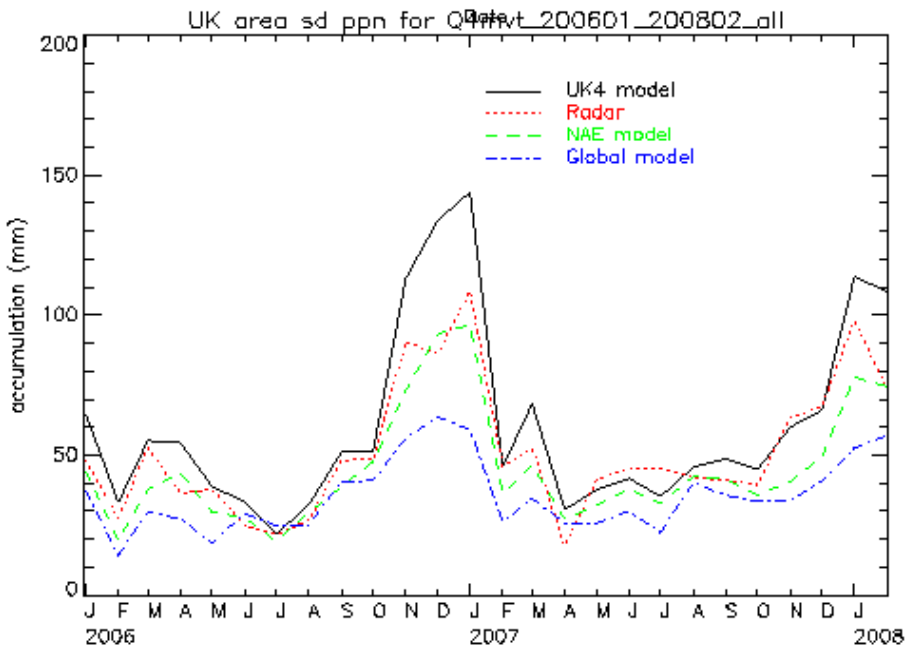


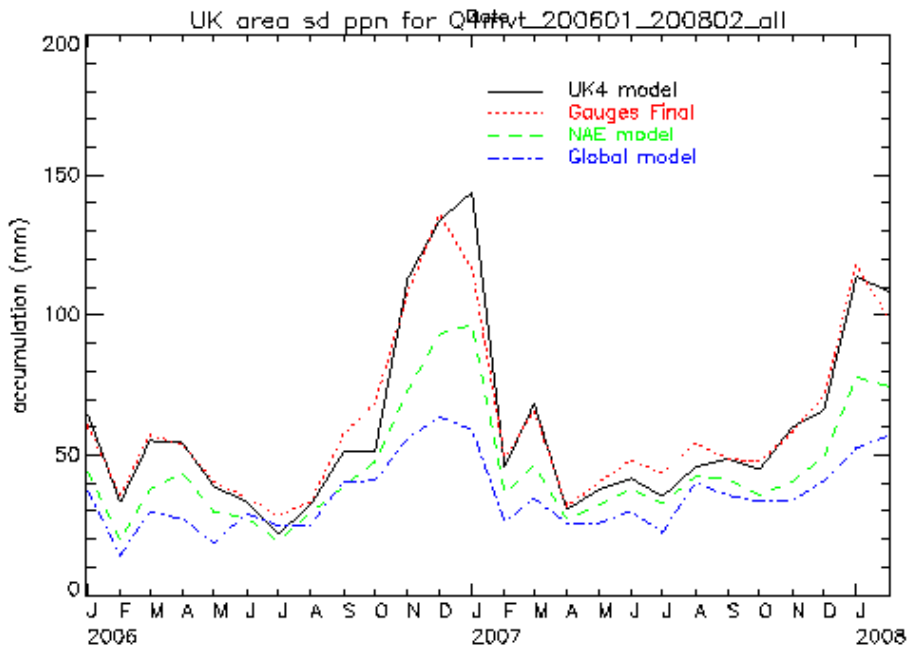
4km



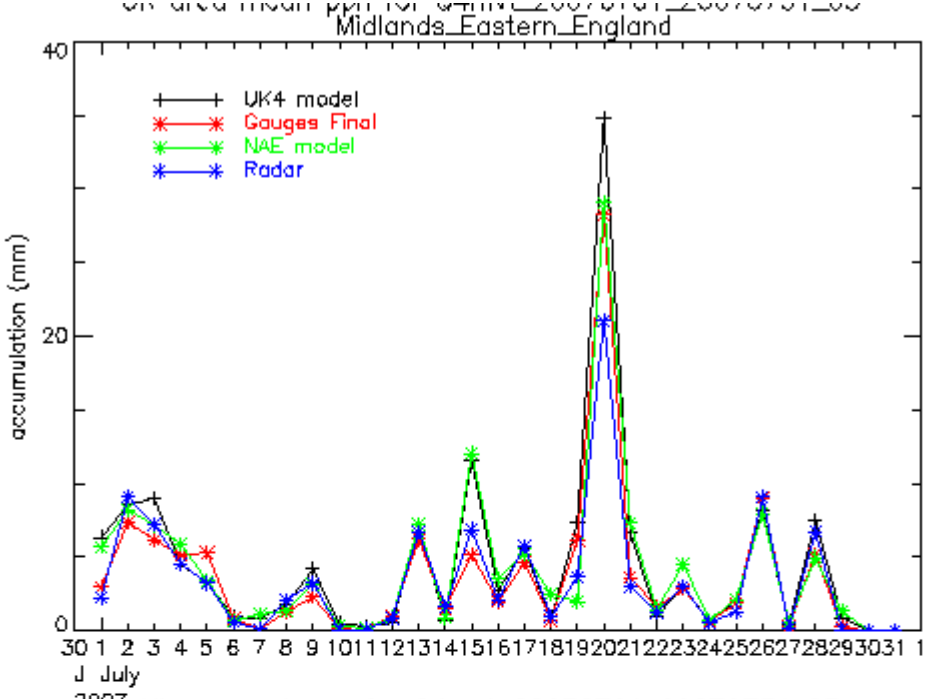


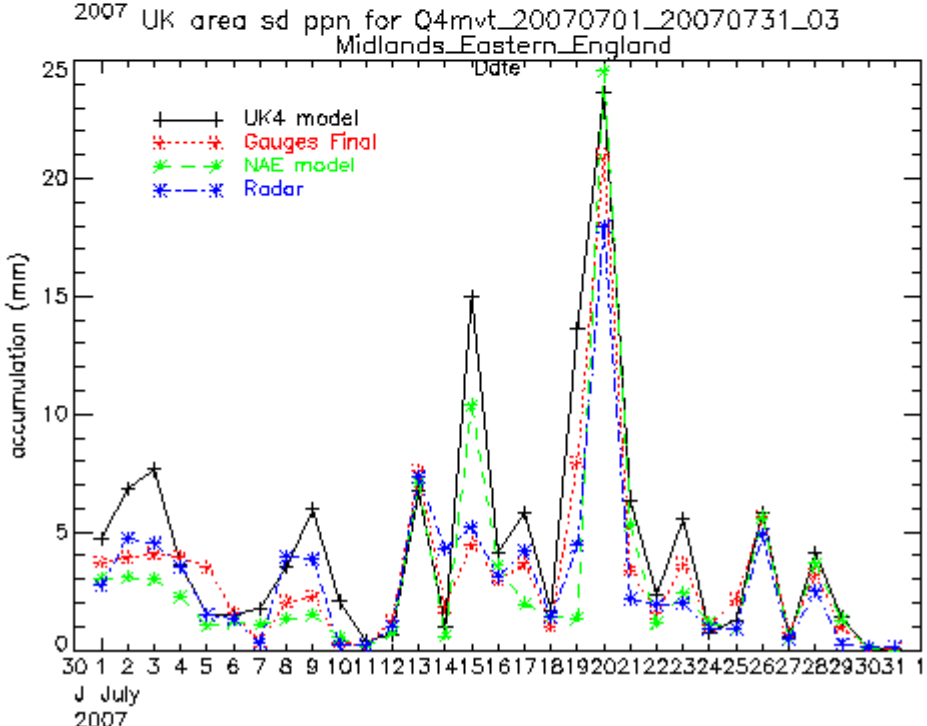






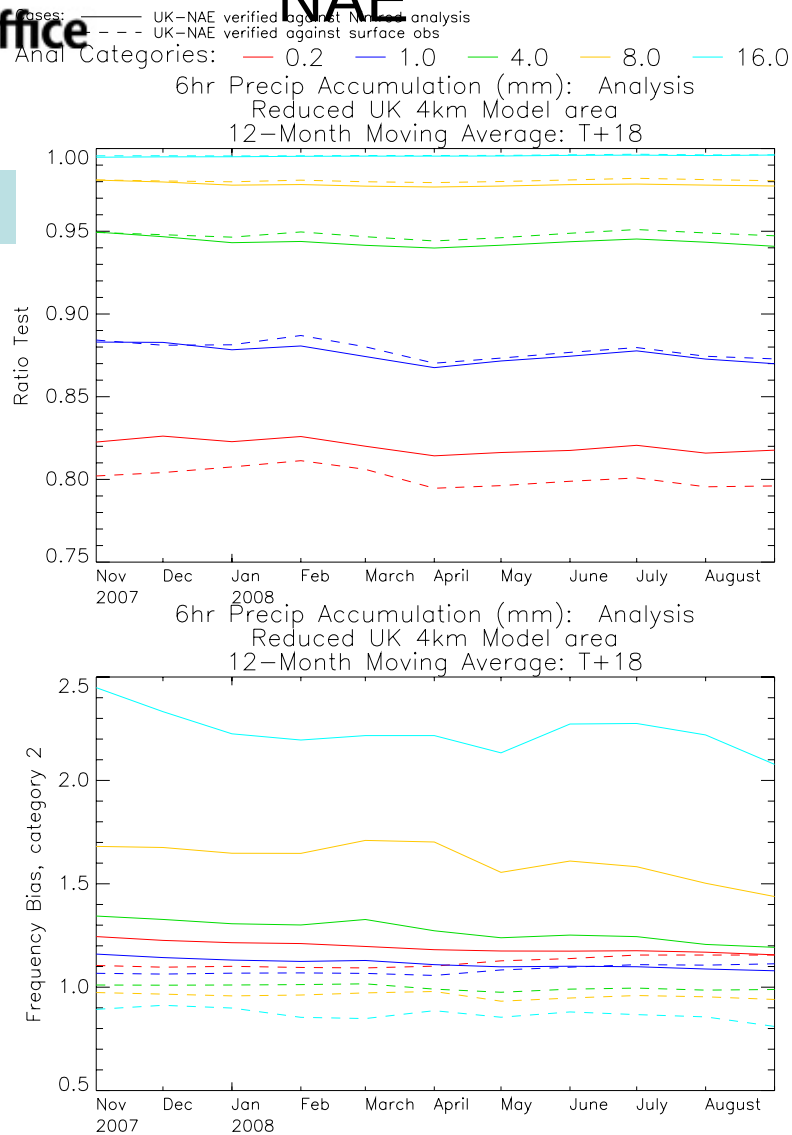
Daily rainfall July07



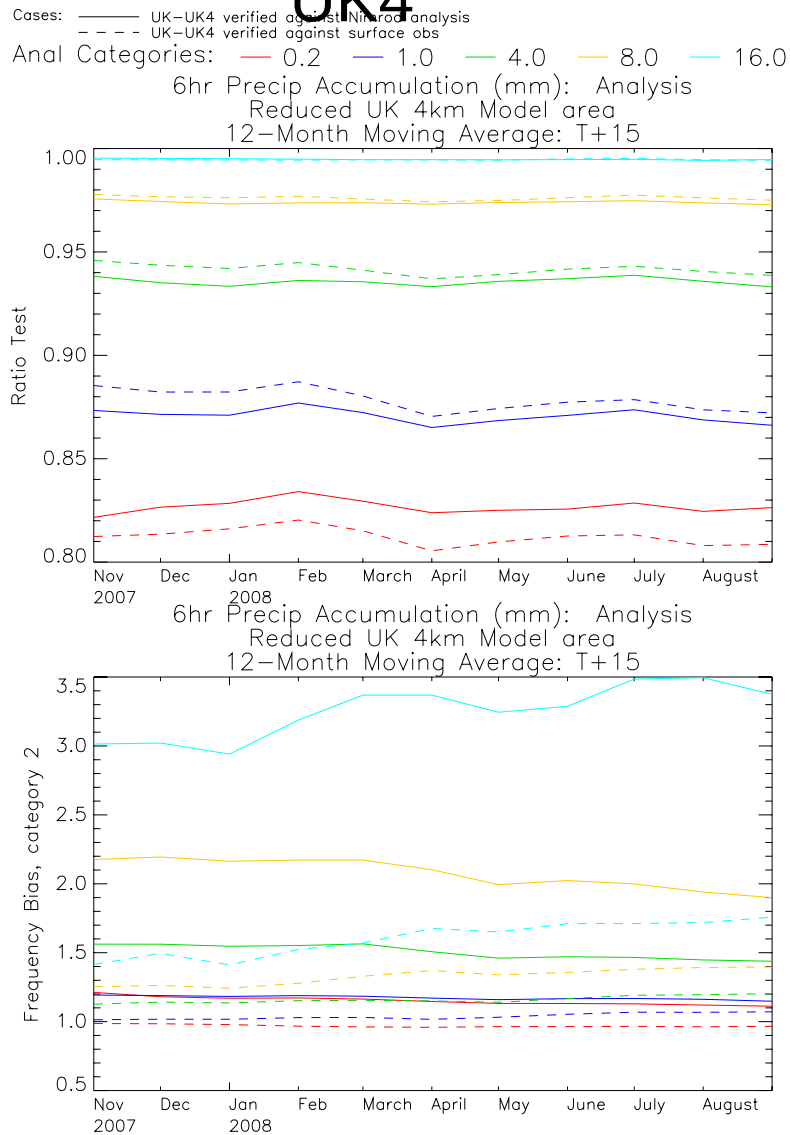


Comparison NAE & UK4 against stations (dashed) & radar composites (full) –Dec06- Sep08

NAE



UK4

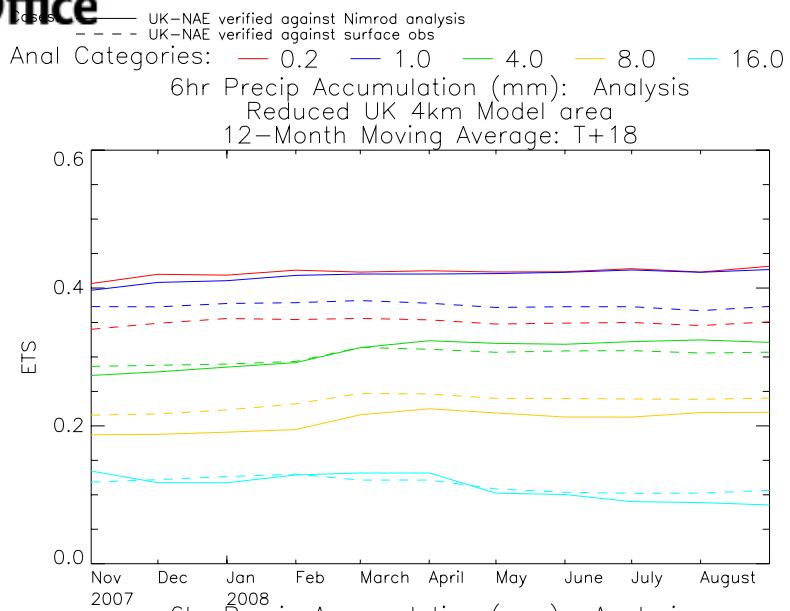


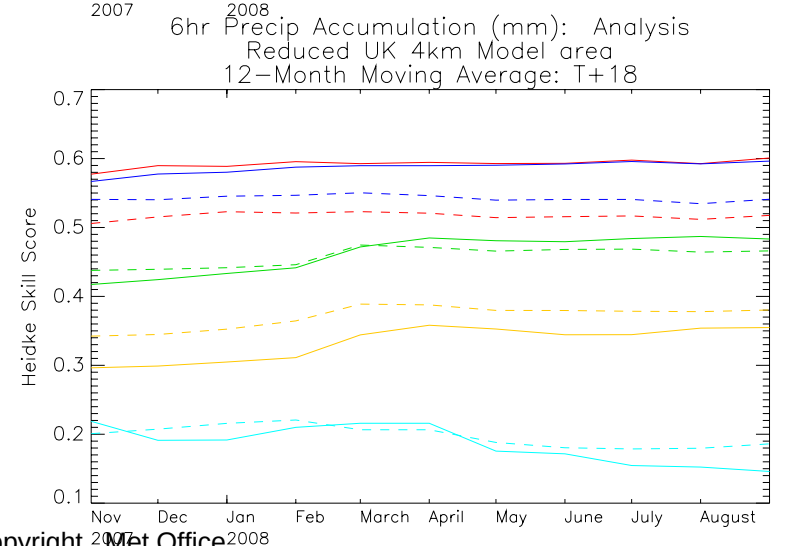
%correct

bias

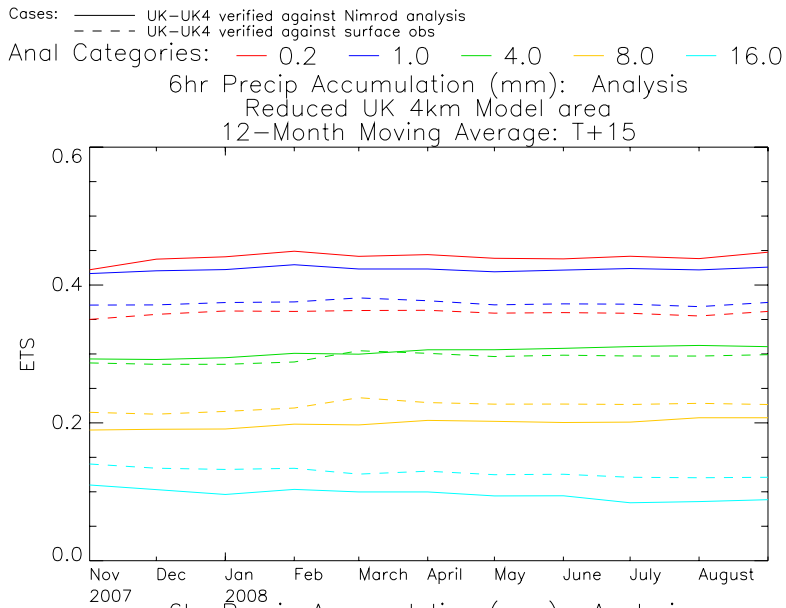
Comparison NAE & UK4 against stations (dashed) & radar composites (full) –Dec06- Sep08

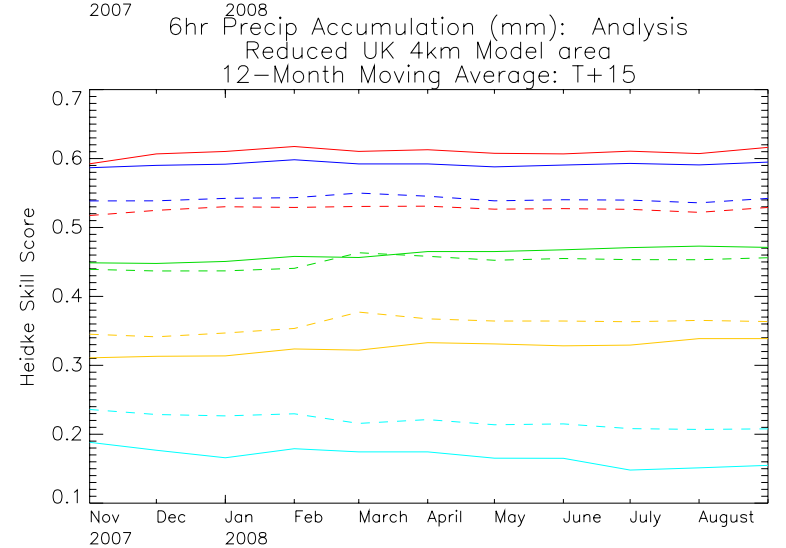
NAE





UK4





ETS

Heidke

“Fuzzy” Verification

Against radar composites

Intensity/scale verification Barbara Casati et al

- Threshold precipitation forecast and analysis X >u
  - Intensity, u
- Form Binary error & analyse with Haar wavelets (scale, l)

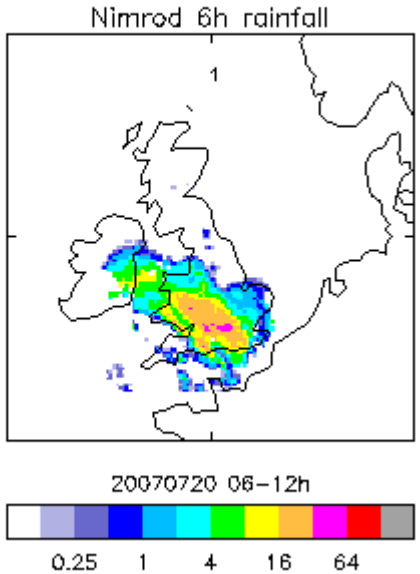
$$E_u = \sum_{l=1}^{L} E_{u,l}$$

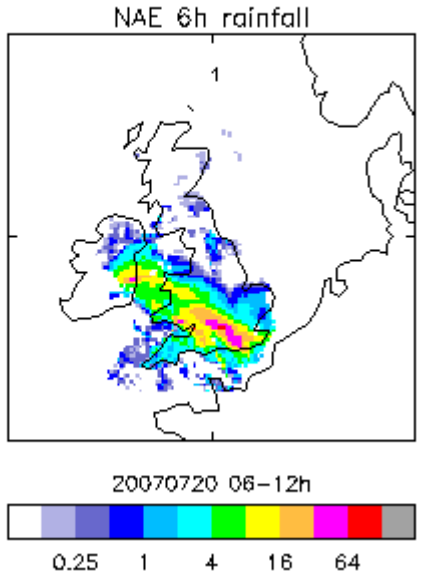
- MSE of binary error sum of each length scale error (wavelet orthogonality)

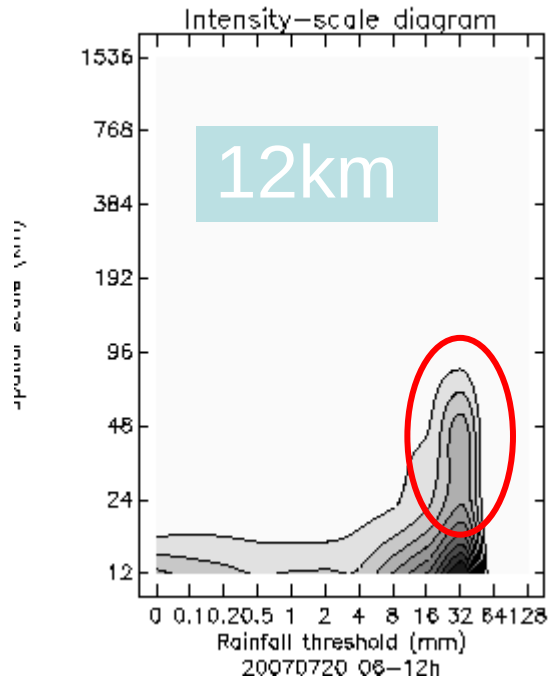
$$MSE_u = \sum_{l=1}^{L} MSE_{u,l}$$

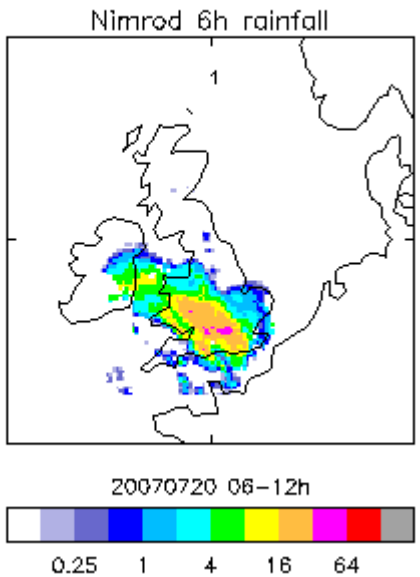
- Skill compared to random forecast with no scale dependency

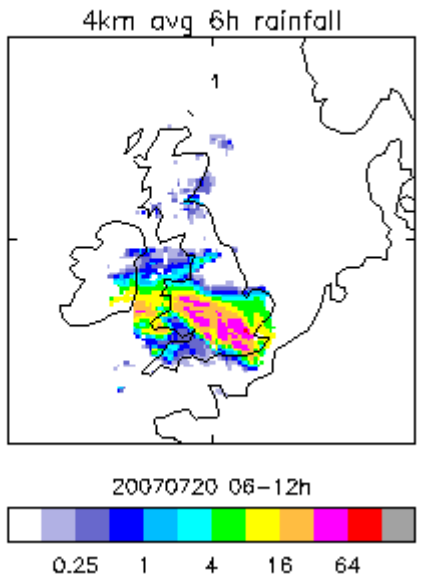
$$SS_{u,l} = 1 - \frac{MSE_{u,l}}{2\varepsilon(1-\varepsilon)/L} \qquad \varepsilon = \frac{a+c}{n} = P(X > u)$$

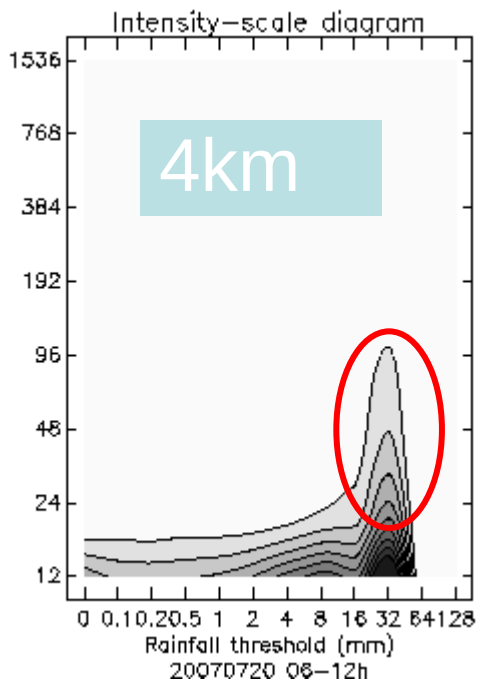






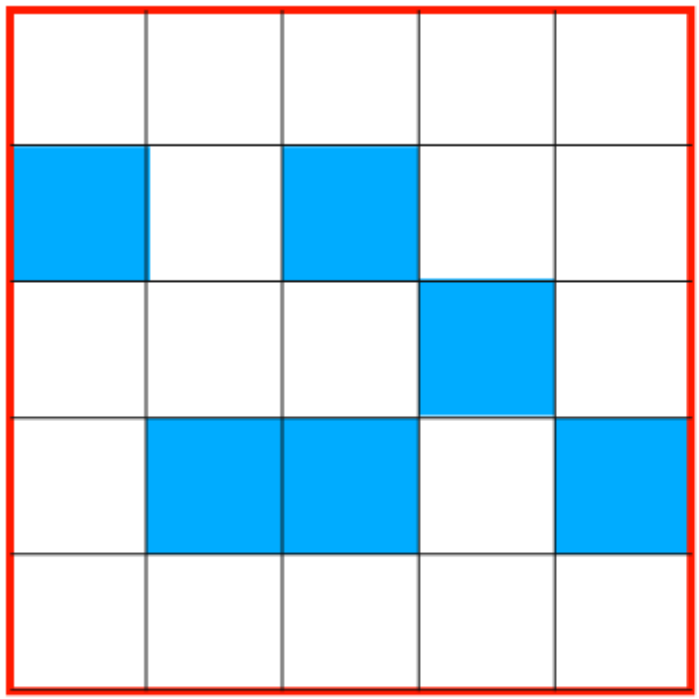






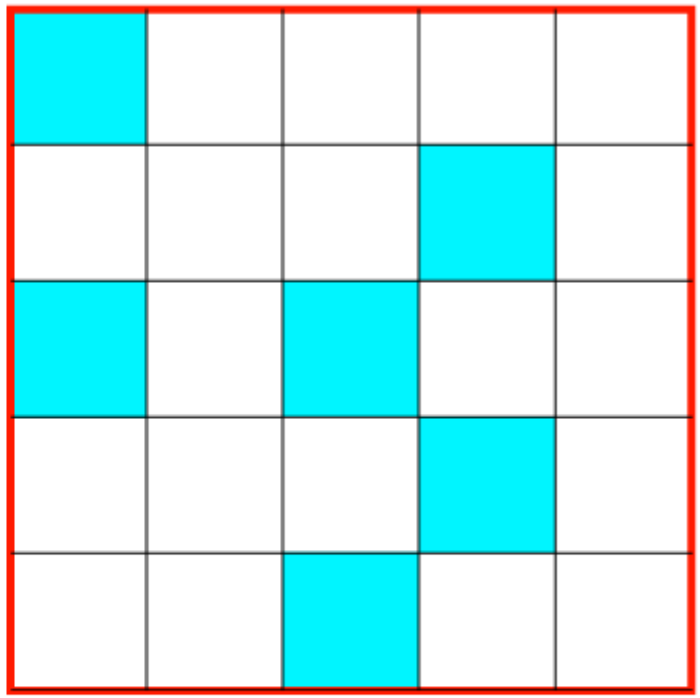
Schematic comparison of fractions (Nigel Roberts)

observed



Fraction = 6/25 = 0.24

forecast



Fraction = 6/25 = 0.24

A score for comparing fractions with fractions (Nigel Roberts)

Brier score for comparing fractions

$$\text{FBS} \quad \text{(Fractions Brier Score)} \qquad = \frac{1}{N} \sum_{j=1}^{N} (p_j - o_j)^2$$

$0 \leqslant p_j \leqslant 1$ forecast fractions

$0 \leqslant o_j \leqslant 1$ radar fractions

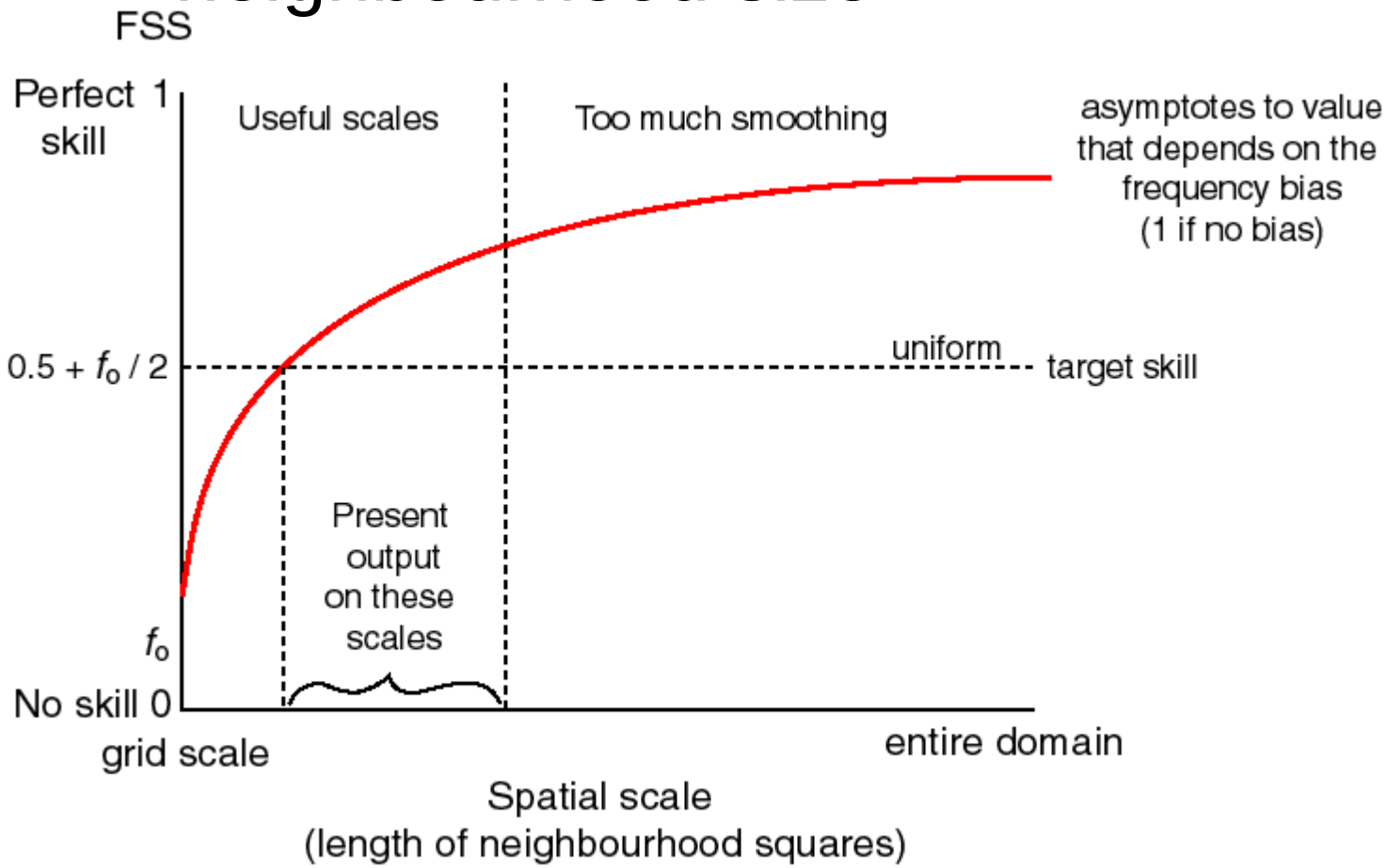
N number of points

Skill score for fractions/probabilities - Fractions Skill Score (FSS)

$$\text{FSS} = 1 - \frac{\text{FBS}}{\frac{1}{N}\left[\sum_{j=1}^{N} (p_j)^2 + \sum_{j=1}^{N} (o_j)^2\right]}$$

No overlap of fractions

Example graph of FSS against neighbourhood size

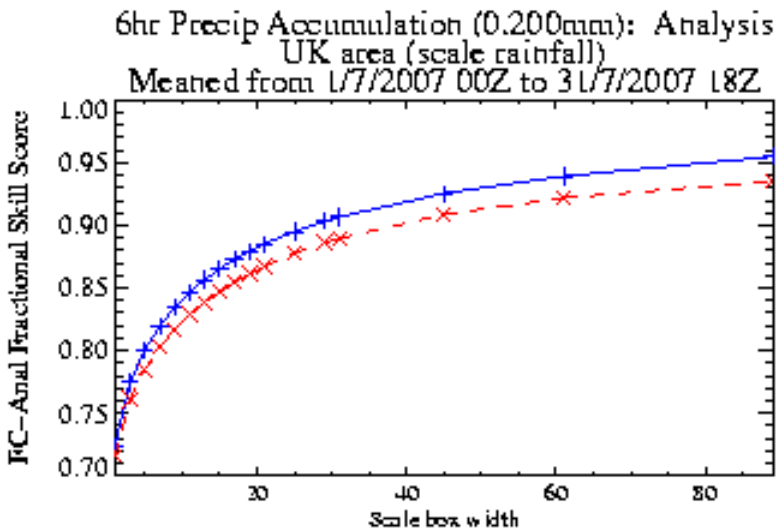


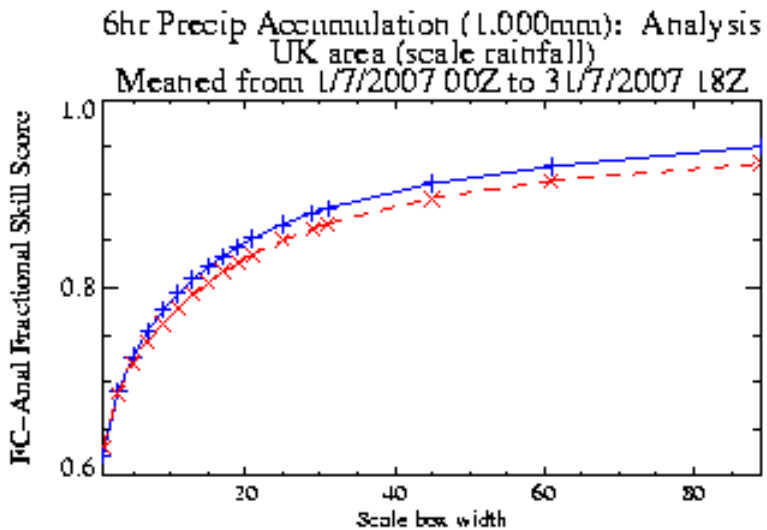
FSS JULY 2007

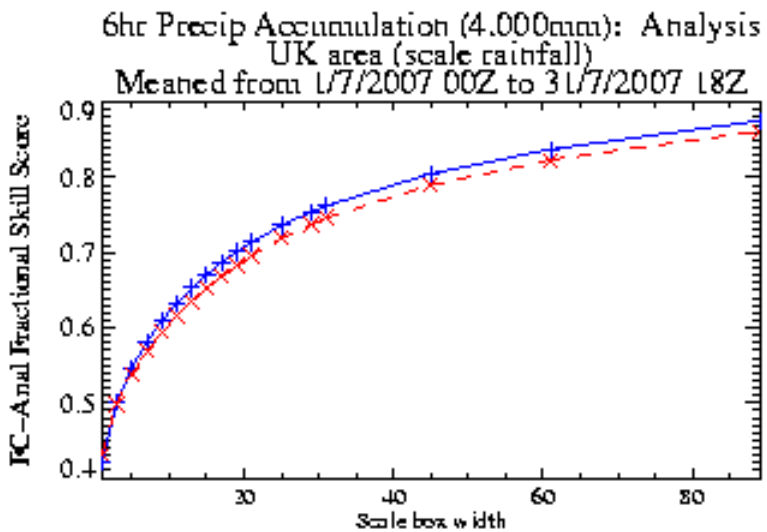
12km

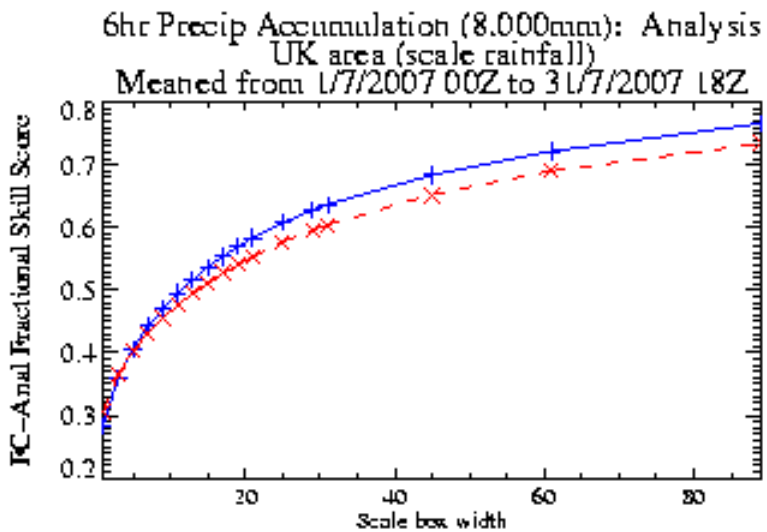
4km

FCRanges: T+9 T+12







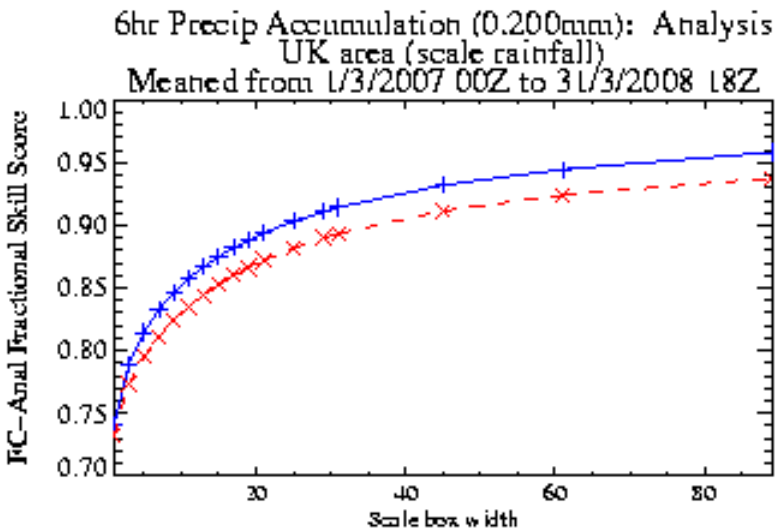


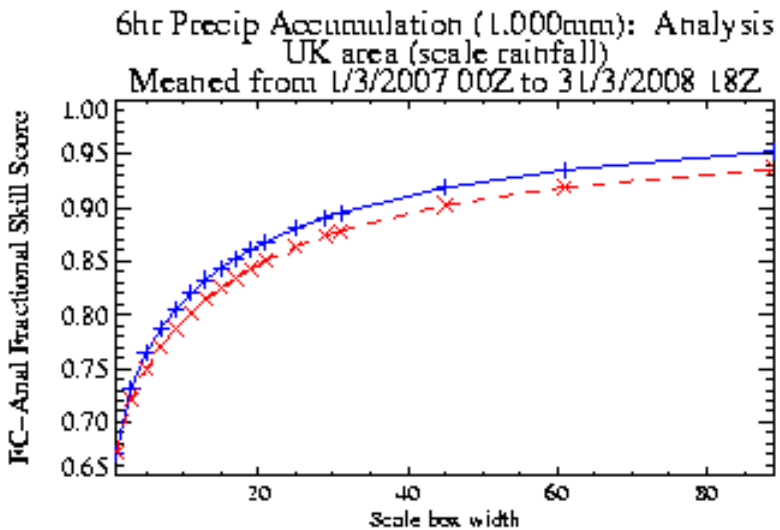
FSS Mar07-Mar08

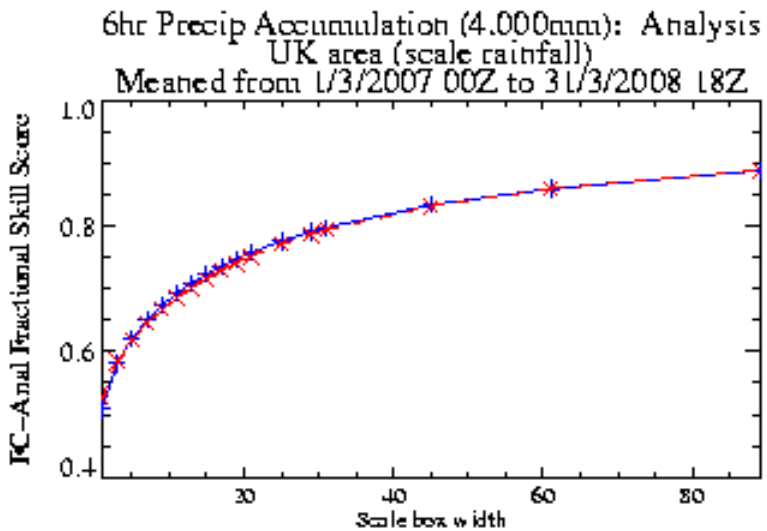
12km

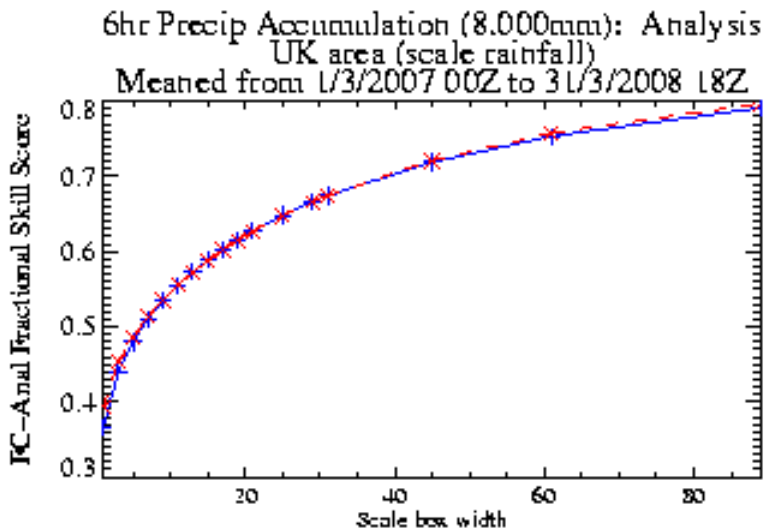
4km

FCRanges: T+9 T+12





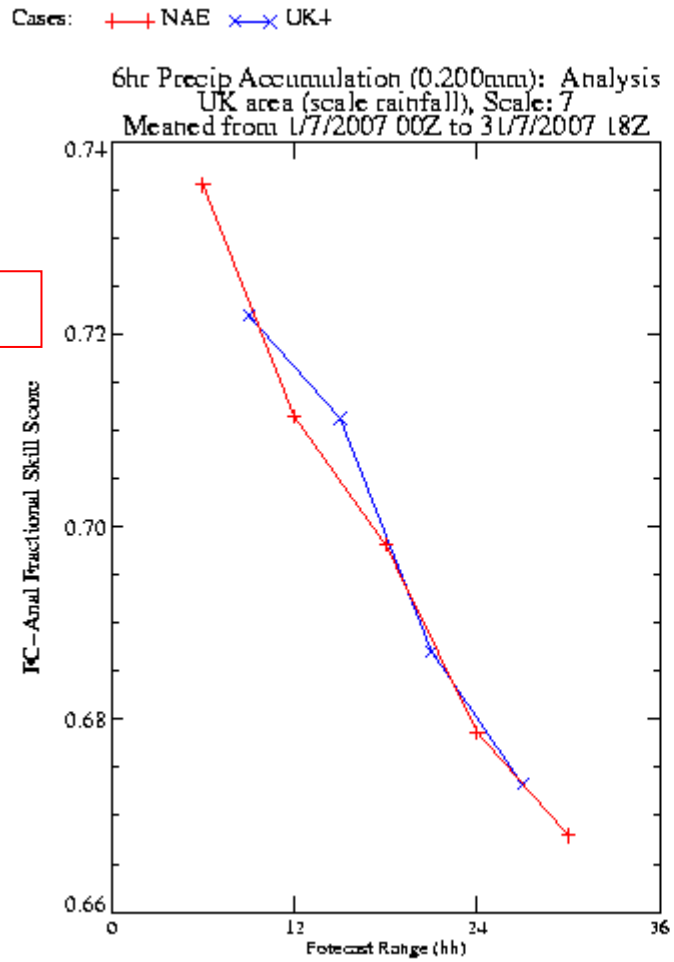


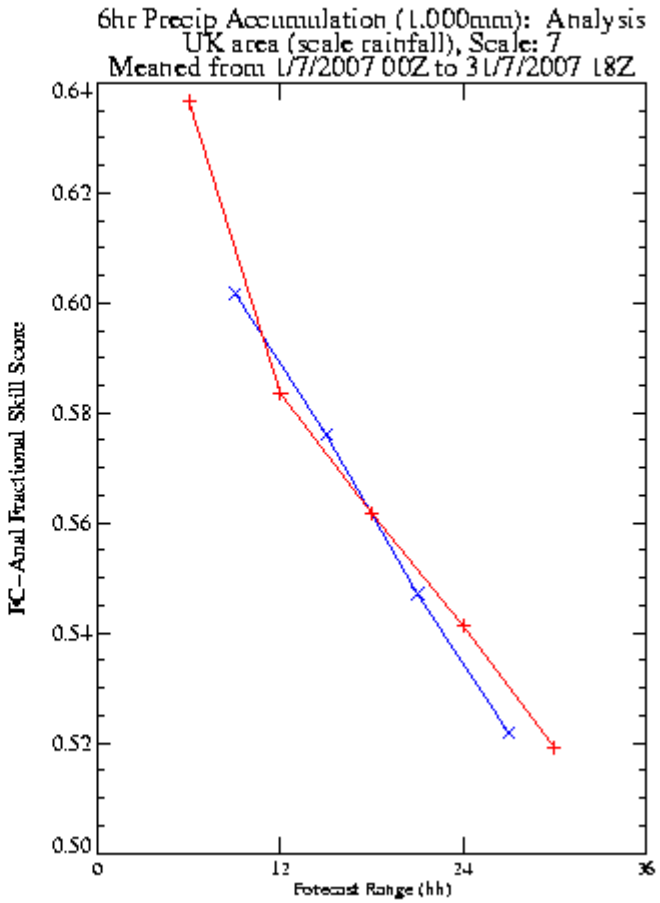


FSS –July 07 -0.2 & 1.0mm thresholds -scale=7*5km

12km

4km





Conclusions

- Standard verification contradictory as resolution increases
  - Nearest grid point – “double penalty effect”
  - Unrepresentative
- Gridded gauges and radar composites can differ by 50-100%
- Daily precipitation compared to dense gauges or radar well forecast – better by 4km resolution
- Fuzzy scores – fractional skill score can identify scales at which forecasts show useful skill
- More work planned to show the effects of uncertainty in observations on verification

Questions and answers