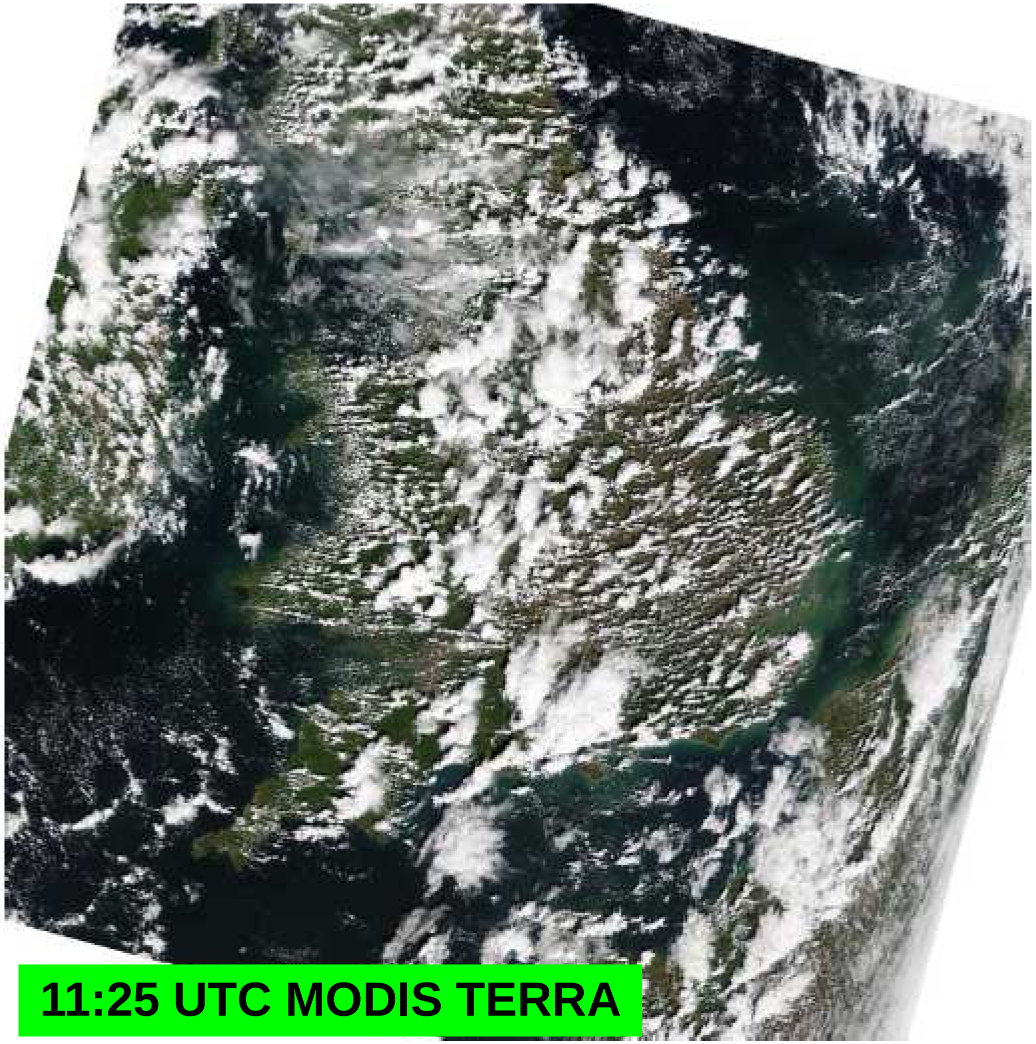
Observations: an inconvenient truth?

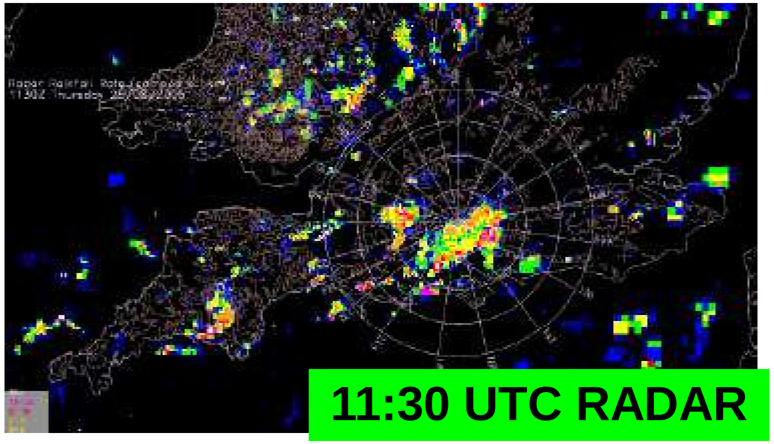
Marion Mittermaier

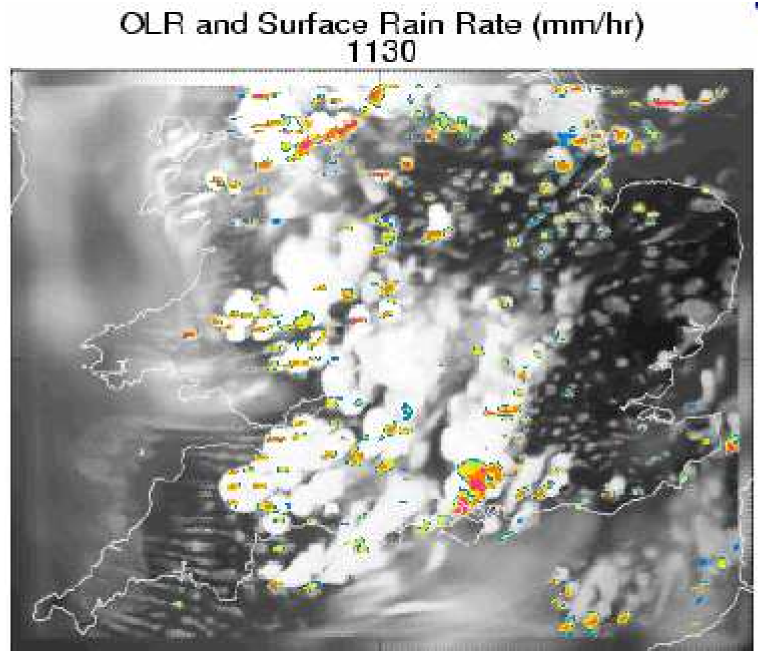
Outline

1. **Spatial** characteristics impact
2. **Observation** characteristics impact
3. **Multiple “truths”** → which way to jump?
4. Unravelling the signal from the noise → can it be done?
5. Conclusions

Cloud and rain, all the same?





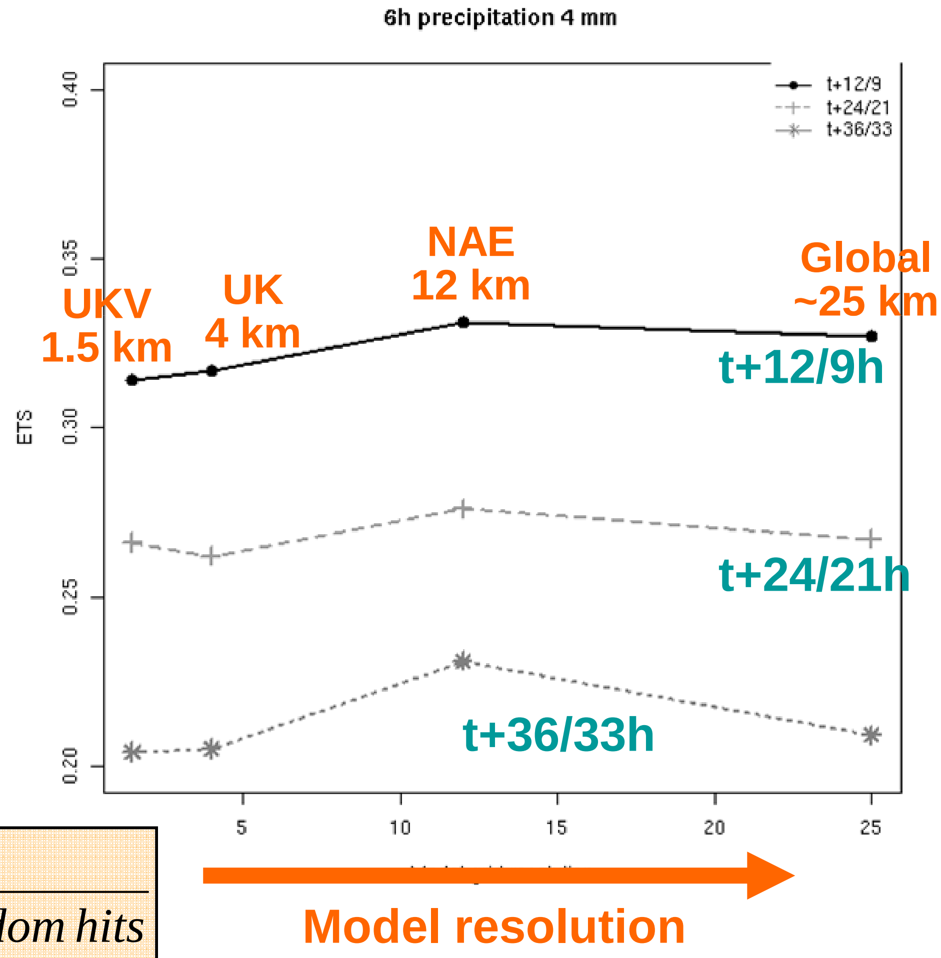


Courtesy of Sue Ballard

Does higher resolution give more skilful forecasts?

Apparently not!   Has it all been a waste of time?

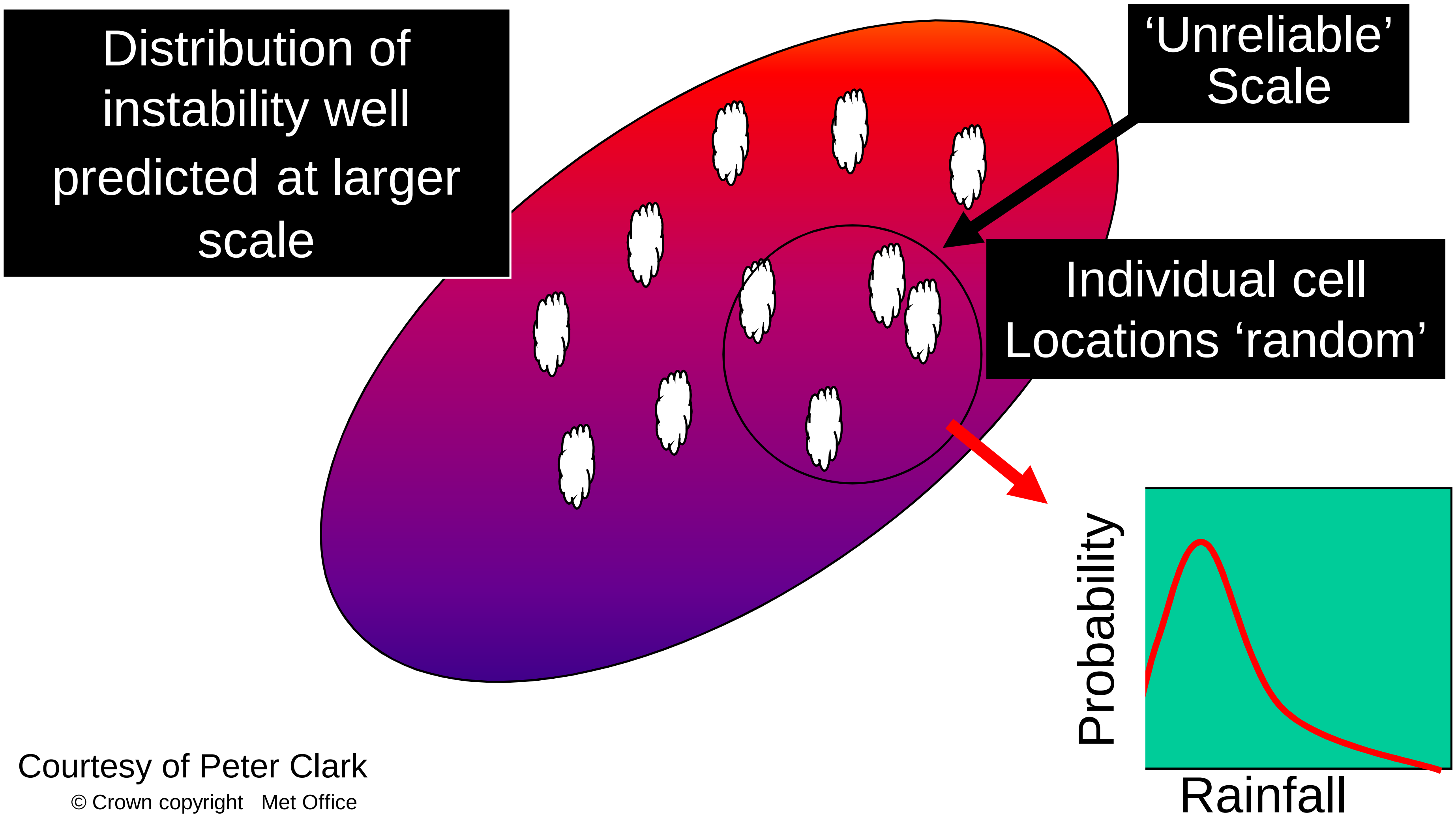
- April to Oct 2010
- Equitable Threat Score (ETS)
- Using Block 03 gauges



$$ETS = \frac{hits - random\ hits}{hits + false\ alarms + misses - random\ hits}$$

M Mittermaier et al. 2012 Met Apps

We shouldn't believe high-resolution (at or near the grid scale)



Courtesy of Peter Clark

Role of observations

- **Essential for verification**, but need to be treated with respect.
- **QC is important**!
- **Forecasts need to be well posed** to facilitate matching with observations.
- **Observations need to be appropriate** to capture the events of interest.
- **Observational uncertainty should be taken into account** in whatever way possible.

Error/uncertainty sources

- Biases in frequency or value
- Instrument error
- Random error or noise
- Reporting errors
- Reporting of errors
- Subjective obs (e.g., STORM data)
- Representativeness error
- Precision error
- Conversion error
- Analysis error
- ***Forecast error***

Radar vs gauge?

Rain gauges

- **Relatively precise and stable**
- Sparse network – not sufficient spatial information
- Point measurement - not a grid box average
- Occasional QC issues: e.g. snow melt
- Accumulation periods too long from many gauges

Radar

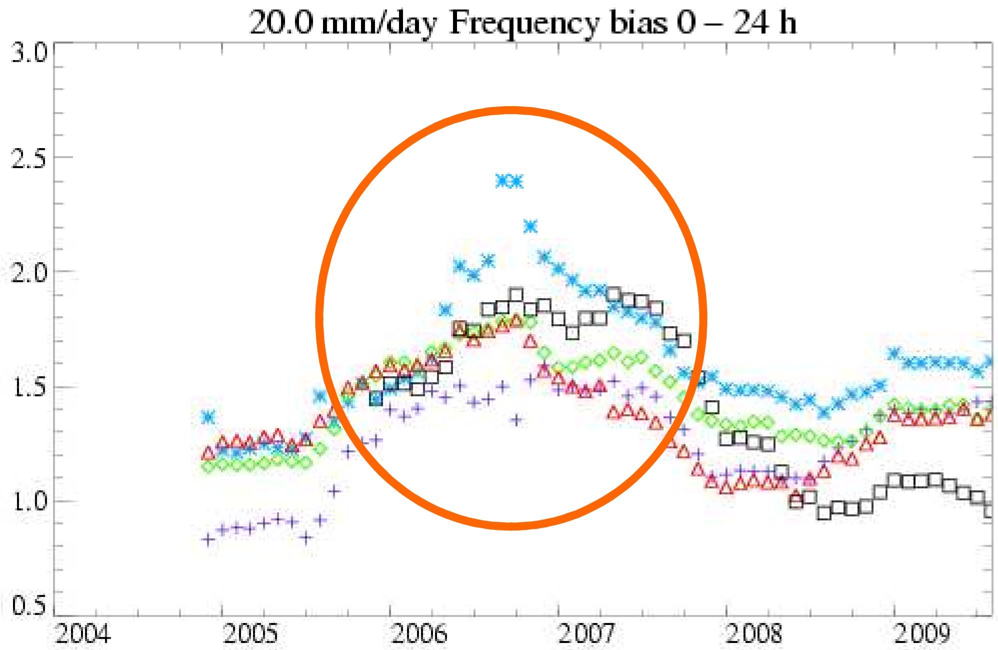
- **Good spatial coverage**
- **Grid square average**
- **Good temporal resolution**
- Assumptions in converting reflectivity to rain
- Clutter, anaprop – can be serious
- **Hardware and software upgraded; enhancements**
- Old network to be upgraded – not stable
- Attenuation in heavier rain
- Orographic enhancement

Nevertheless – if the forecasts looked like radar we'd be delighted

Courtesy of Nigel Roberts

Long-term forecast monitoring using radar-rainfall I

- The European Model Intercomparison of Precipitation (EMIP) showed **the power of using several models for monitoring the radar baseline**.

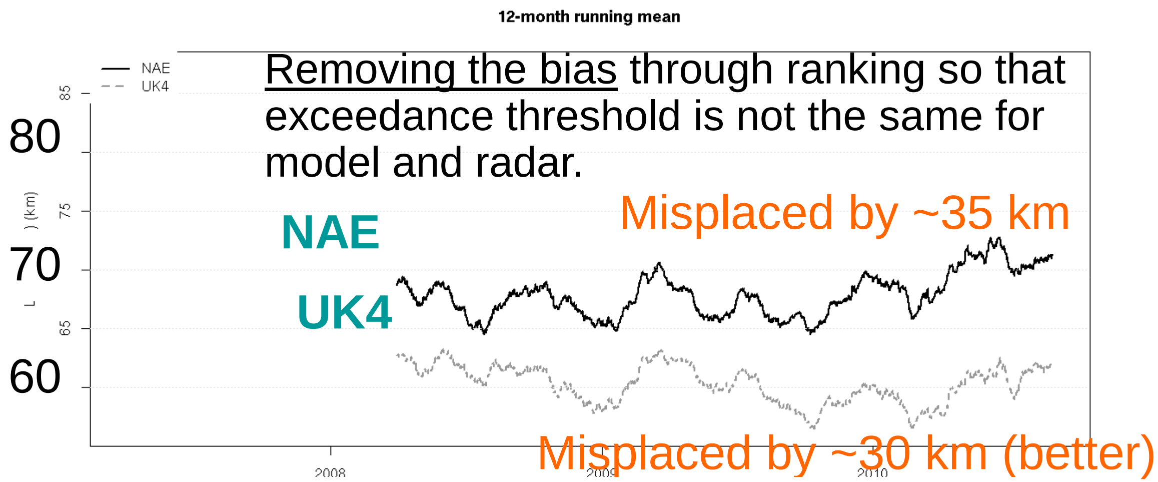


Traced to an issue of 5-min data used for hourly accumulations being deleted before the hour ended, so hourly accumulations only consisted of 45 min or 9 5-min slices.

Long-term forecast monitoring using radar-rainfall II

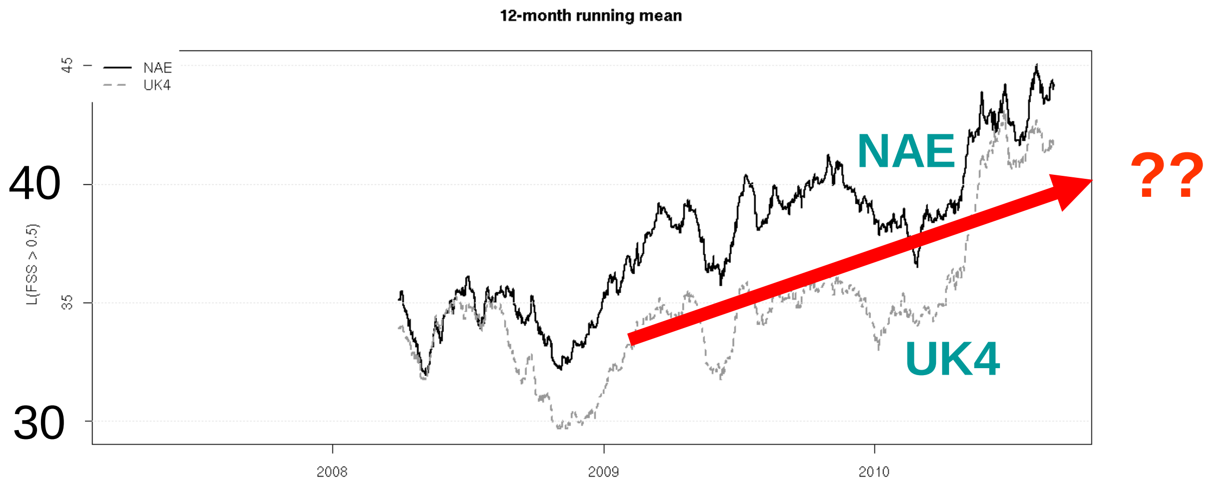
The expectation is that through model improvements (FSS>0.5) DECREASES over time….. or at least stays constant

10% threshold



Metric is impacted through the physical exceedance threshold applied at the grid scale.

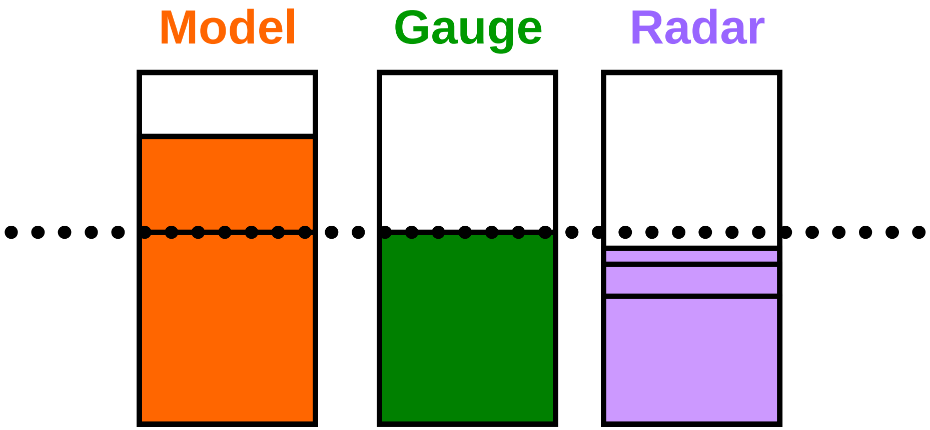
0.5 mm/6h



From Mittermaier *et al* 2012

What is going on here?

Counting the exceedances above a physical threshold e.g. 4 mm/h



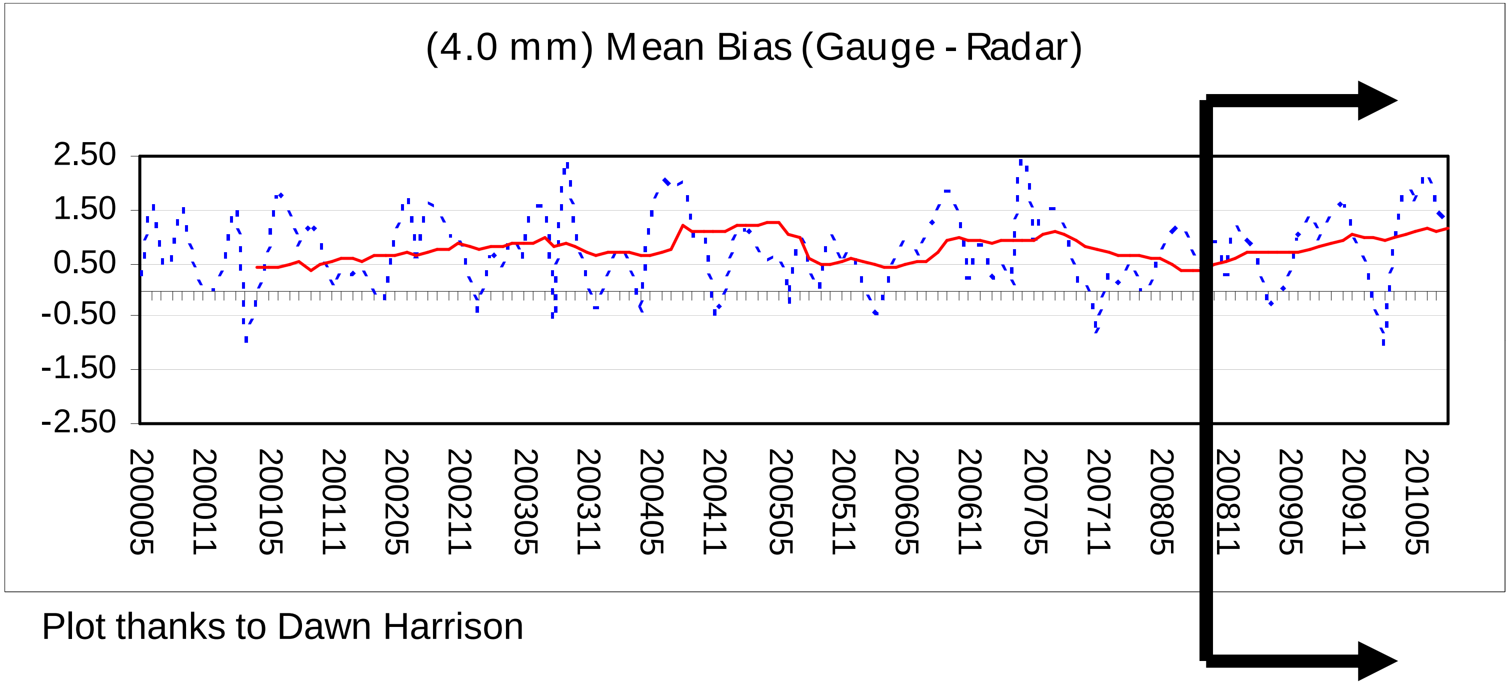
- **Model** configurations change several times a year, each associated with a possible impact on precipitation biases.

- **Gauges** are taken as relatively stable.

- **Radar** has a bias relative to gauges, and given the system complexity they are subject to fluctuations in output on much shorter time scales, and systematic trends.

- When using radar to verify model forecasts these two biases are superimposed and interact.

Gauge-radar bias against calibrating gauges



Plot thanks to Dawn Harrison

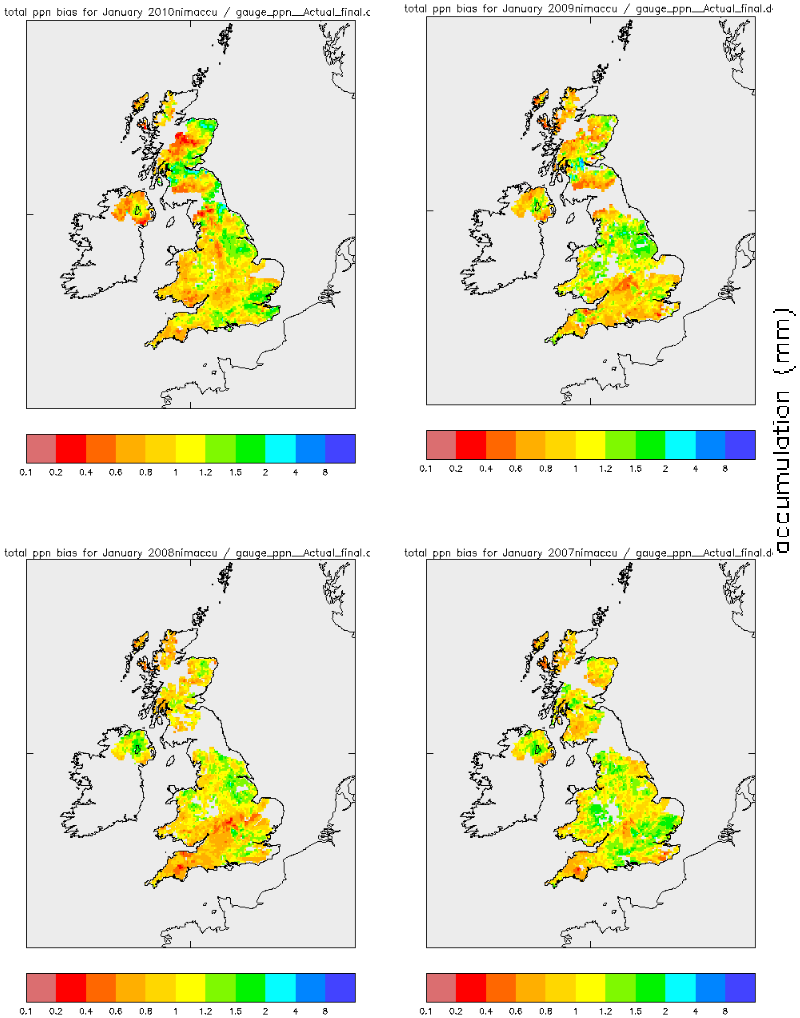
Caveats:
- Calibrating gauges not representative.
- Some radars have none in domain!

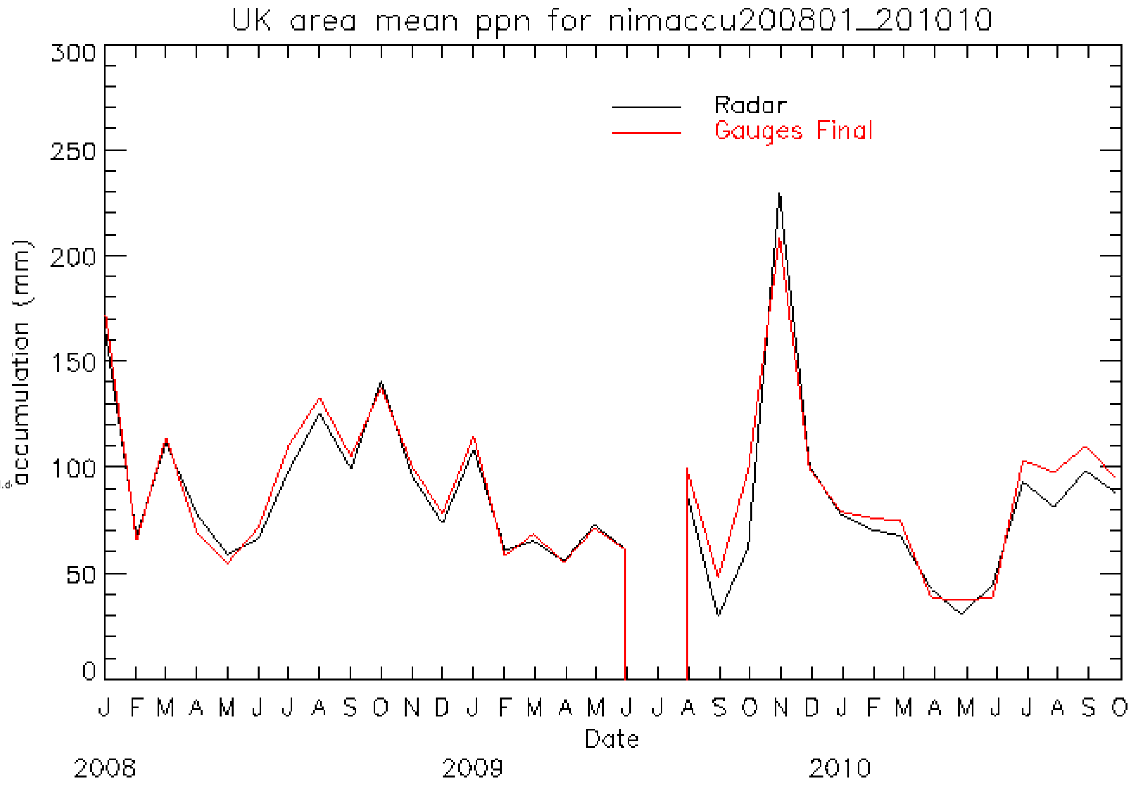
- A gradual increase in the bias towards **greater under-estimation by radar** means that fewer events breach a physical exceedance threshold, introducing a bias through the observations into the model frequency bias and scores.

Monthly maps and time series

CAVEAT: not equally matched.
Bias highly variable in space.

Bias Radar/Gauge January



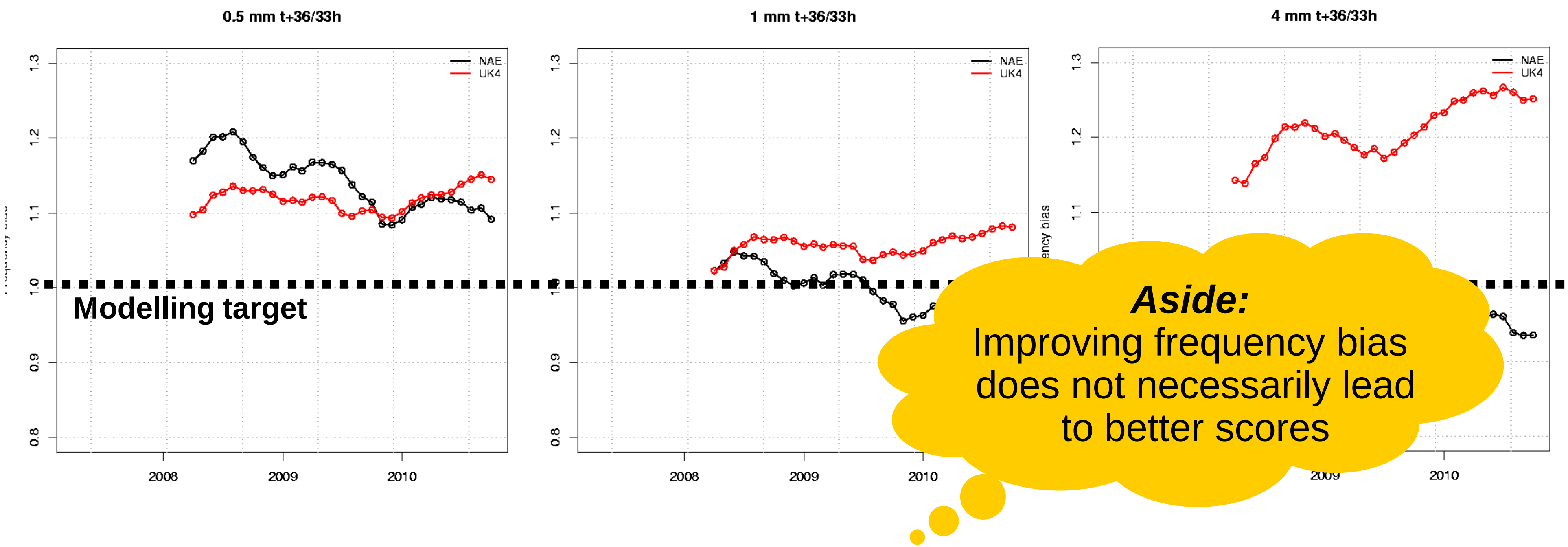


Radar more likely to be "under".

All plots Clive Wilson

Model bias against gauges

12-month means

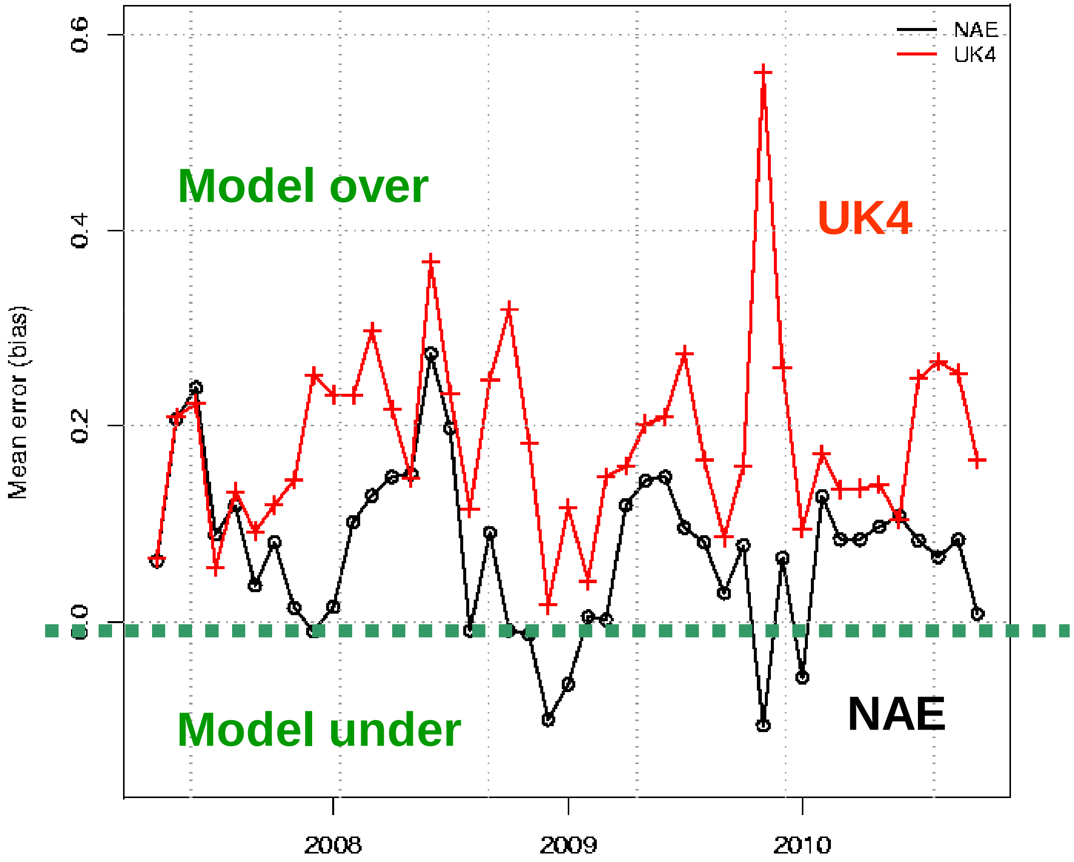


- Gradual improvement in NAE bias.
- Under-estimation of NAE for larger thresholds (expected)
- Over-estimation of UK4 at larger thresholds (expected). Worsening trend possibly not expected?

Model bias against gauges 2

(calculated more like the gauge-radar bias)

- Monthly ME values
- Not conditional (so slightly different to radar-gauge metric)
- In millimetres



Key findings

- No observations source is perfect, or complete.

- **The power in model inter-comparisons stems from spotting similar trends** that point to a characteristic of the baseline. One does not expect them to behave in exactly the same way as they are not at the same resolution.

- Despite the use of frequency thresholds **the lack of stability of a radar baseline could jeopardise the use of radar for long-term monitoring** for precipitation forecast skill, except in a comparative sense.

- The way **observation type (characteristics) affect verification statistics** poses a dilemma when it comes to interpretation of results: which way should I be tuning my forecasts? What is more right? Care is needed.

- **Disentangling systematic model behaviour from mixed observations signals** after-the-fact is virtually impossible. Be aware and understand before you start!

Thanks for listening!

Mittermaier M., N. Roberts and S. A. Thompson, 2012: A long-term assessment of precipitation forecast skill using the Fractions Skill Score. *Meteorol. Apps*., DOI= 10.1002/met.296