Current verification practices with a particular focus on dust

Marion Mittermaier and Ric Crocker

Outline

1. **Guide** to developing verification studies
2. **Observations** → at the root of it all
3. Grid-to-point, point-to-grid, grid-to-grid? **Traditional verification** and the **double penalty effect**
4. When and why do we need **new spatial methods**?
5. The appeal of evaluating synoptic evolution → **object-based methods** come into their own
6. Dust!

Simple guide for developing verification studies

1. Consider the **users**…
   - … of the forecasts
   - … of the verification information
2. What **aspects of forecast quality are of interest** for the user?
3. **Develop verification questions to evaluate those** aspects/attributes

Simple guide for developing verification studies 2

4. **Identify observations that represent the *event*** being forecast, including the
    - Element (e.g., temperature, precipitation)
    - Temporal resolution
    - Spatial resolution and representation
    - Thresholds, categories, etc.
5. **Identify multiple *verification attributes* that can provide answers** to the questions of interest
6. **Select *measures and graphics*** that appropriately measure and represent the attributes of interest
7. **Identify a *standard of comparison*** that provides a reference level of skill (e.g., persistence, climatology, old model)

Matching forecasts and observations

- **May be the *most difficult (and time consuming)* part of the verification process!**
- Many factors need to be taken into account, e.g.
  - Identifying observations that represent the forecast event

    Example: Precipitation accumulation over an hour at a point
  - For a gridded forecast there are many options for the matching process

    →Point-to-grid

    →Match obs to closest gridpoint

    →Grid-to-point

    →Interpolate?

    →Take largest value?

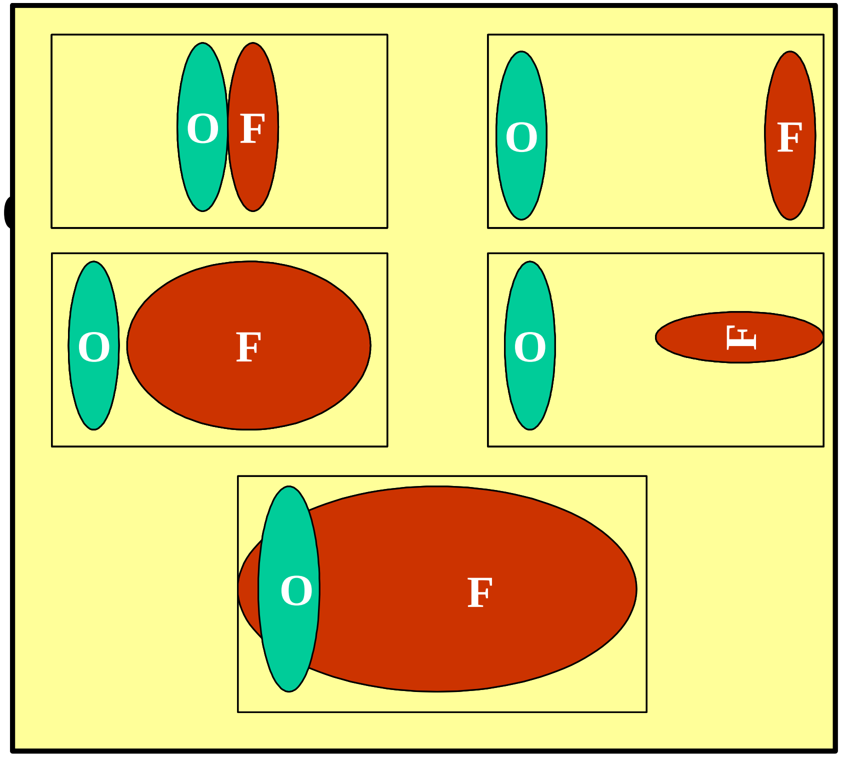
Role of observations

- **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**

The double penalty



Closeness not rewarded

Detail is penalised unless exactly correct

- higher resolution is more detailed!

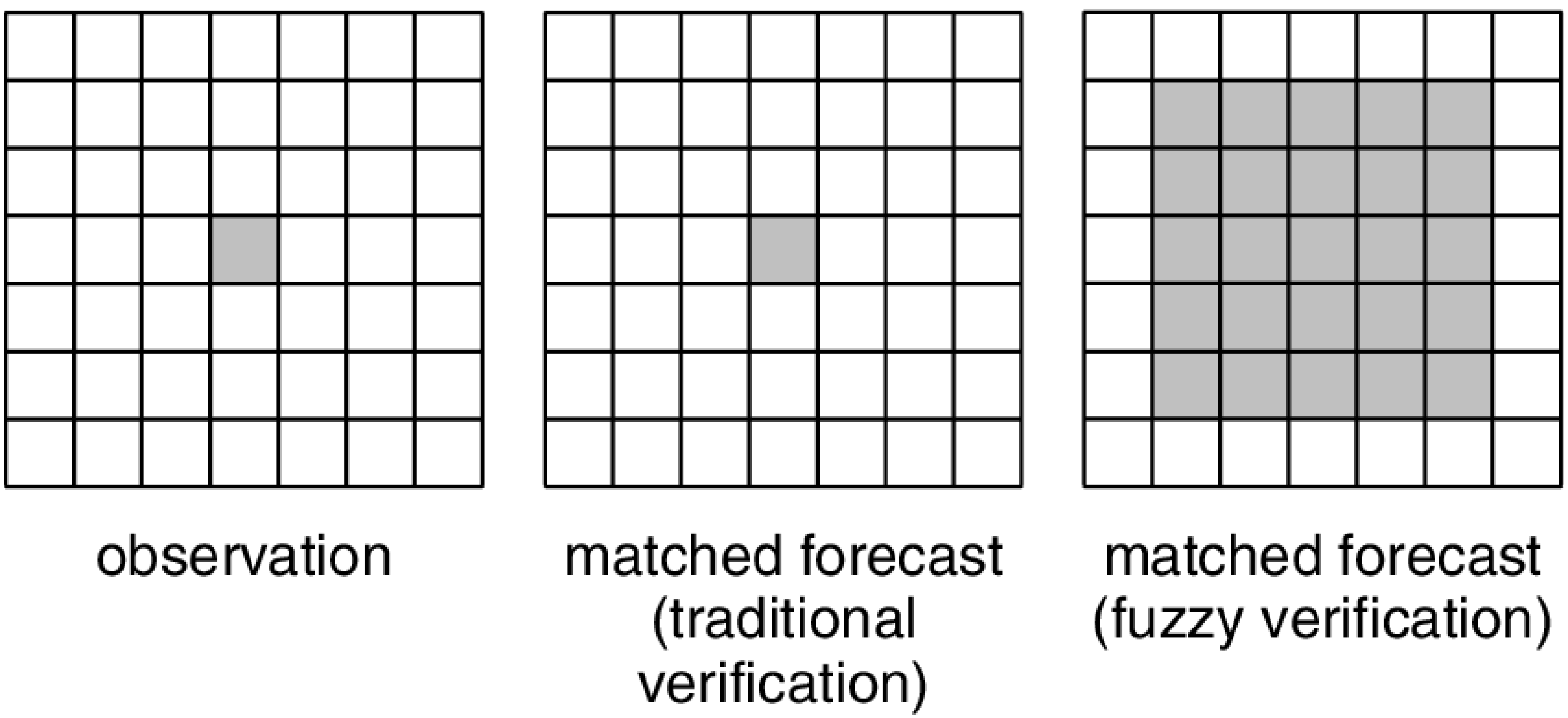
CSI = 0 for first 4;

CSI > 0 for the 5th

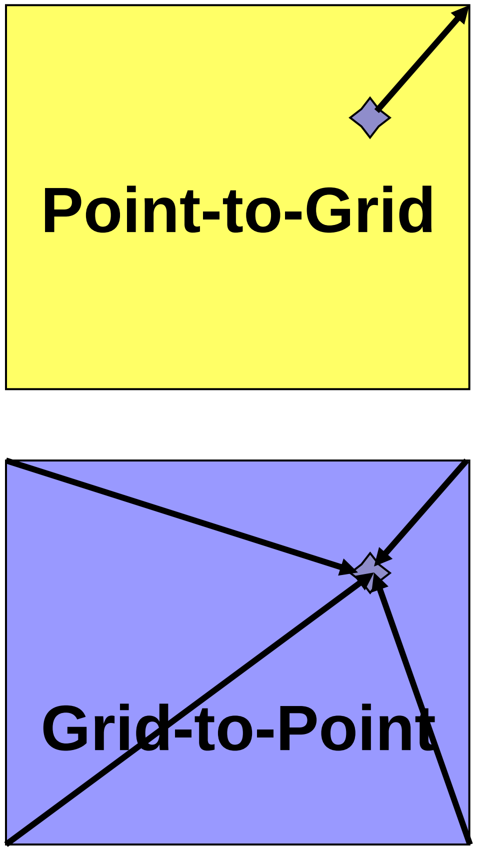
$$CSI = \frac{hits}{hits + false\,alarms + misses}$$

Traditional v neighbourhood verification

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



Matching forecasts and observations



Matching approach can impact verification results and interpretation

Impact of land-sea points?

On biases?

Matching forecasts and observations 2

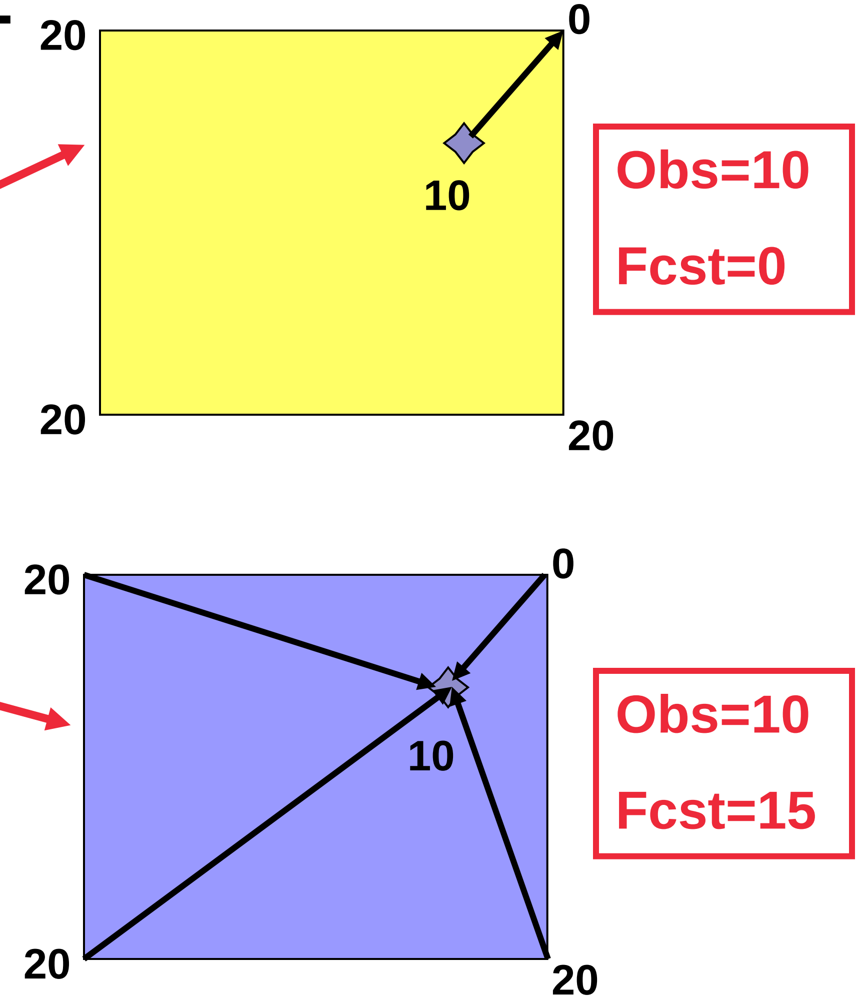
Example:

- Two approaches:
  - Match rain gauge to nearest gridpoint

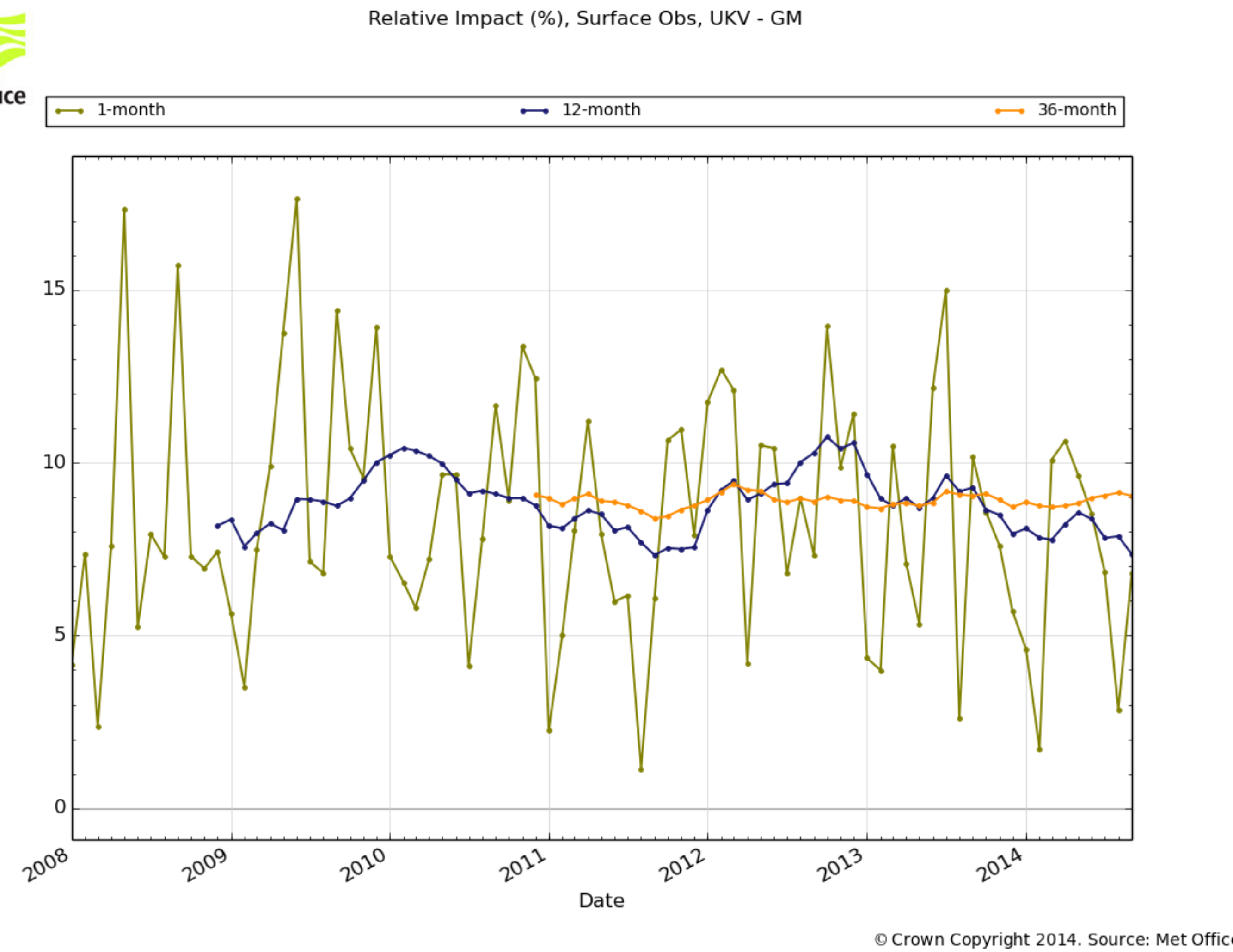
  or

  - Interpolate grid values to rain gauge location
    - *Crude assumption: equal weight to each gridpoint*
- Differences in results associated with matching:

"representativeness" error



Forecast accuracy: the relative index

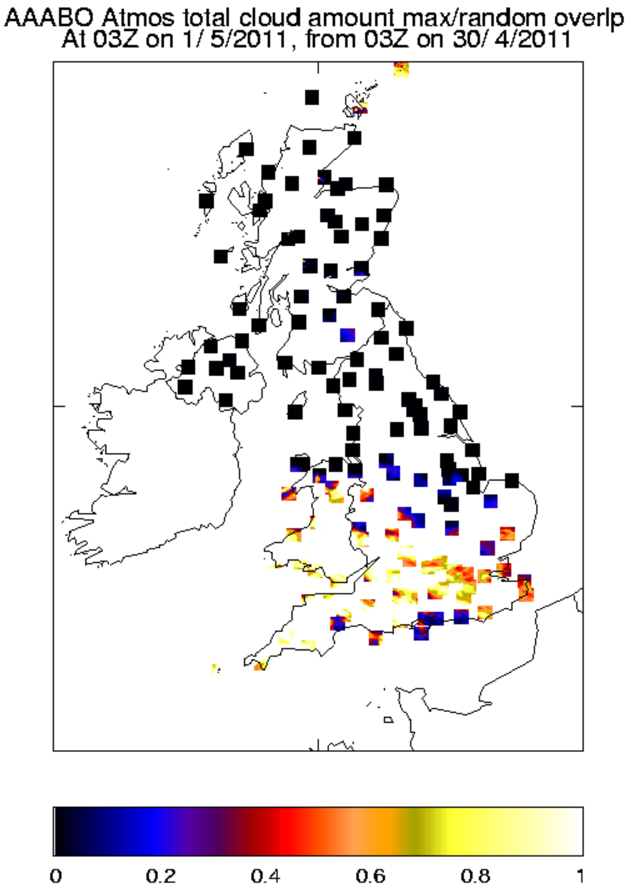


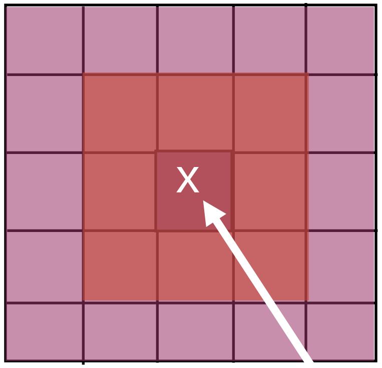
												ar-2014
Surface (1.5m) Temperature												7587
Vector Wind Difference												7591
Surface (1.5m) Visibility												5016
Total Cloud Amount/Cover												7526
Cloud Base Height (given 2.5 Okta												2921
6hr Precipitation Accumulation												3251
Overall impact												**7736**
												ar-2015
Surface (1.5m) Temperature												
Vector Wind Difference												
Surface (1.5m) Visibility	0.5731	0.5567	0.5508	0.5361	0.5465	0.3597						
Total Cloud Amount/Cover	0.8327	0.8861	0.9249	0.8685	0.8744	0.8547						
Cloud Base Height (given 2.5 Oktas Cloud Cover)	0.3535	0.4039	0.4093	0.4461	0.4841	0.5129						
6hr Precipitation Accumulation	0.3178	0.3159	0.2903	0.2658	0.2437	0.2400						
Overall impact	**8.2245**	**8.5624**	**8.3726**	**7.8213**	**7.8746**	**7.3555**						

Spatial sampling

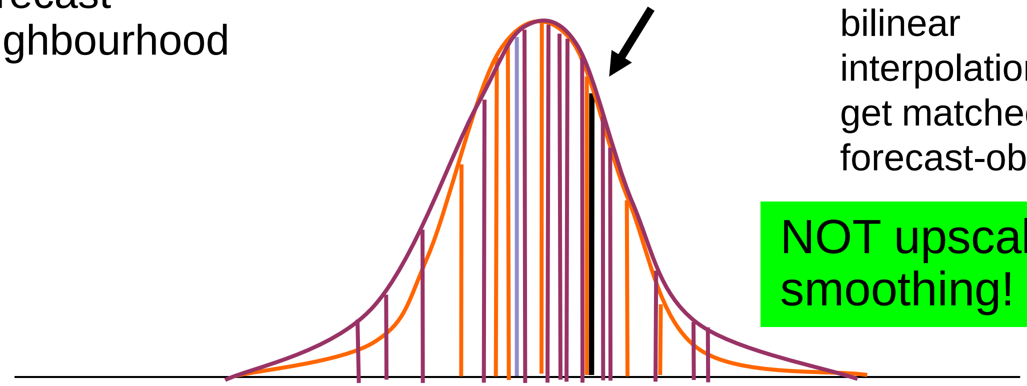
- Make use of spatial verification methods which compare **single observations** to a **forecast neighbourhood around the observation location**. → **SO-NF**

17 x 17





Forecast neighbourhood



- 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!

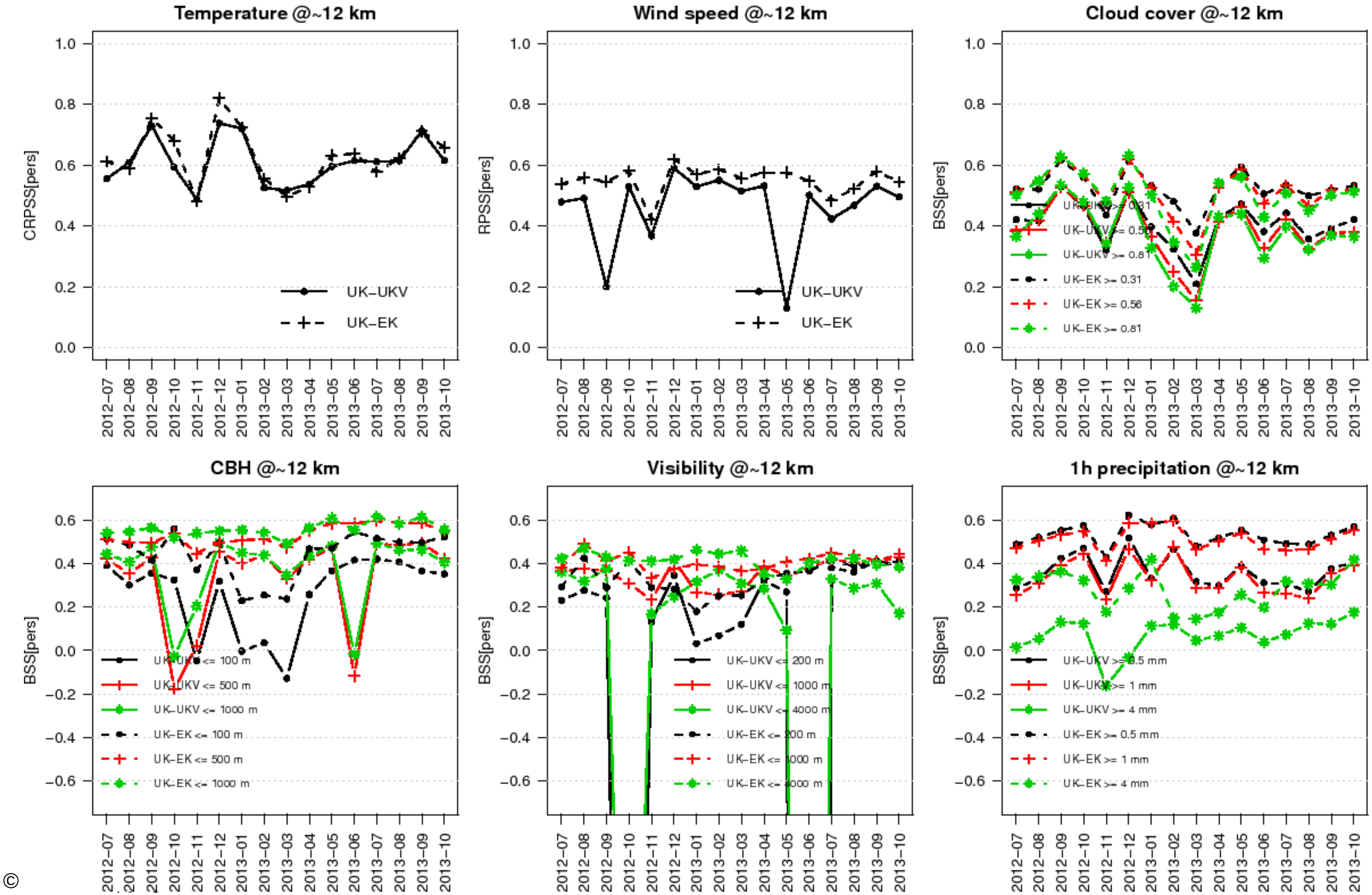
Only ~130 1.5 km grid points in >500 000 domain used to assess entire forecast! Note the variability in the neighbourhoods.

Time series – skill against persistence

MOGREPS-UK @ 2.2 km
UKV @ 1.5 km

Met Office

MOGREPS-UK is already outperforming UKV



Synoptic evolution: Feature-based assessment

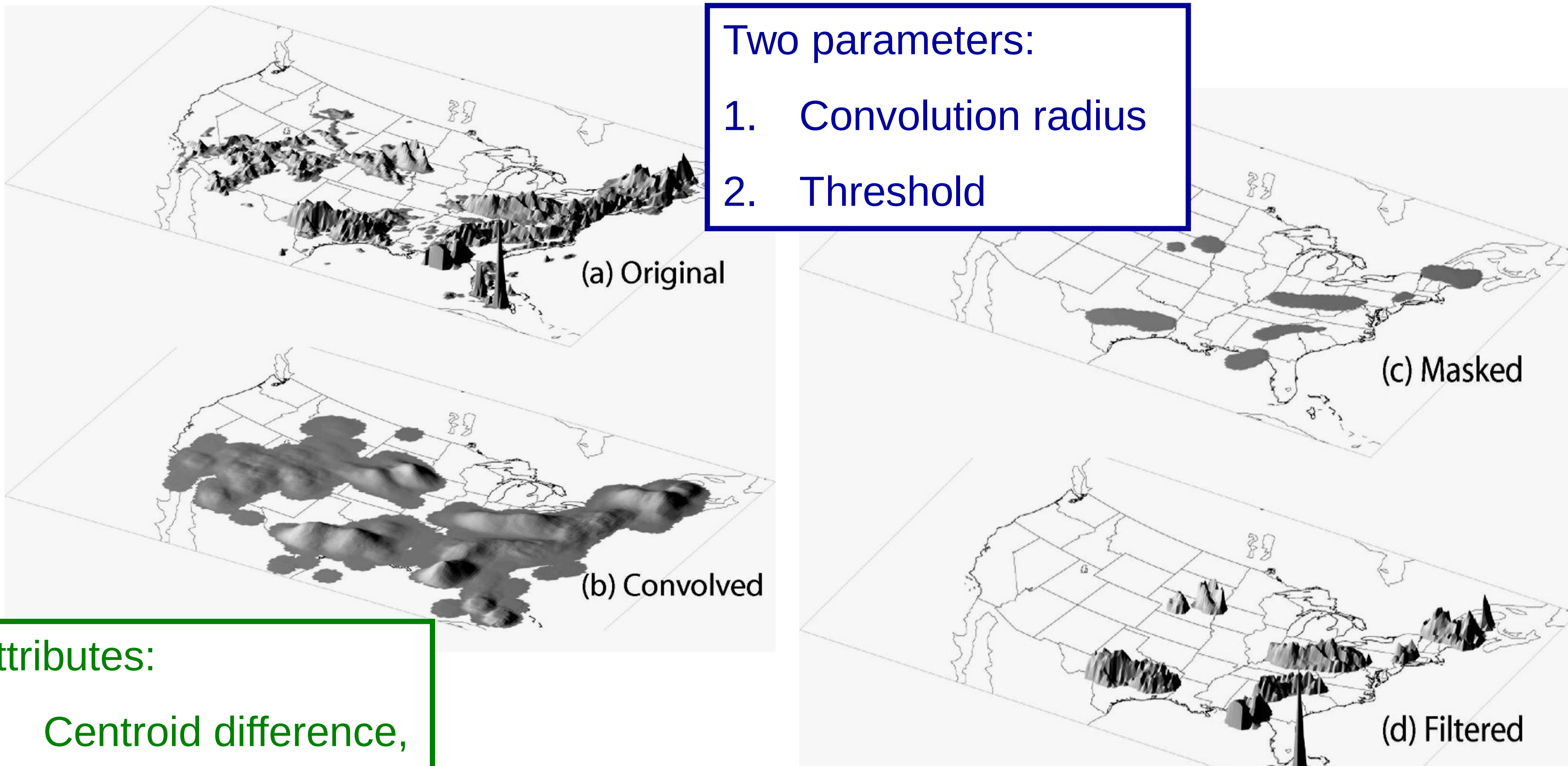
MODE – Method for Object-based Diagnostic Evaluation

Highly configurable

Davis et al., *MWR*, 2006

Two parameters:

1. Convolution radius
2. Threshold

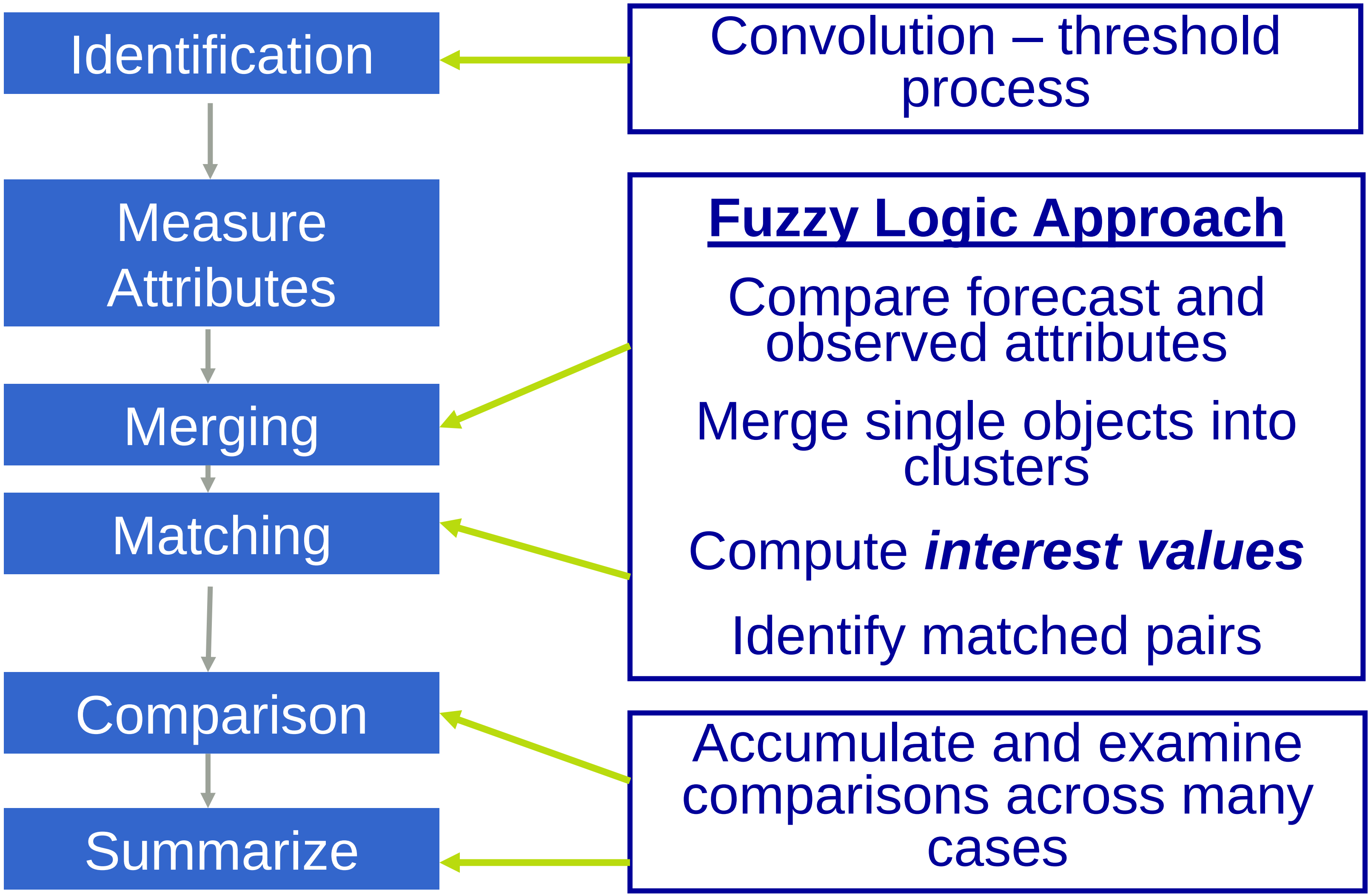


Attributes:

- Centroid difference,
- Angle difference,
- Area ratio etc

Focus is on *spatial* properties, especially the *spatial* biases

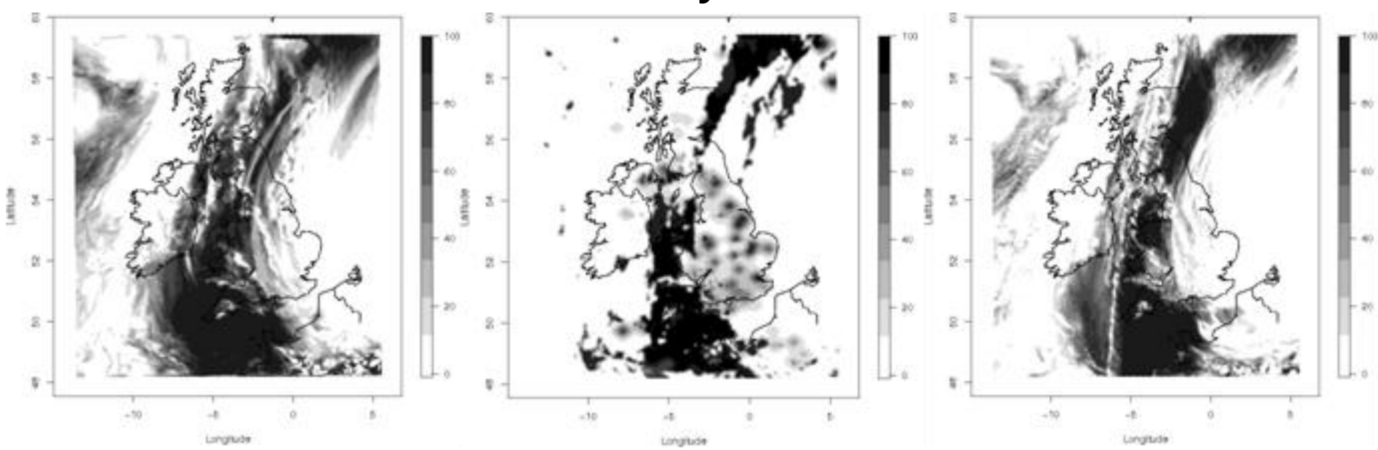
MODE methodology

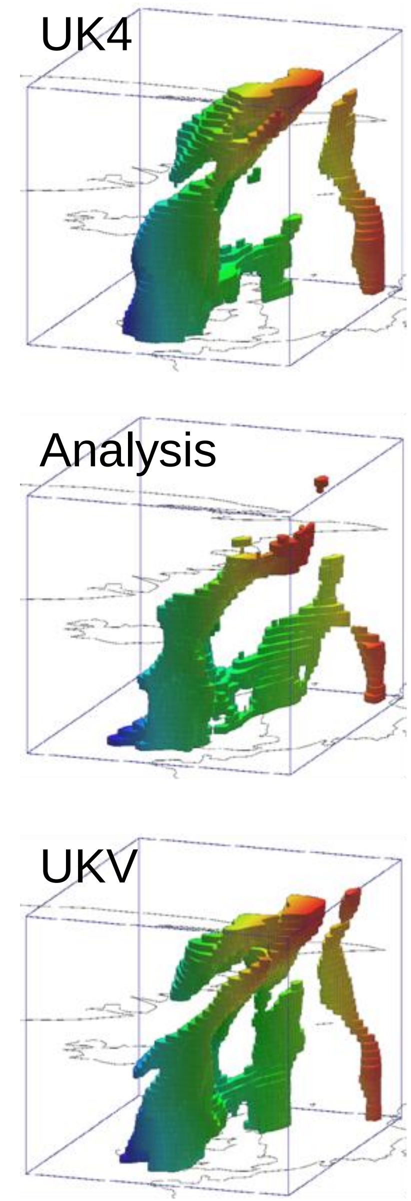


MODE example: cloud

- Tracking cloud free areas
- Using UKPP cloud analysis

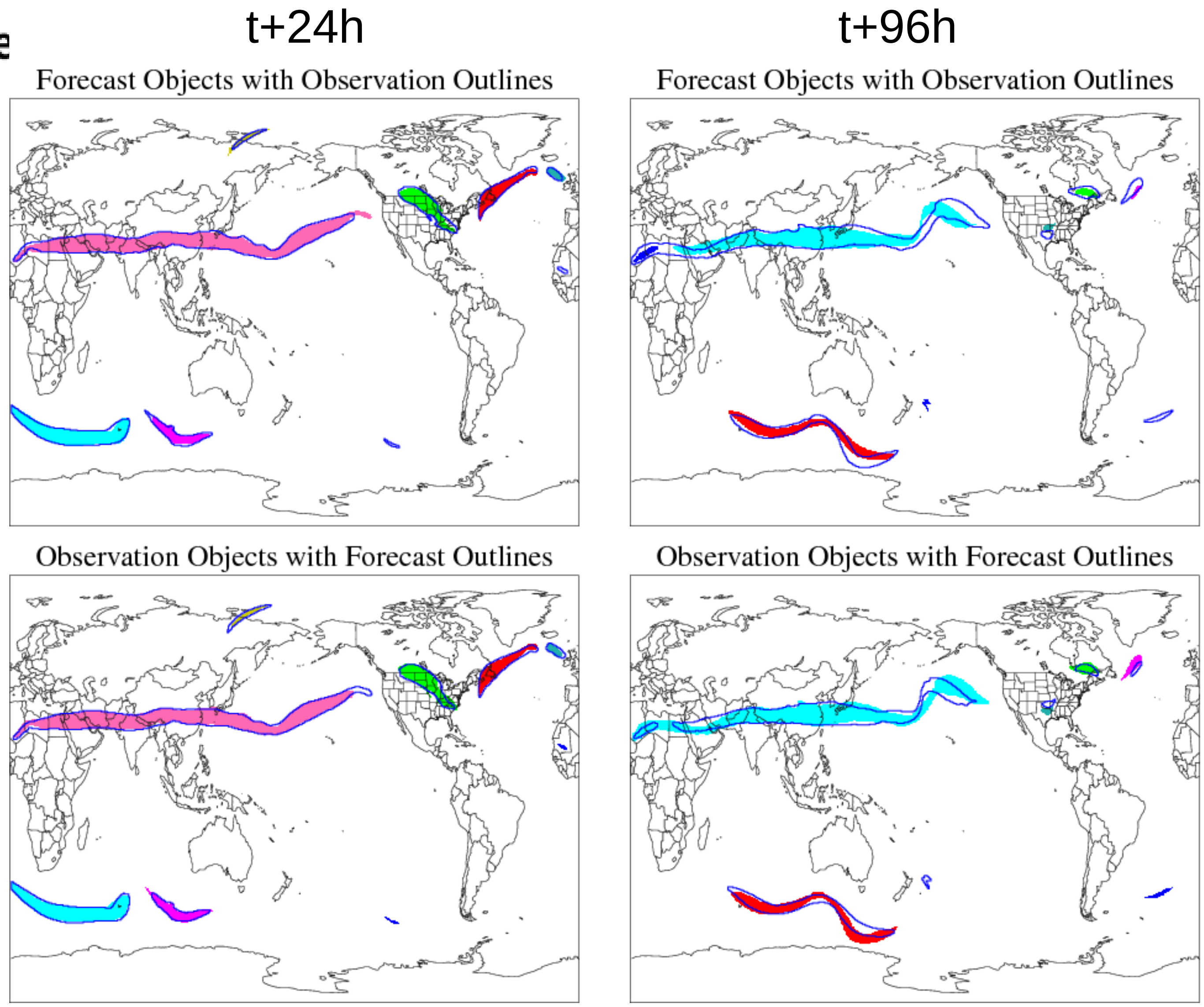
UK4 Analysis UKV





Mittermaier and Bullock, 2013)

MODE example: 250 hPa jets

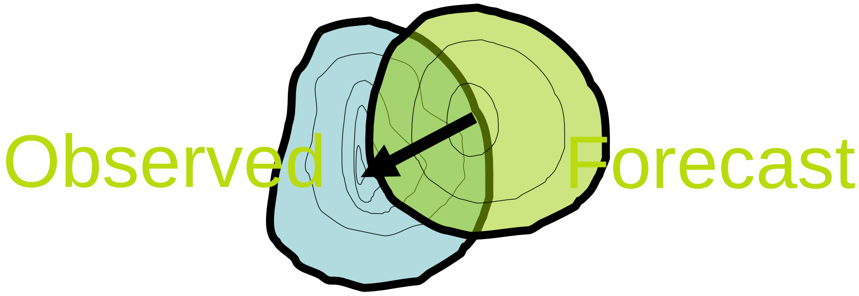


Evaluation of dust

Entity-based approach (CRA)

Ebert and McBride (2000)

- Define entities using threshold (Contiguous Rain Areas)
- Horizontally translate the forecast until a *pattern matching* criterion is met:
  - minimum total squared error between forecast and observations
  - maximum correlation
  - maximum overlap
- The displacement is the vector difference between the original and final locations of the forecast.



CRA details

- MSE*total* = MSE*disp* + MSE*vol* + MSE*pattern*

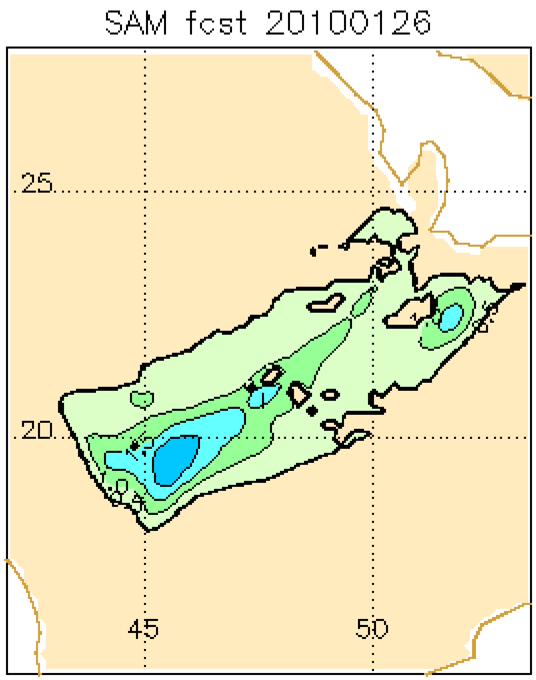
- The difference between the MSE before and after the feature has been displaced gives
  MSE*disp*= MSE*total* – MSE*shifted*

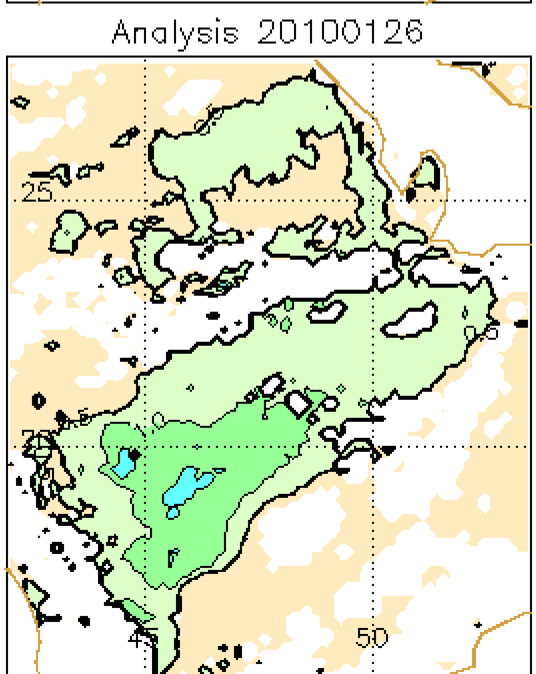
- The volume error represents the bias in mean intensity
  MSE*vol* = (F – X)^2 where F and X are the mean forecast and observed values after displacement

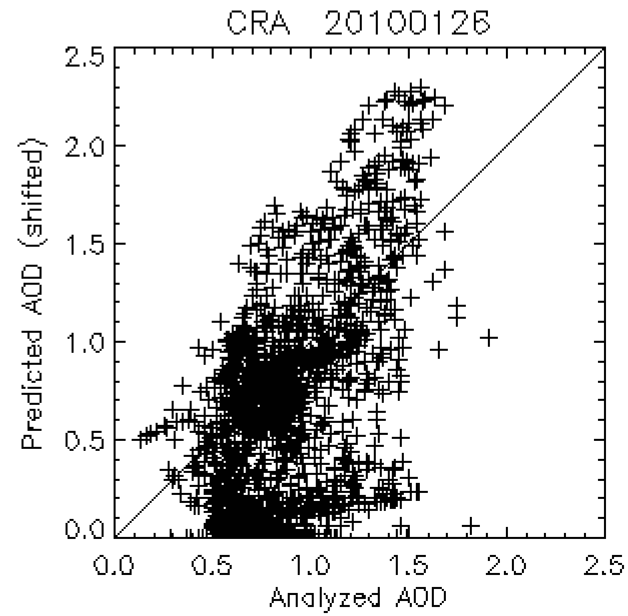
- The pattern error accounts for differences in the fine structure with MSE*pattern* = MSE*shifted* – MSE*vol*

Comparison details

- 12 km Unified model configuration called the “South Asia Model” or SAM.
- Compared to AERONET aerosol optical depth (AOD) and AOD product from Spinning Enhanced Visible and Infrared Imager (SEVIRI) instrument on MSG.
- SEVIRI viewing area only covers ~half of model domain.
-
- AOD can not be calculated for areas contaminated with cloud, or over the sea.
- Both SEVIRI and forecast put on 0.15deg grid.







SAM 12-12 fcst 20100126 n=1975
(15.50°,42.25°) to (27.65°,53.20°)
Verif. grid=0.150° CRA threshold=0.5 AOD

	Analysed	Forecast
# gridpoints ≥0 AOD	1918	1113
Average AOD ()	0.84	0.58
Maximum AOD ()	1.91	2.30

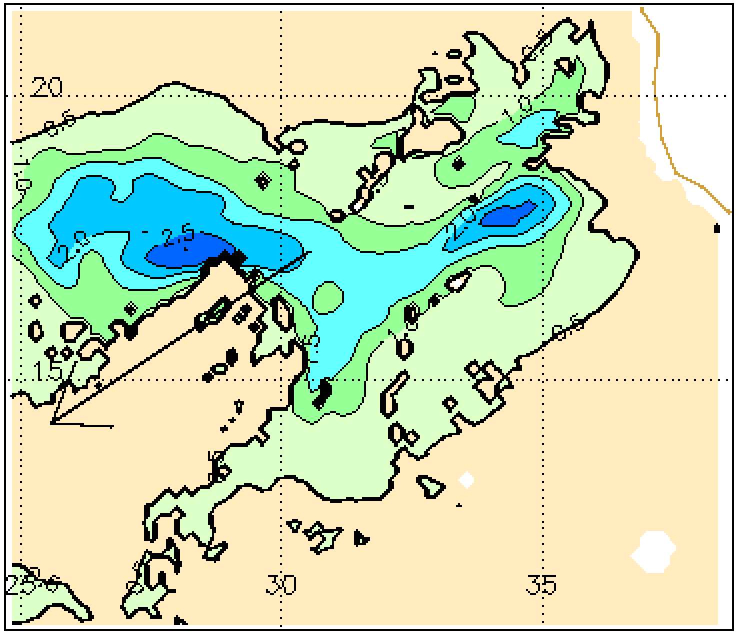
Displacement (E,N) = [0.00°,0.30°]

	Original	Shifted
RMS error ()	0.49	0.51
Correlation coefficient	0.516	0.575

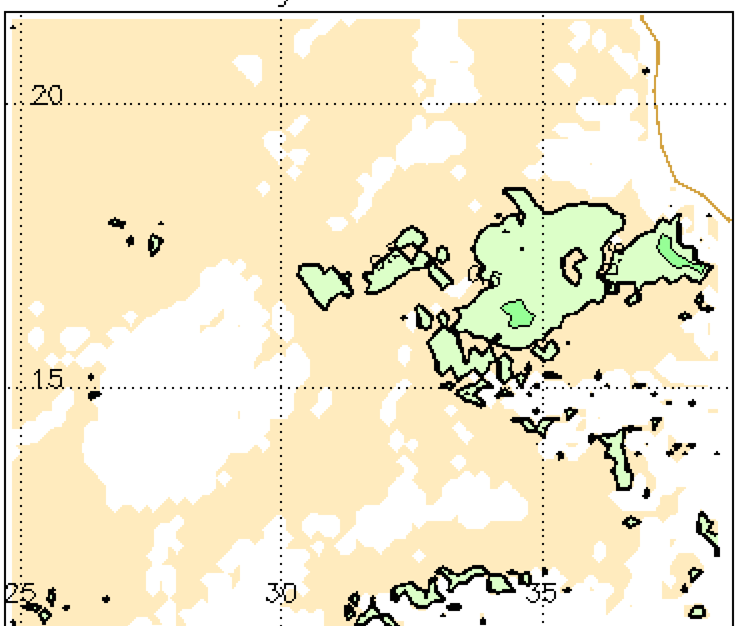
Error Decomposition:

Displacement error	6.9%
Volume error	20.4%
Pattern error	72.7%

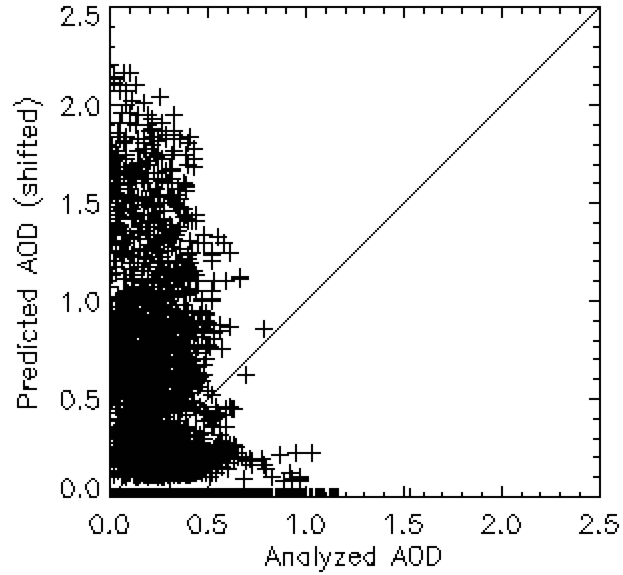
SAM fcst 20100218



Analysis 20100218



CRA 20100218



SAM 12-12 fcst 20100218 n=3147
(10.85°,25.00°) to (21.35°,38.35°)
Verif. grid=0.150° CRA threshold=0.5 AOD

	Analysed	Forecast
# gridpoints ≥0 AOD	441	2626
Average AOD ()	0.29	0.42
Maximum AOD ()	1.53	2.21

Displacement (E,N) = [4.95°,3.00°]

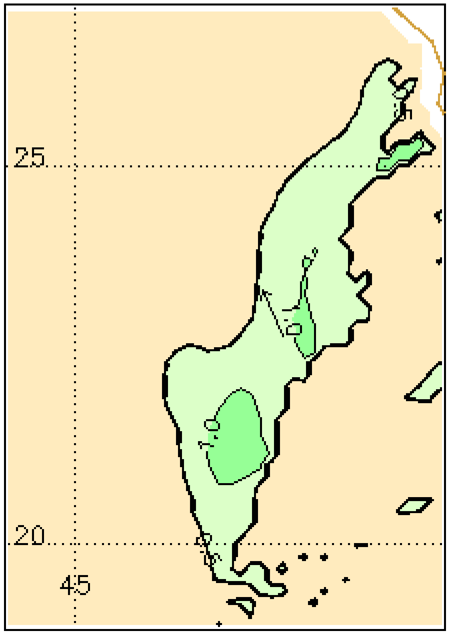
	Original	Shifted
RMS error ()	0.95	0.60
Correlation coefficient	0.075	−0.325

Displacement may be wrong – >25% of fcst removed

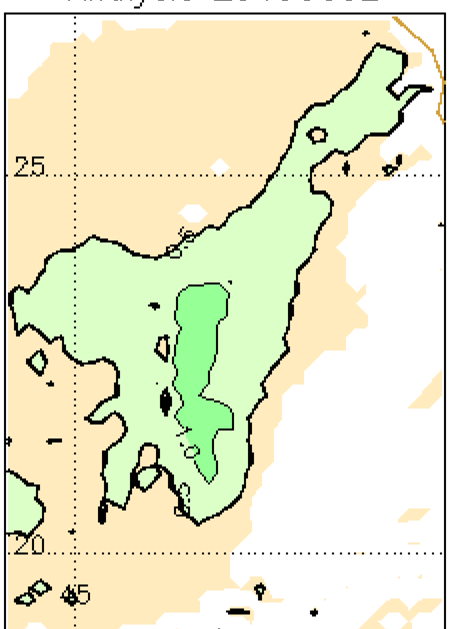
Error Decomposition:

Displacement error	60.8%
Volume error	1.7%
Pattern error	37.5%

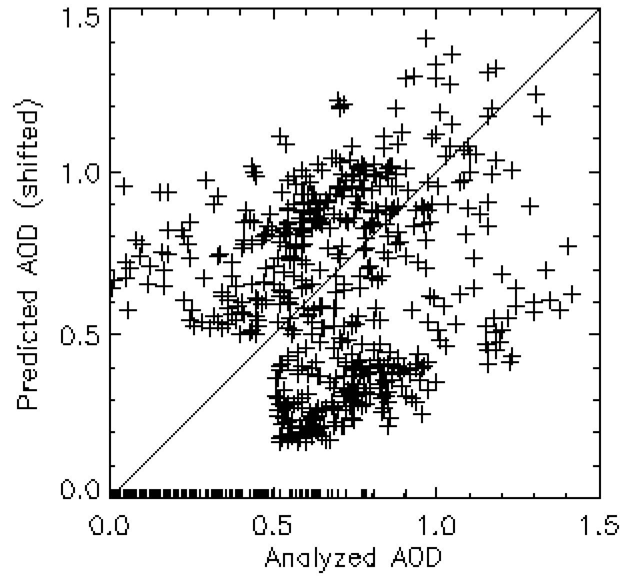
SAM fcst 20100302



Analysis 20100302



CRA 20100302



SAM 12-12 fcst 20100302 n=742
(19.10°,44.20°) to (26.90°,50.05°)
Verif. grid=0.150° CRA threshold=0.5 AOD

	Analysed	Forecast
# gridpoints ≥0 AOD	523	376
Average AOD ()	0.61	0.49
Maximum AOD ()	1.42	1.41

Displacement (E,N) = [0.30°,−0.60°]

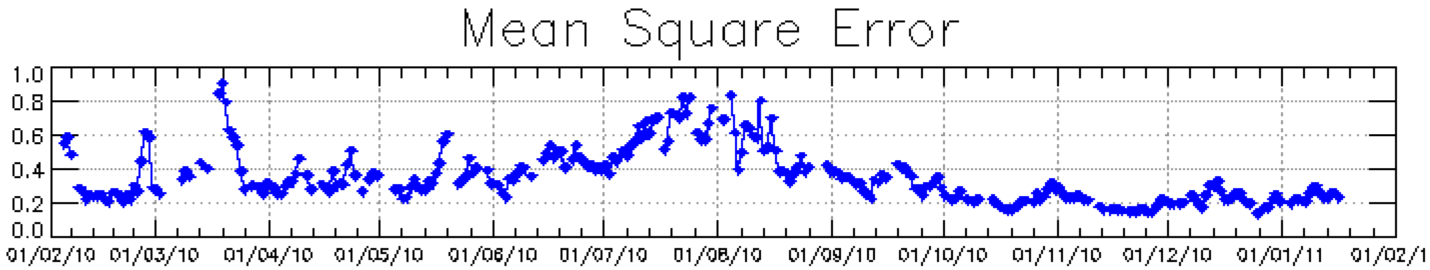
	Original	Shifted
RMS error ()	0.46	0.36
Correlation coefficient	−0.319	0.471

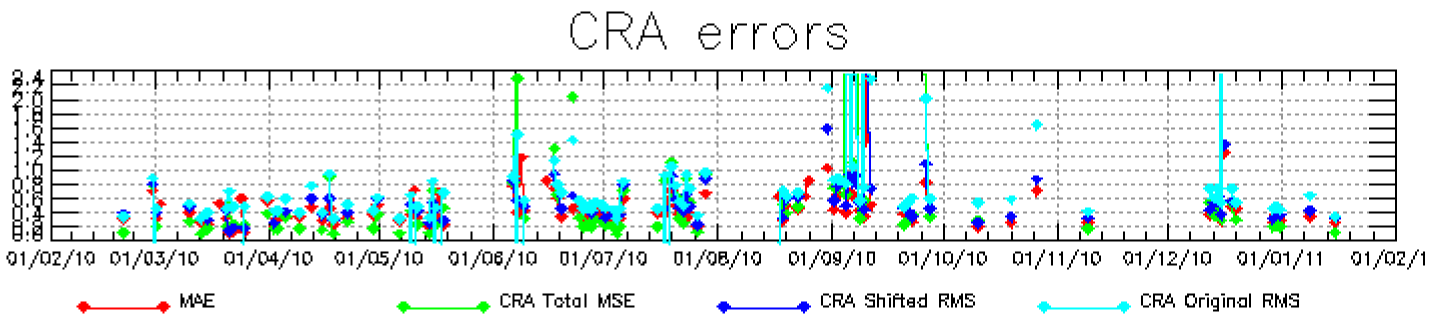
Displacement may be wrong − correlation not signif.

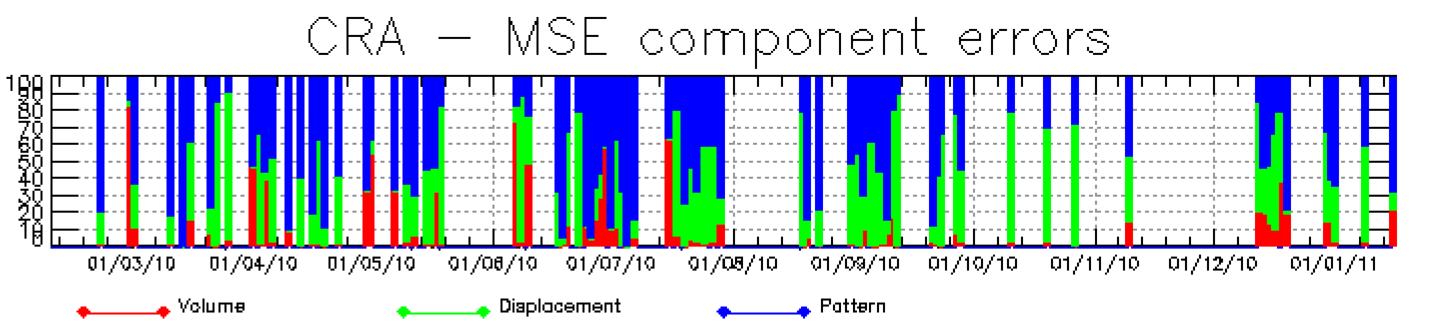
Error Decomposition:

Displacement error	37.0%
Volume error	6.9%
Pattern error	56.2%

CRA time series

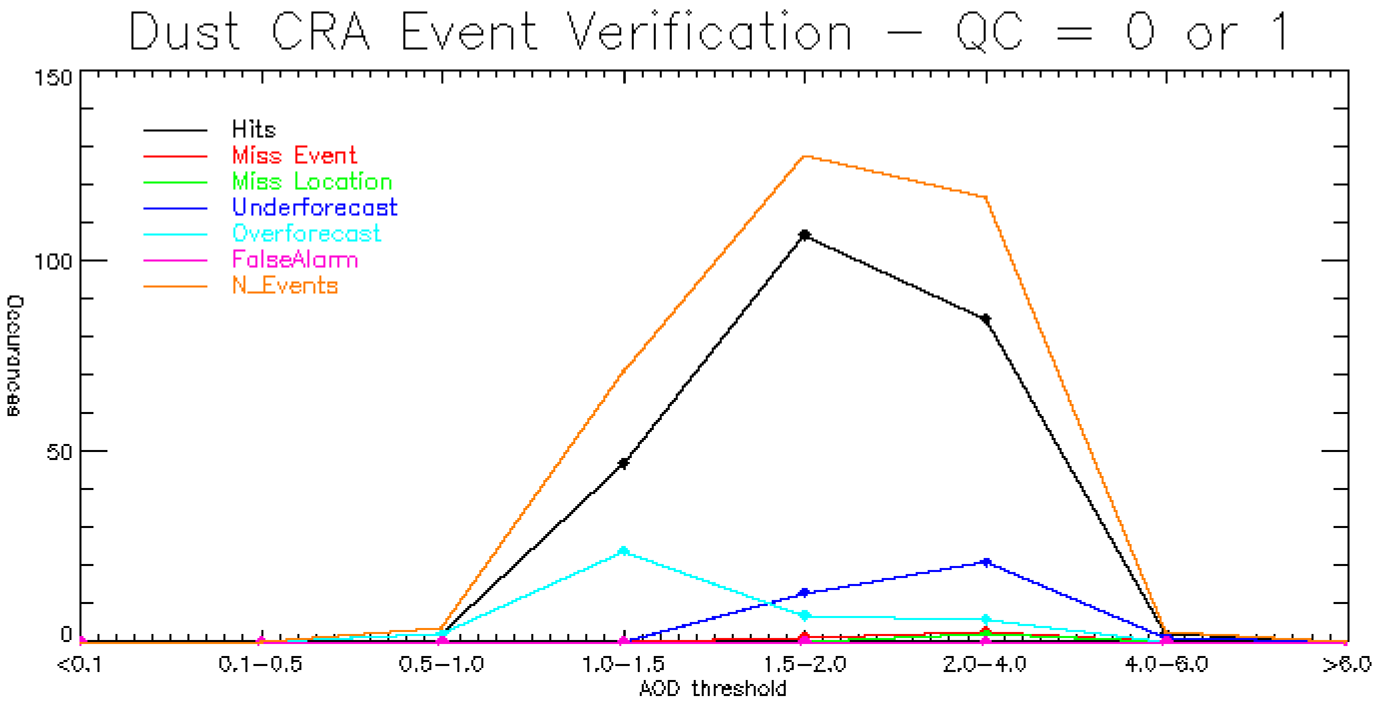




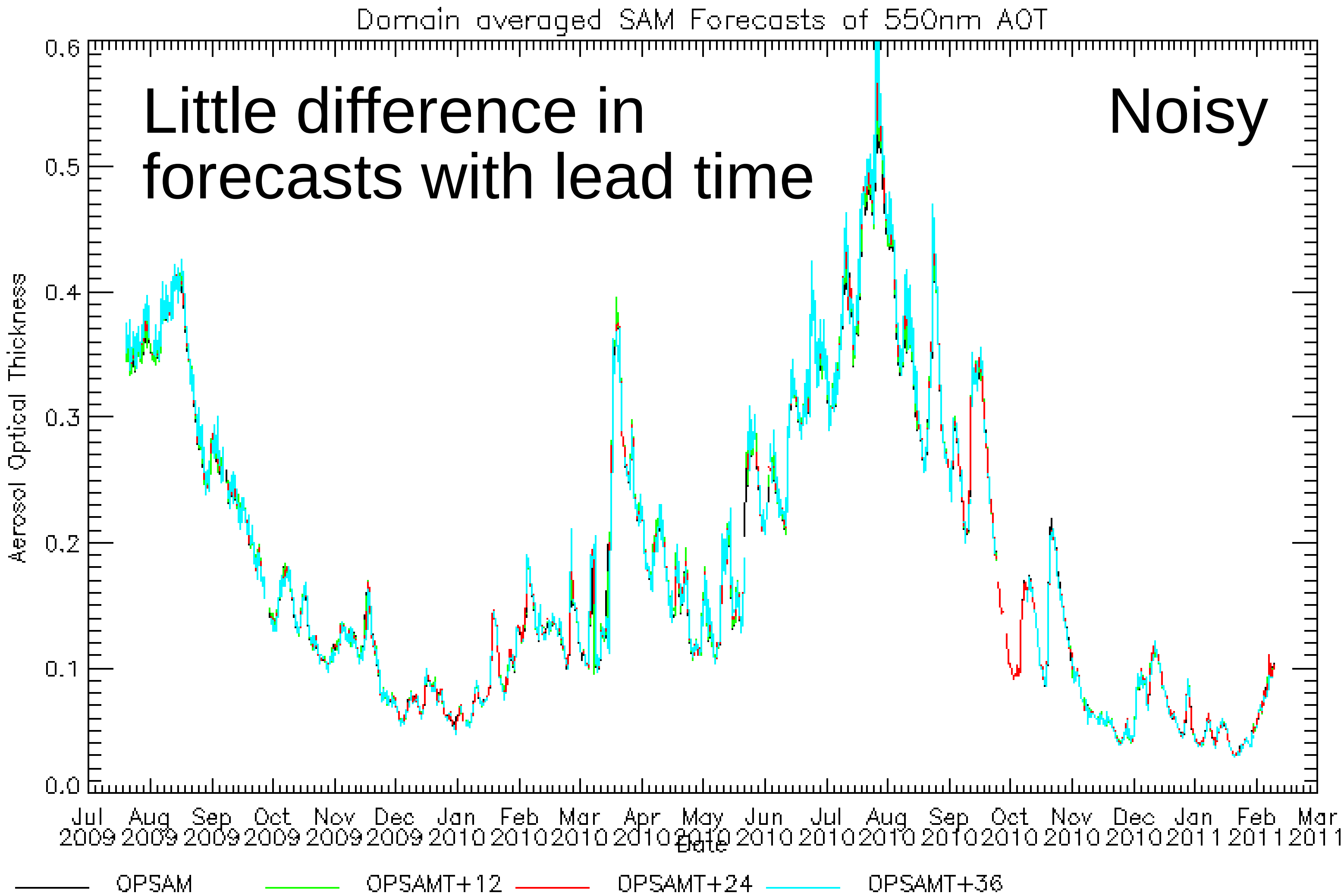


Contingency table for matching CRA objects

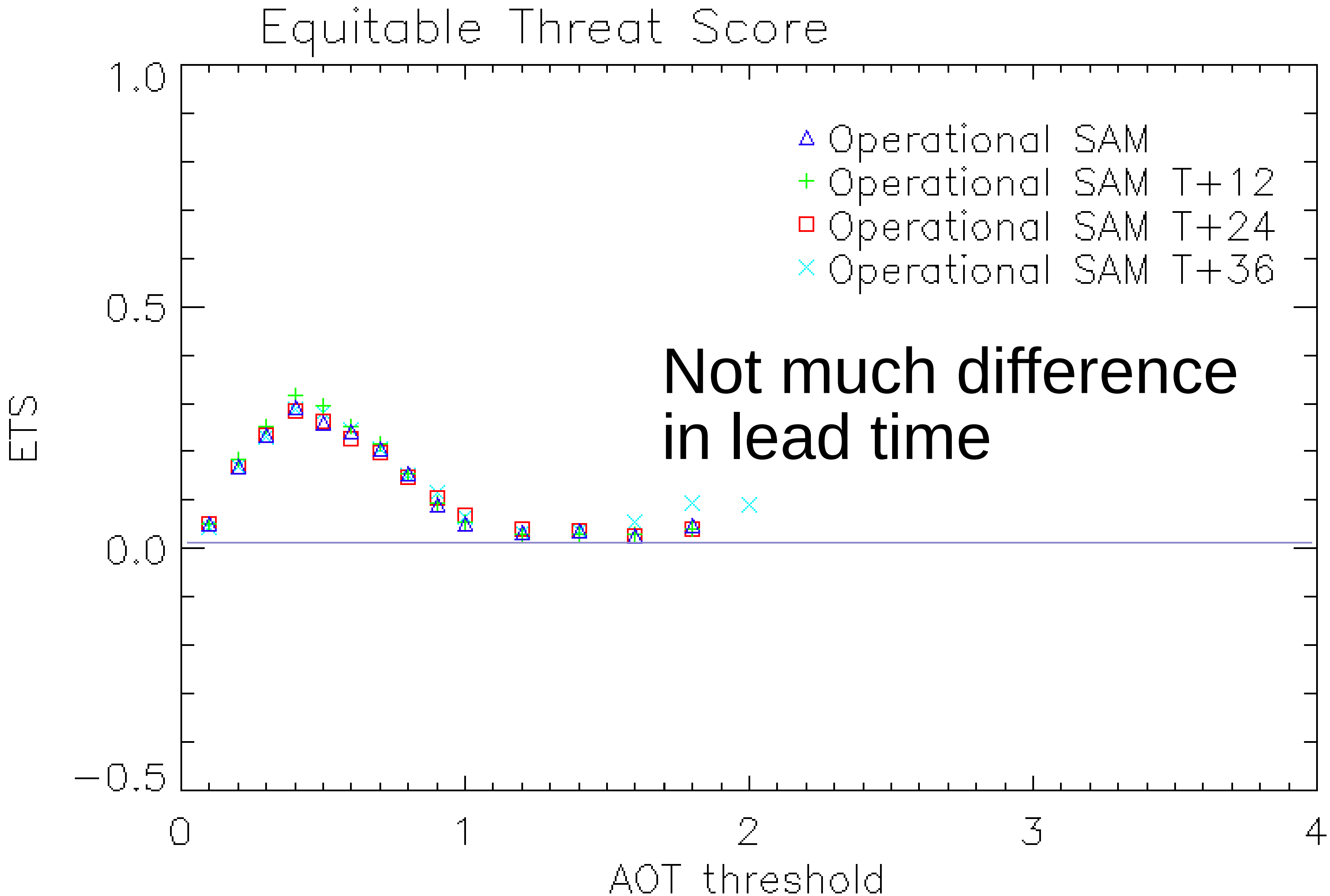
- By defining criteria for the centre of mass of isolated objects, a contingency table can be created to see how the model performs as a function of threshold



AERONET –site-specific verification



AERONET ETS @ Mussafa



Not much difference in lead time

Conclusions

Conclusions

- Point and grid verification, will not necessarily give the same answer. Neither will traditional methods and new spatial methods.

- New methods can be really useful but not all new spatial methods are equally useful for specific applications.

- A tangible impact of the double penalty effect will depend on the parameter and the resolution of your forecast/observation.

- Understanding the limitations of your observations is critical.

- The most challenging aspect of a verification process is making the decision on what to do and how to do it. Calculating the actual scores is often trivial. Data preparation and choice of methodology are the most important aspects to get right.

Questions?