Diagnostic evaluation of precipitation forecasts at multiple spatial scales

Barbara Brown, John Halley Gotway,
Chris Davis, and Lacey Holland

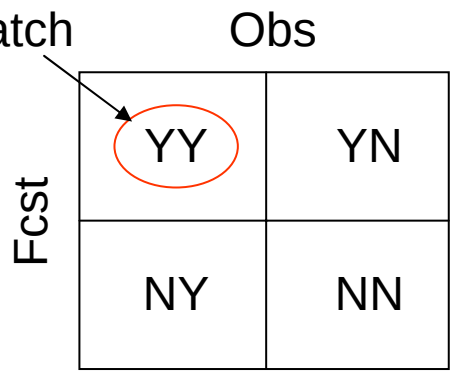
National Center for Atmospheric Research (NCAR)
Boulder CO USA
bgb@ucar.edu

February 2007

Collaborators: Randy Bullock, Eric Gilleland, Paul Kucera, and David Ahijevych

Traditional "Measures"-Based Approach

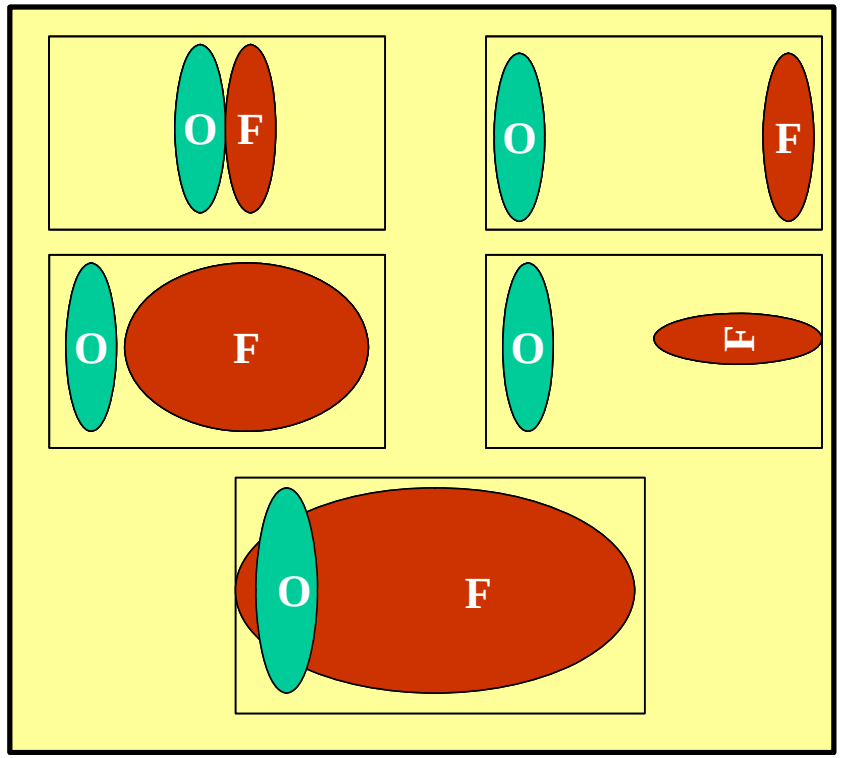
Consider forecasts and observations of some dichotomous field on a grid:



Critical Success Index
CSI=YY/(YY+NY+YN)

Equitable Threat Score

ETS=(YY-ε)/(YY+NY+YN-ε), where
ε=success due to chance

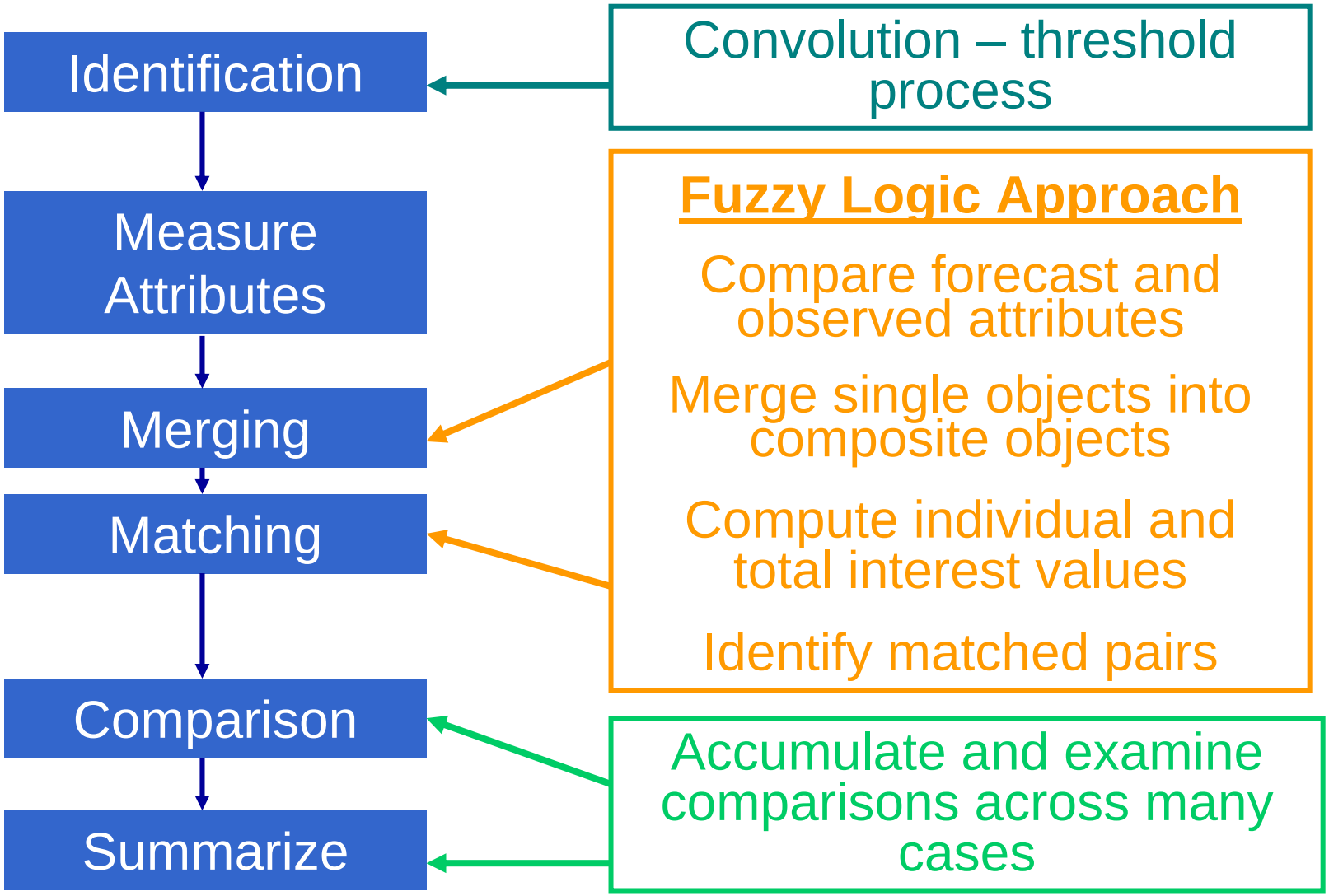


CSI = 0 for first 4;

CSI > 0 for the 5th

Non-diagnostic and utra-sensitive to small errors in simulation of localized phenomena!

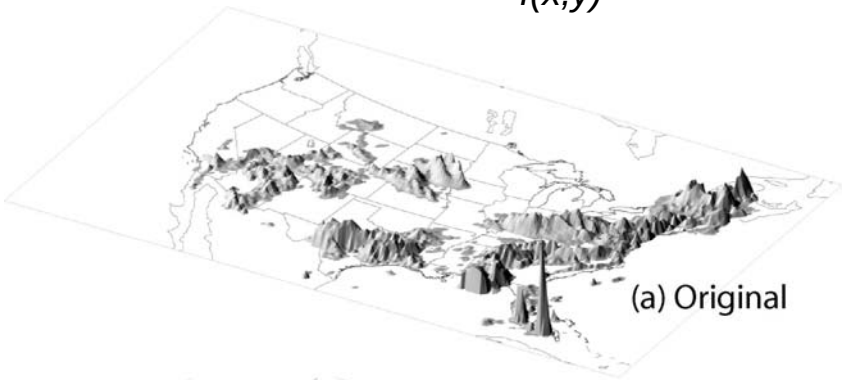
MODE*: Object-based approach

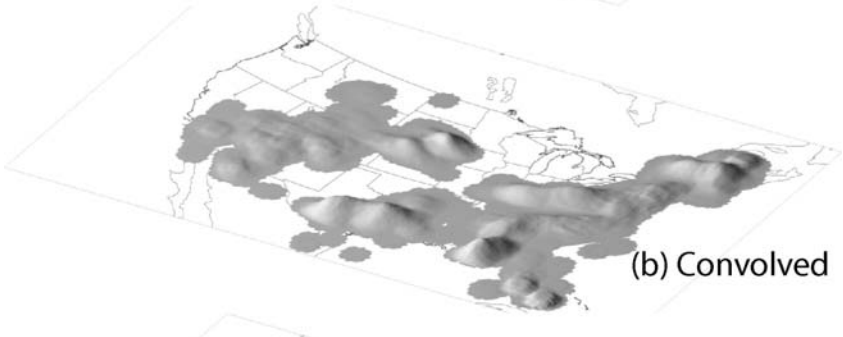


*Method for Object-based Diagnostic Evaluation

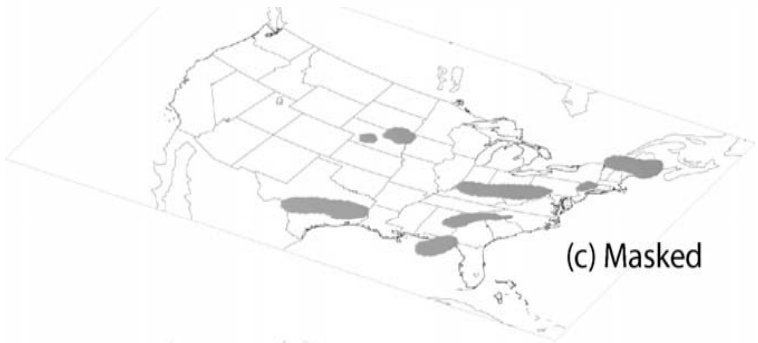
Object identification

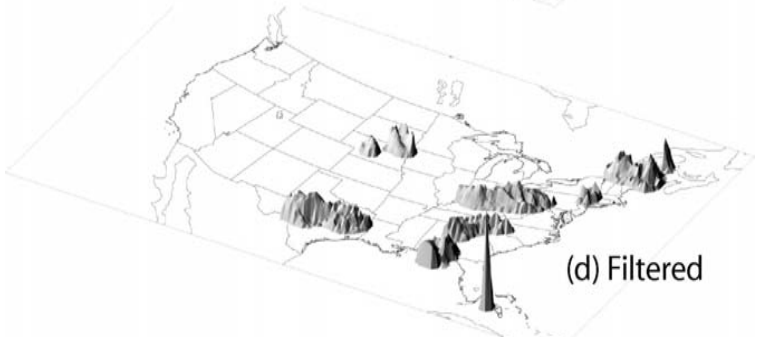
f(x,y)





$$h(x,y) = \begin{cases} 1 & \text{if } g(x,y) \geq T \\ 0 & \text{otherwise} \end{cases}$$





Restore original field where h(x,y) =1

$$g(x,y) = \sum_{(u,v)\in G} \phi(u,v)\, f(x-u, y-v)$$

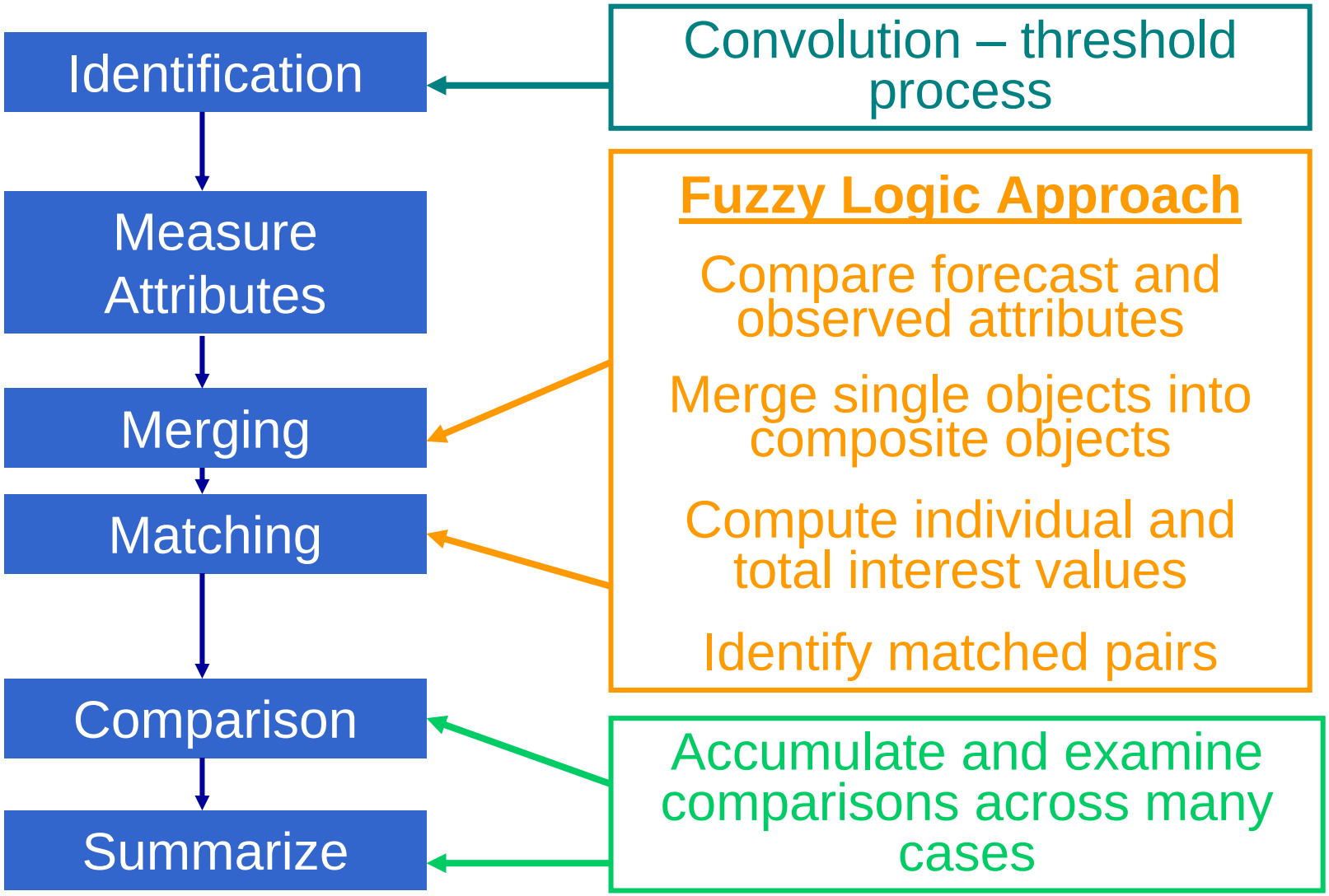
$$\phi(x,y) = \begin{cases} H & \text{if } x^2 + y^2 \leq R^2 \\ 0 & \text{otherwise} \end{cases}$$

$$\pi R^2 H = 1.$$

2 parameters:

1. Convolution radius
2. Threshold

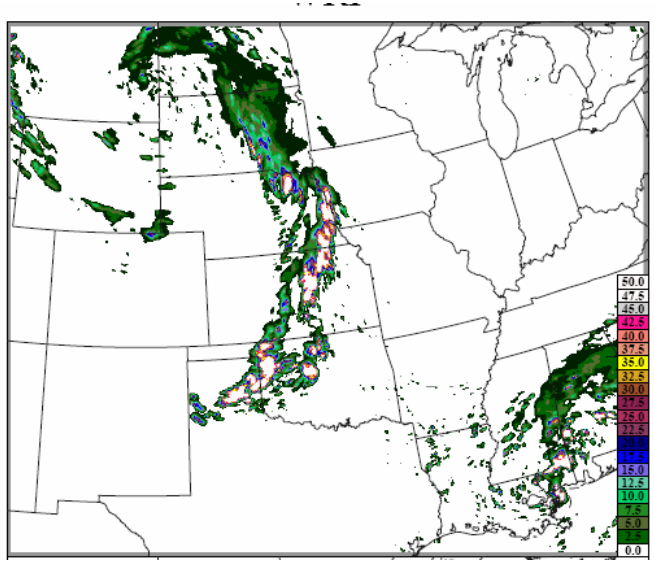
MODE*: Object-based approach



*Method for Object-based Diagnostic Evaluation

Observations and model

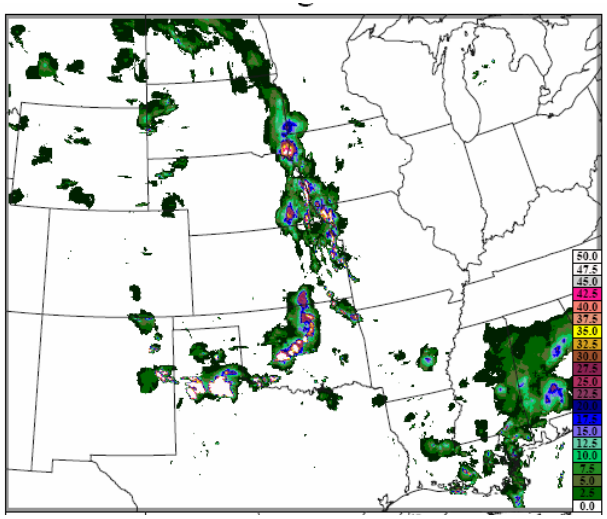
- *Forecasts*: Weather Research and Forecasting (WRF) model
  - Advanced Research WRF (ARW), 4-km grid spacing
  - Forecasts initialized at 0000 UTC from Eta initial conditions
  - 24-h lead
  - 1-h precipitation accumulation
  - 18 April – 4 June, 2005; 9 cases selected for extensive study
  - Study Domain: United States, Rocky Mountains (west) to Appalachian Mountains (east)
- *Observations*: Multi-sensor hourly accumulated precipitation
  - Stage II on 4-km grid

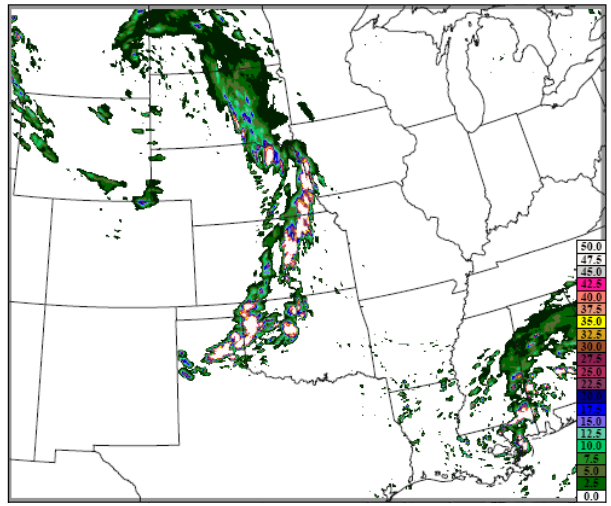


Stage II precipitation estimate; 1 June 2005, 0000 UTC

Object-based example: 1 June 2005

WRF (24-h)

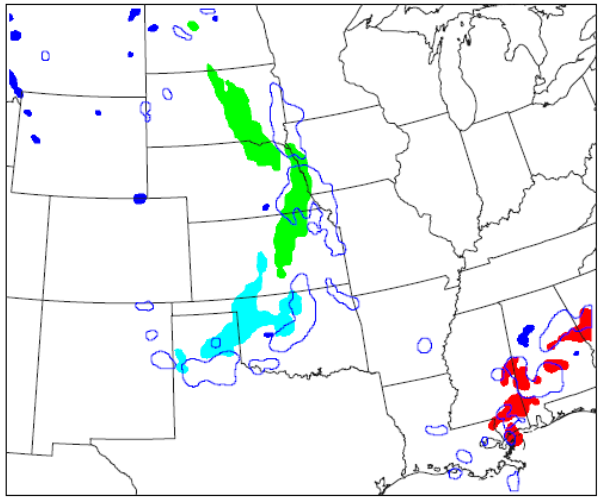


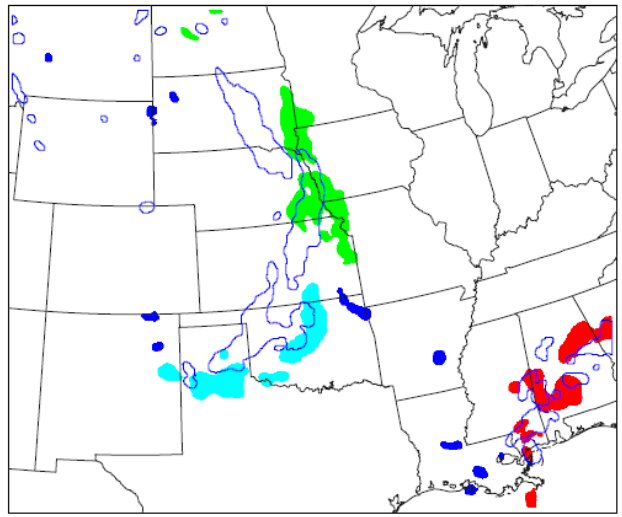


Stage II

WRF Objects with StageII Outlines

StageII Objects with WRF Outlines

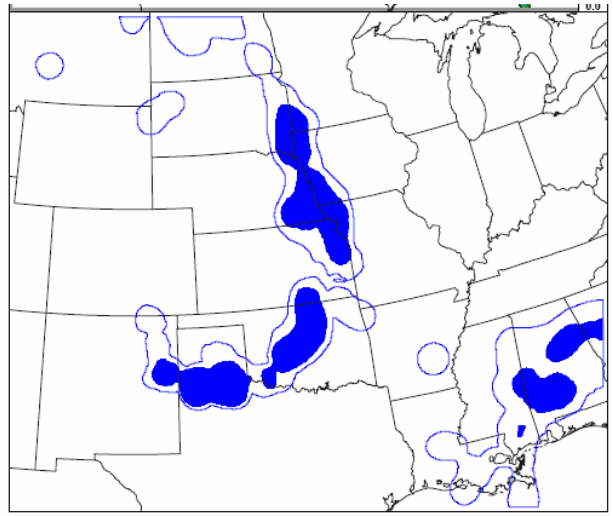


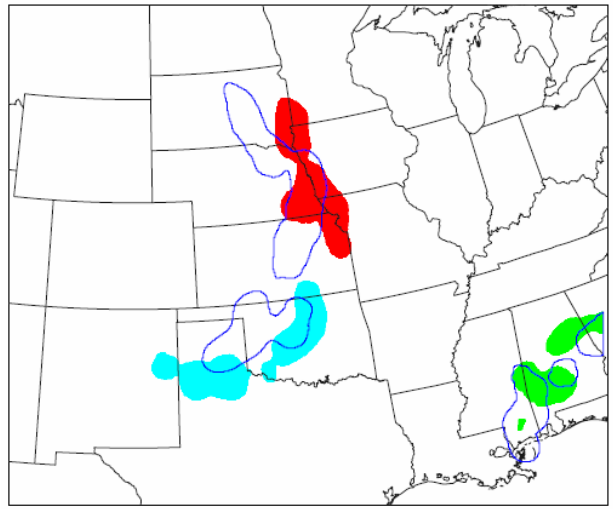


Radius = 5 grid squares, Threshold = 0.05”

Issues: Matching and merging

- Evaluation of matching and merging procedures
  - Two-step process merges observed objects separately from forecast objects
  - But – 2-step process leads to non-optimal matches between forecast and observed objects
  - Double-threshold, single step procedure appears to be most robust, provide most reasonable results

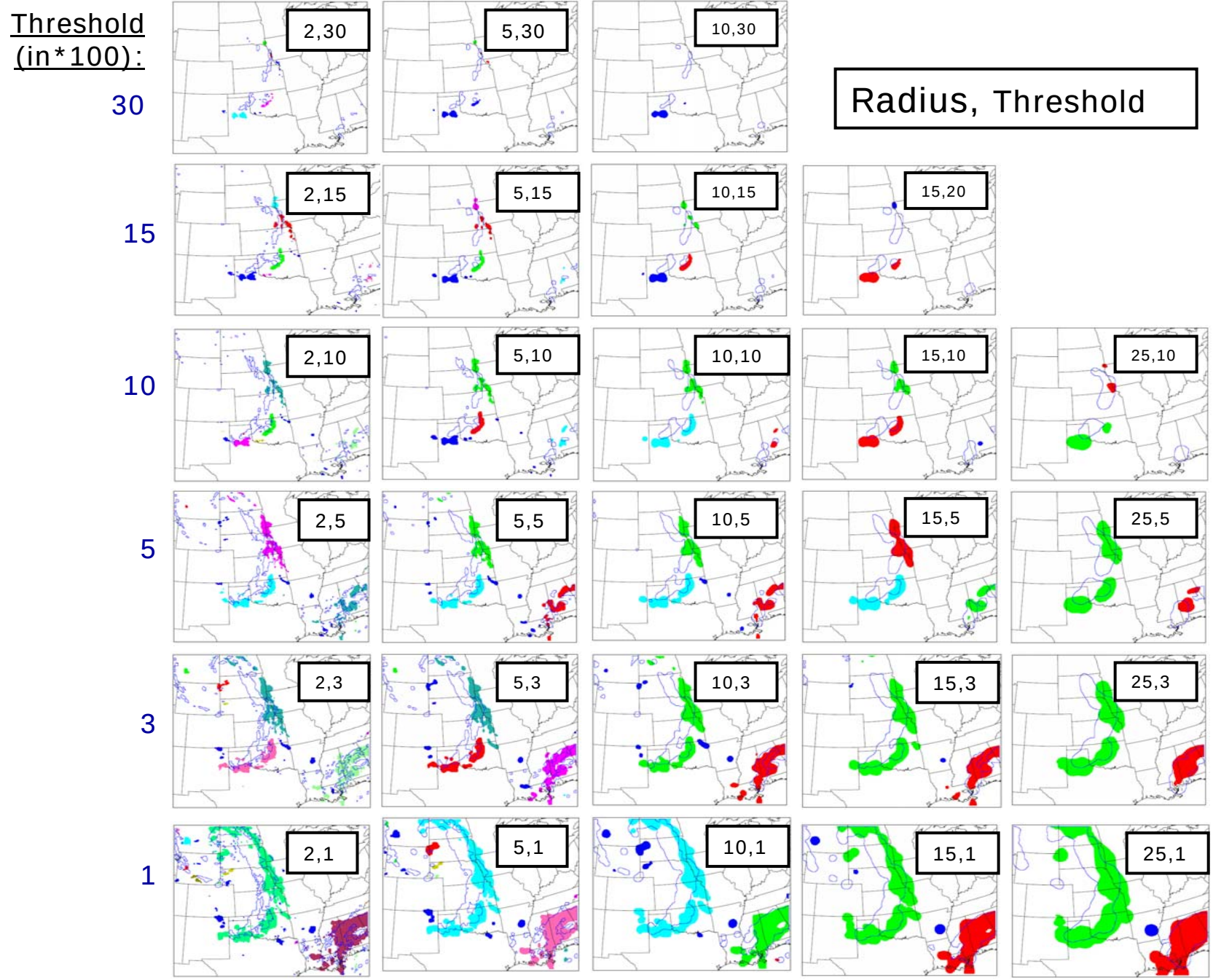




Issues: Object identification and scale

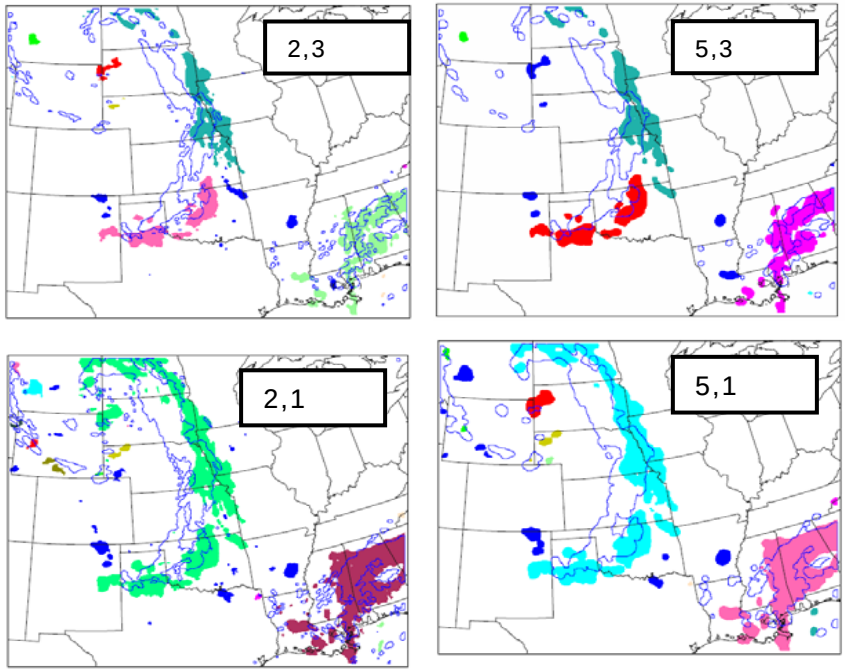
- How should object identification parameters (radius, threshold) be selected?
- Alternative question: What scale(s) are appropriate and meaningful?

Goal: Examine impacts of scale on object and matching properties

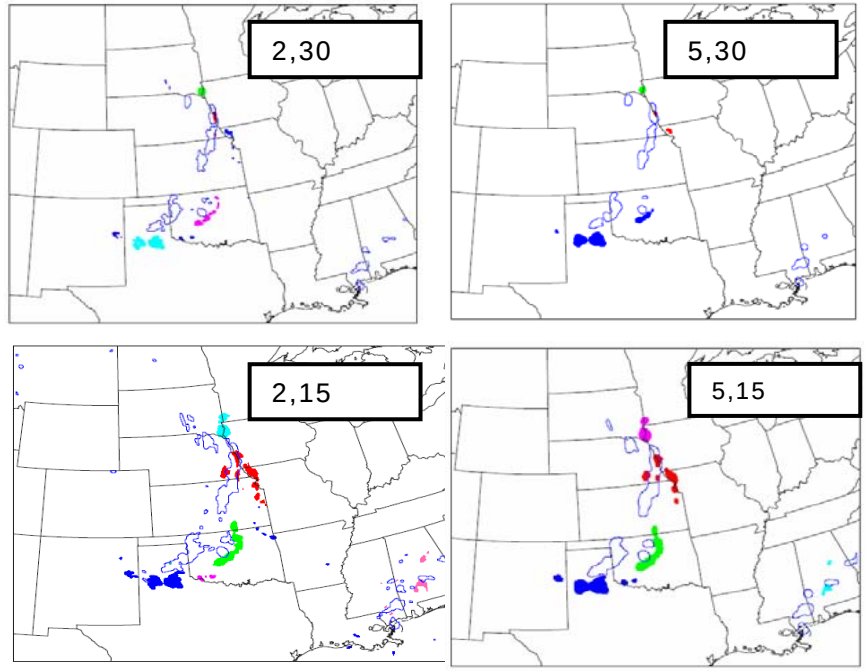

Threshold (in*100):
30
15
10
5
3
1
Radius, Threshold
2,30
5,30
10,30
2,15
5,15
10,15
15,20
2,10
5,10
10,10
15,10
25,10
2,5
5,5
10,5
15,5
25,5
2,3
5,3
10,3
15,3
25,3
2,1
5,1
10,1
15,1
25,1
Radius (grid boxes):
2
5
10
15
25

Scale features

Small radius, Low threshold



Small radius, High threshold

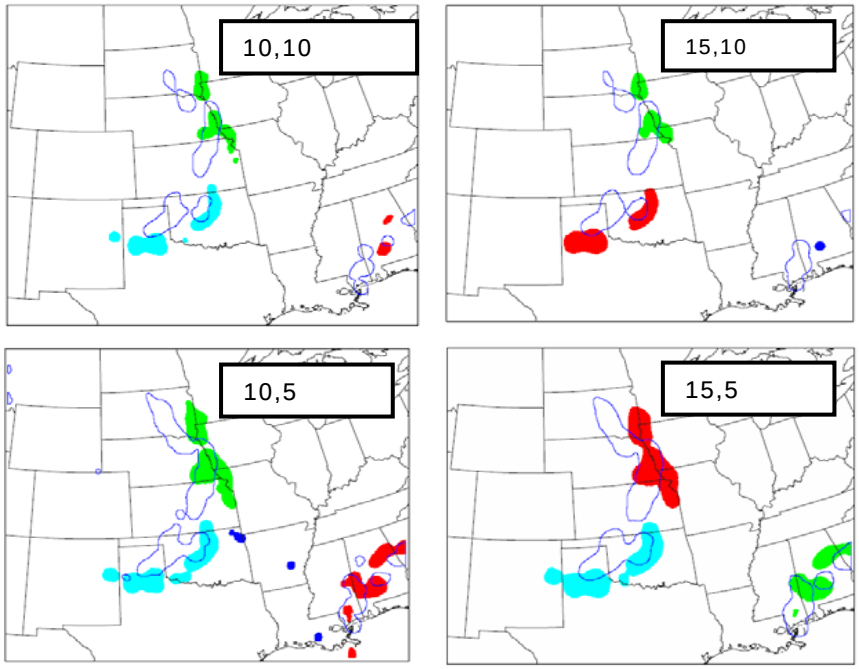


Both are characterized by sharp features, complexity.

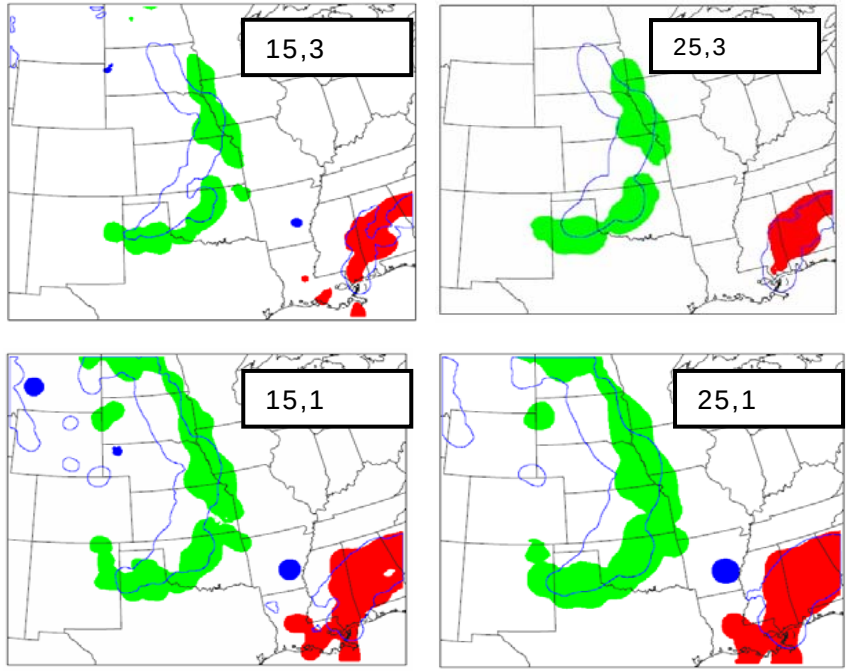
High threshold results in many small objects, fine-scaled features

Scale features

Medium radius, Medium threshold



Large radius, Low threshold

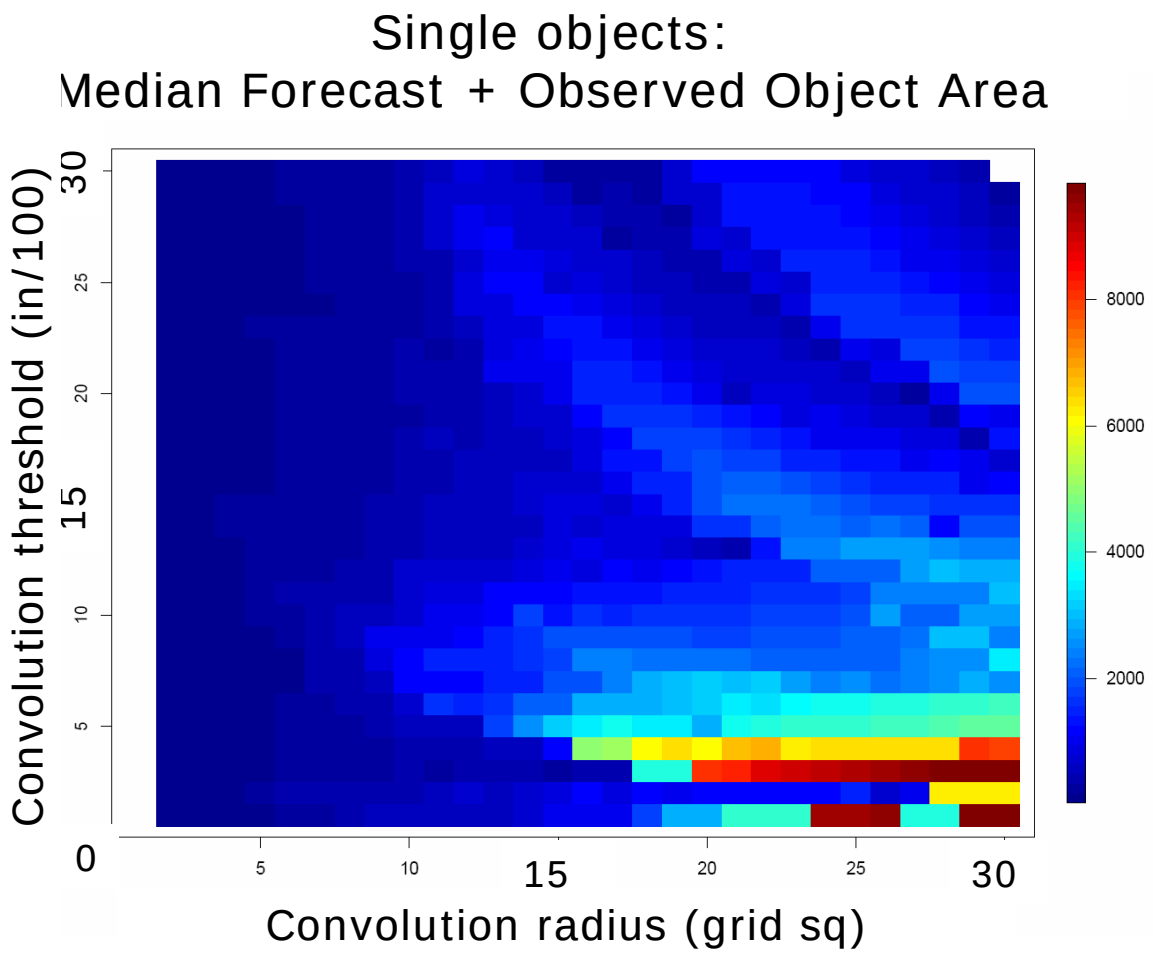


Large radius creates large blobby objects.

Medium parameters focus on more intense rain areas.

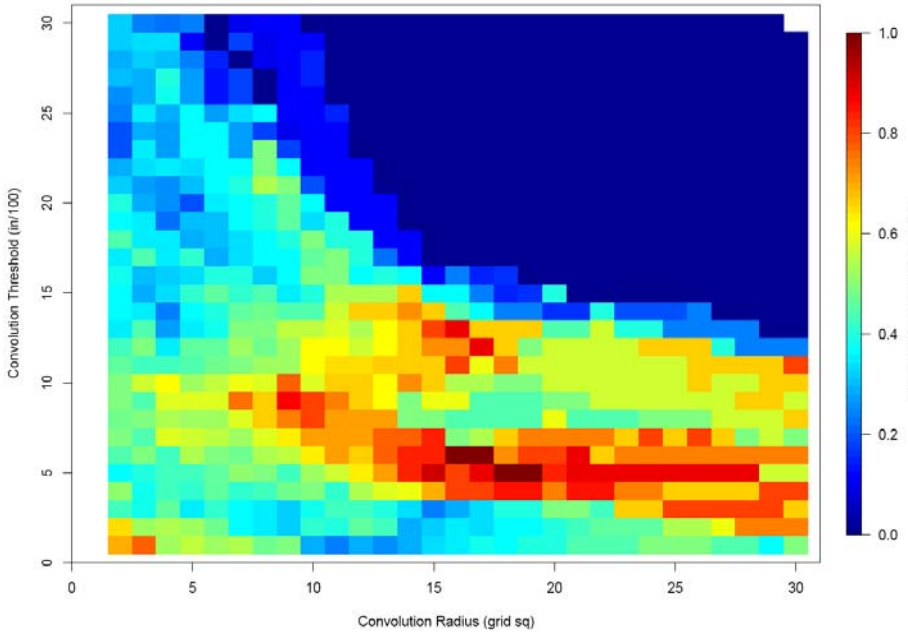
Verification "Quilts"

- Forecast performance attributes as a function of spatial scale
- Similar to charts developed by Casati, Marzban, Ebert

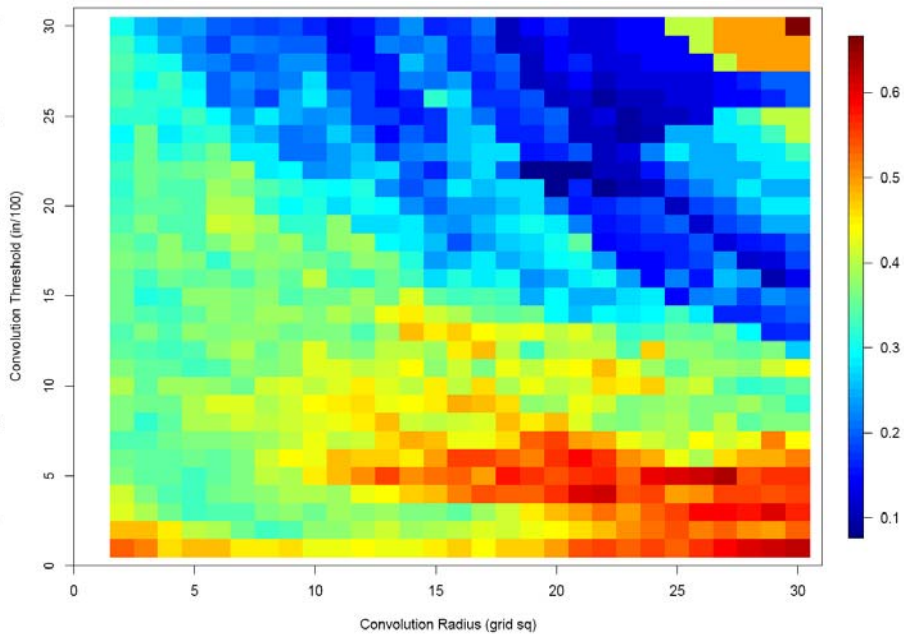


Percent of single objects matched

1 June 2005



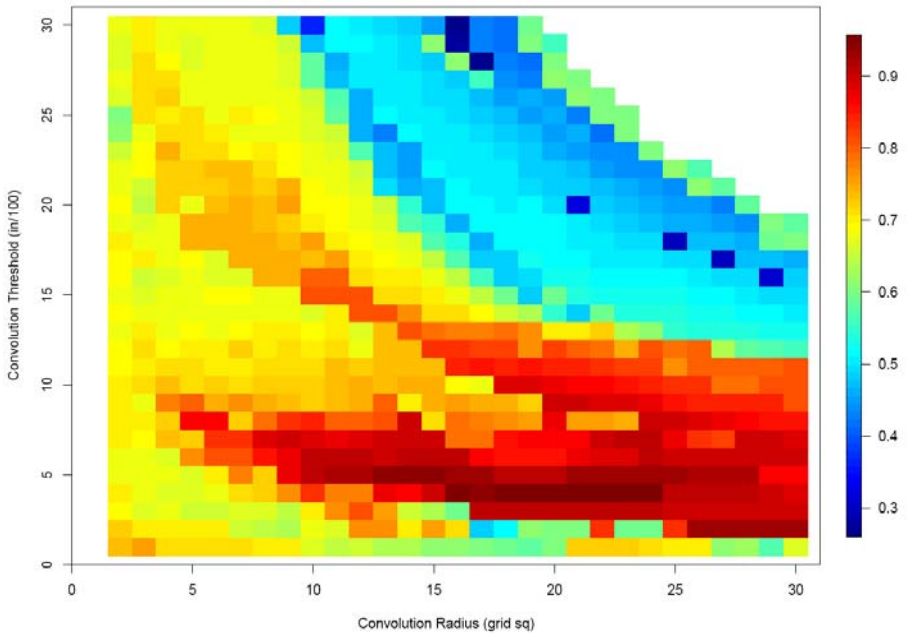
9 Cases



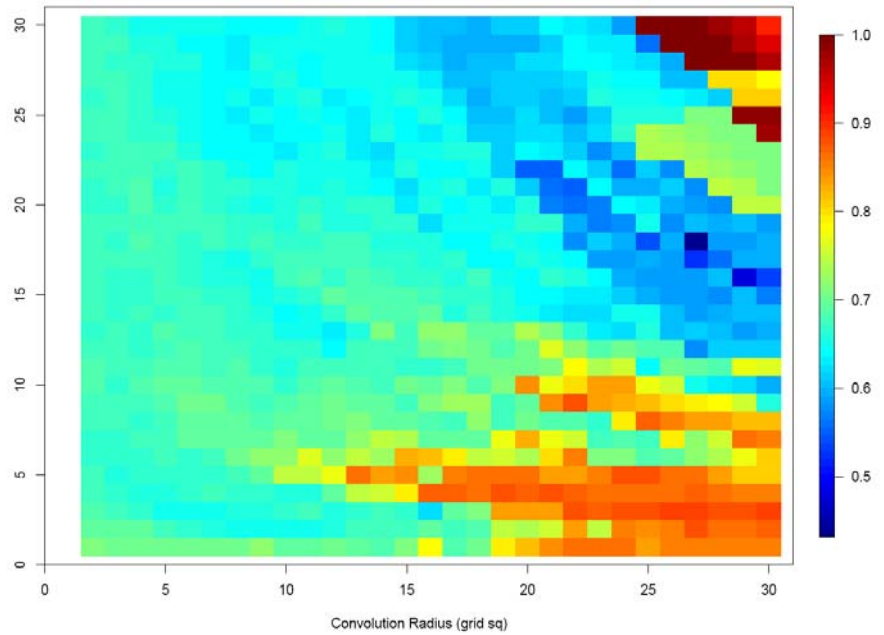
- Region with large radius and large threshold has low rate of matches, except for most extreme values
- Region with moderate values of radius and low threshold (around 5) is the scale with best potential for object matching

Measure of matching strength

1 June 2005



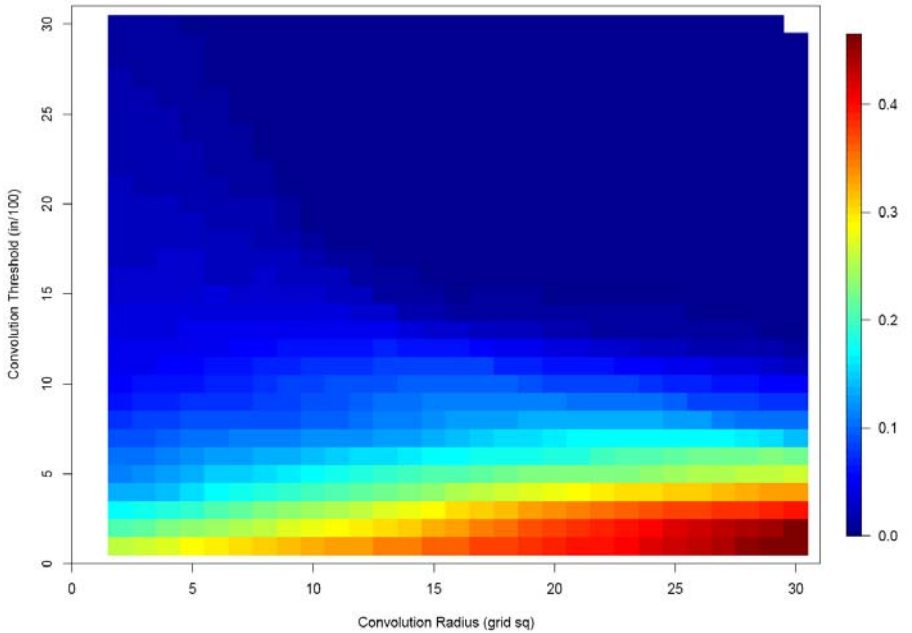
9 Cases



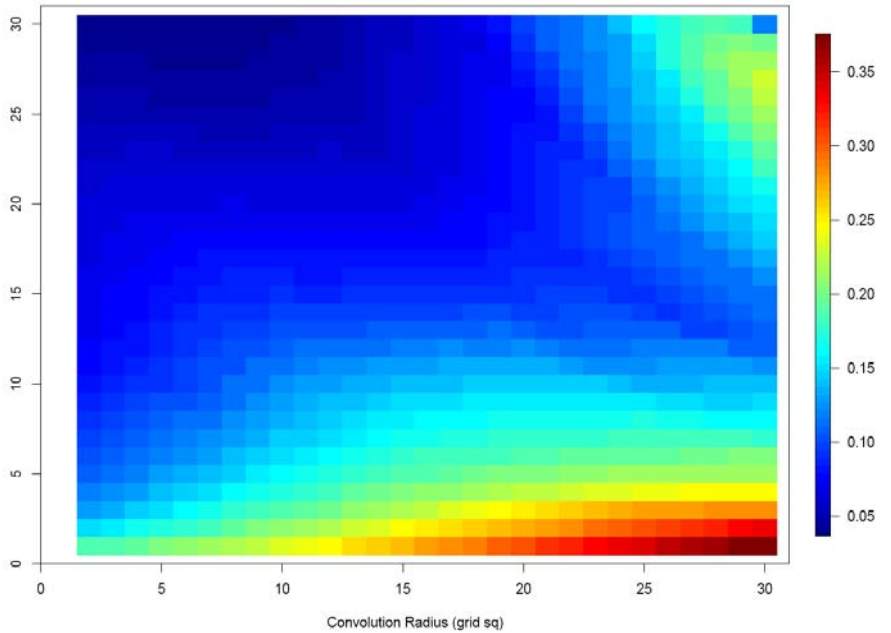
- Region with moderate values of radius and low threshold (around 5) is the scale with best potential for object matching
- A measure of skill?

Critical success index

1 June 2005



9 Cases



- Highly dependent on radius
  - Largest values for smooth objects
  - Less dependence on threshold

Conclusions

- Matching capabilities are – not surprisingly – highly dependent on scale of objects
- Verification "quilts" help define scales with potential skill
- The appropriate question should be

  Which scales are reasonable to
  - Examine in the context of users' applications?
  - Provide a meaningful evaluation of forecast skill (and other attributes)?

  Thus – it is more appropriate to examine objects associated with several representative sets of parameters, rather than focusing on a single set.