Method for Object-based Diagnostic Evaluation (MODE)

A *features-based* spatial forecast verification technique

Eric Gilleland (*et al., see penultimate slide*)
National Center for Atmospheric Research (NCAR), Boulder, Colorado.

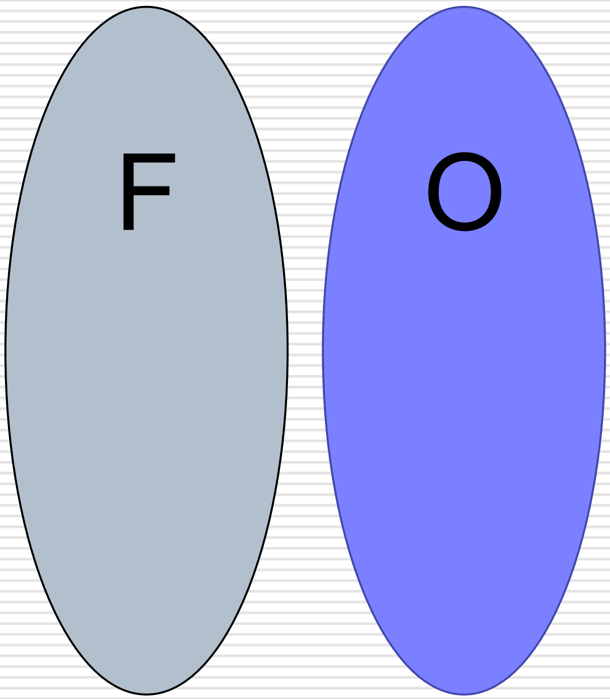
Forecast Verification Issues

- Incorporation of Uncertainty in:
  - verification statistics
  - observations
- Diagnostic and user-relevant verification
- Spatial forecast verification
  - Verification of high-resolution forecasts
- Verification of *non-traditional* forecasts
  - probabilistic
  - ensemble forecasts
- Verification for extreme or rare events
- Properties of verification measures
  - Propriety, Equitability

Forecast Verification Issues

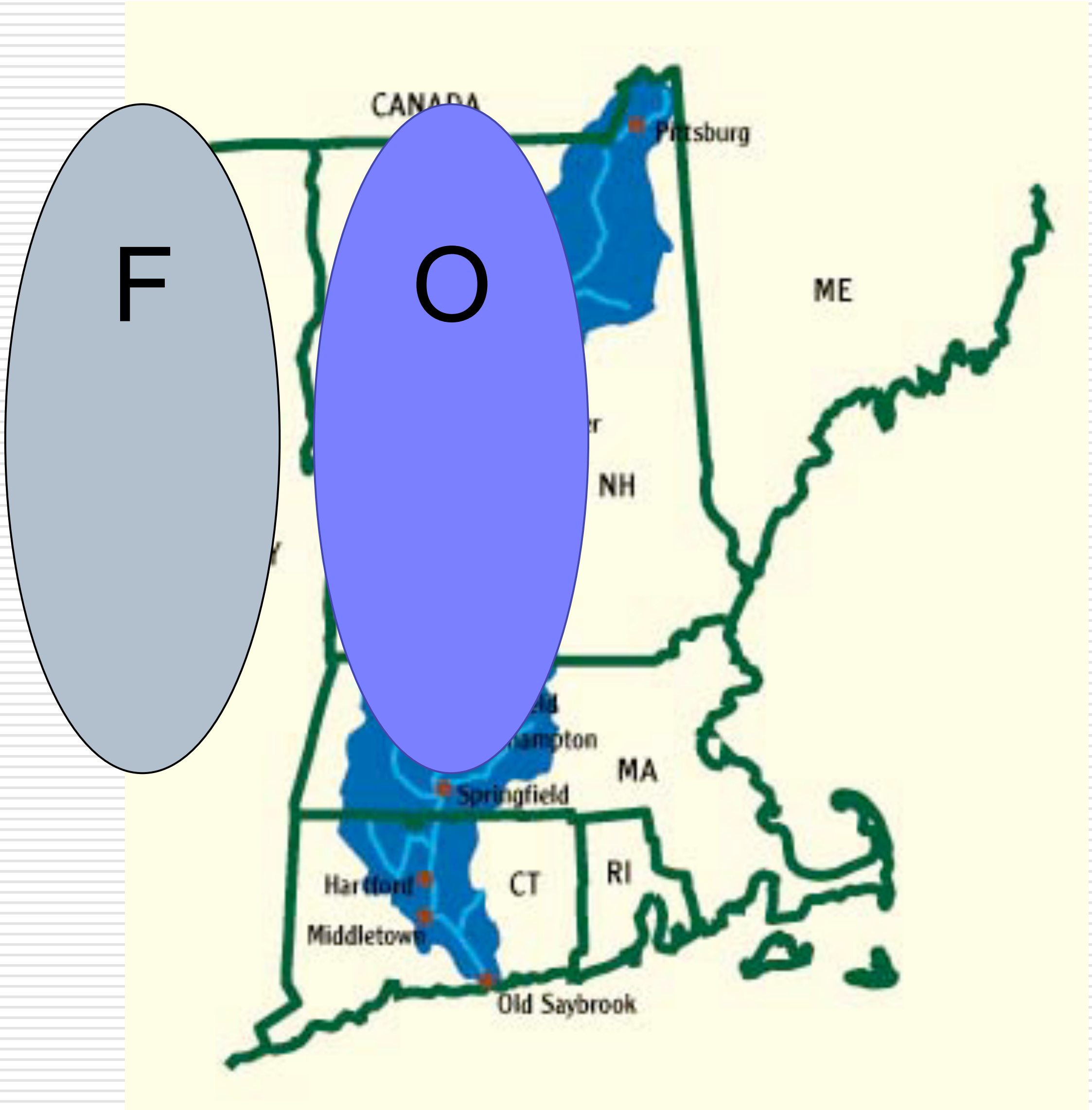
- Incorporation of Uncertainty in:
  - verification statistics
  - observations
- Diagnostic and user-relevant verification
- Spatial forecast verification
  - Verification of high-resolution forecasts
- Verification of *non-traditional* forecasts
  - probabilistic
  - ensemble forecasts
- Verification for extreme or rare events
- Properties of verification measures
  - Propriety, Equitability

User-relevant verification: Good forecast or Bad forecast?

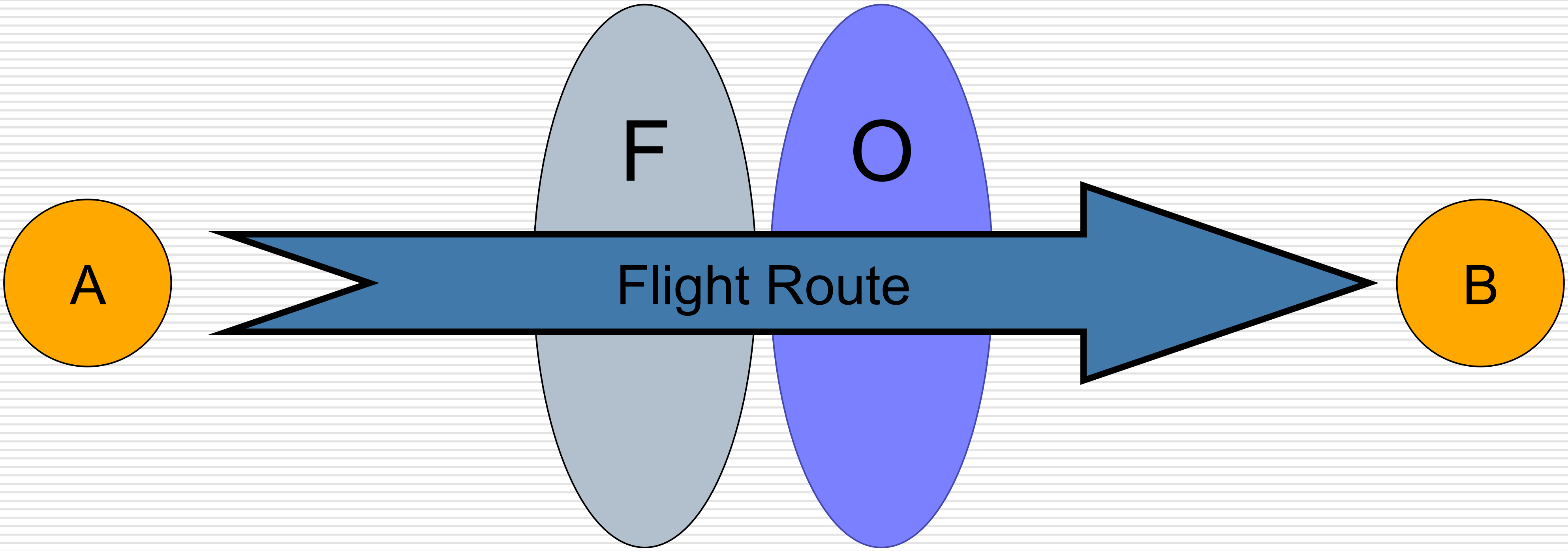


User-relevant verification: Good forecast or Bad forecast?

If I'm a water manager for this watershed, it's a pretty bad forecast…



User-relevant verification: Good forecast or Bad forecast?



If I'm an aviation traffic strategic planner…

It might be a pretty good forecast

Different users have different ideas about what makes a good forecast

Diagnostic and user relevant forecast evaluation approaches

- Provide the link between weather forecasting and forecast value
- Identify and evaluate attributes of the forecasts that are meaningful for particular users
  - Users could be managers, forecast developers, forecasters, decision makers
  - Answer questions about forecast performance in the context of users' decisions
  - Example questions: How do model changes impact user-relevant variables? What is the typical location error of a thunderstorm? Size of a temperature error? Timing error? Lead time?

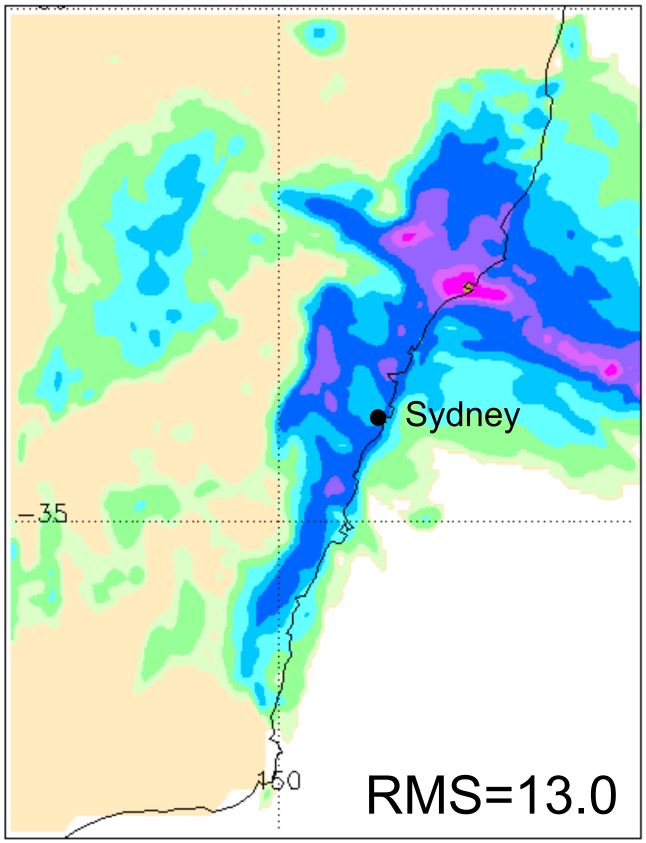
Diagnostic and user relevant forecast evaluation approaches (cont.)

- Provide more detailed information about forecast quality
  - What went wrong? What went right?
  - How can the forecast be improved?
  - How do 2 forecasts differ from each other, and in what ways is one better than the other?

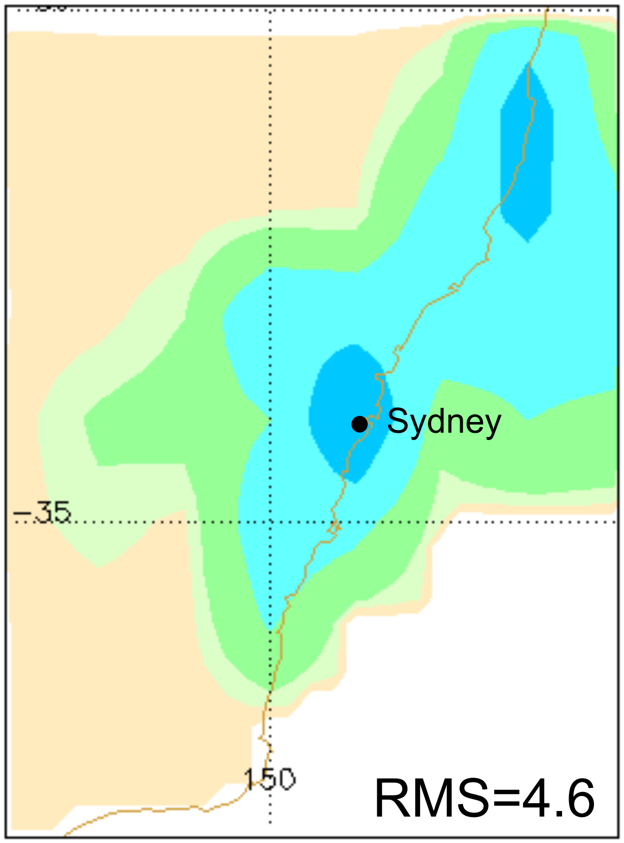
High vs. low resolution

Which rain forecast is better?

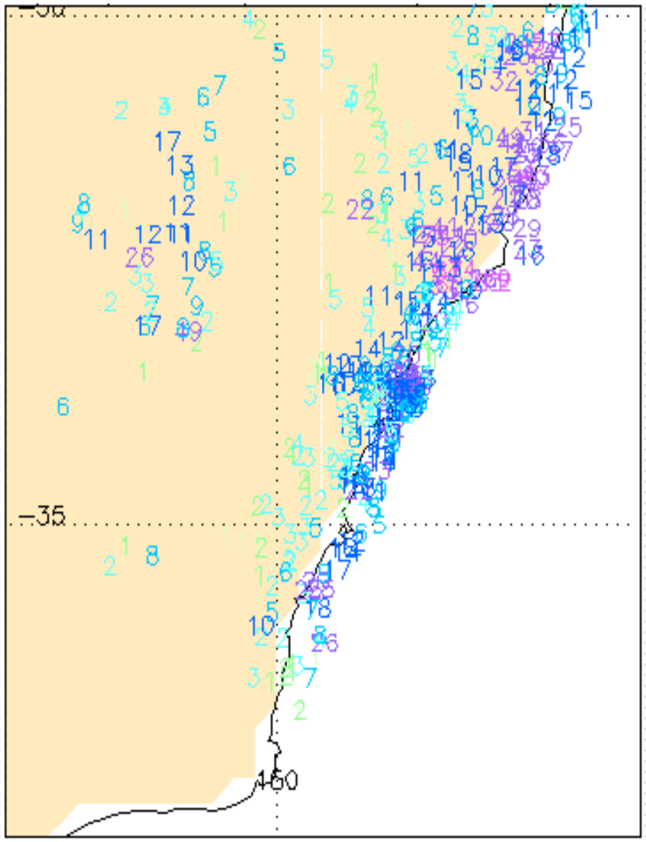
Mesoscale model (5 km) 21 Mar 2004



Global model (100 km) 21 Mar 2004



Observed 24h rain

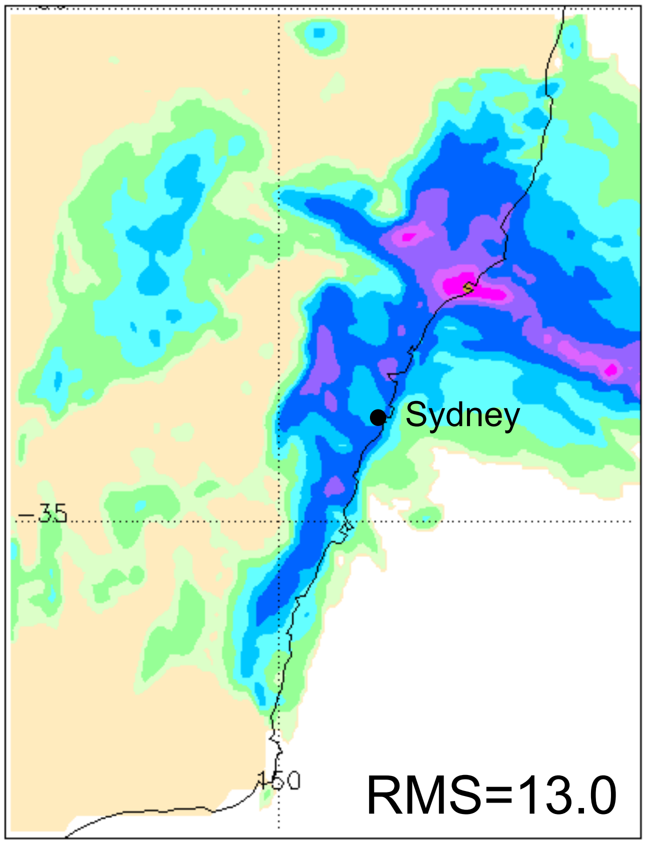


From E. Ebert

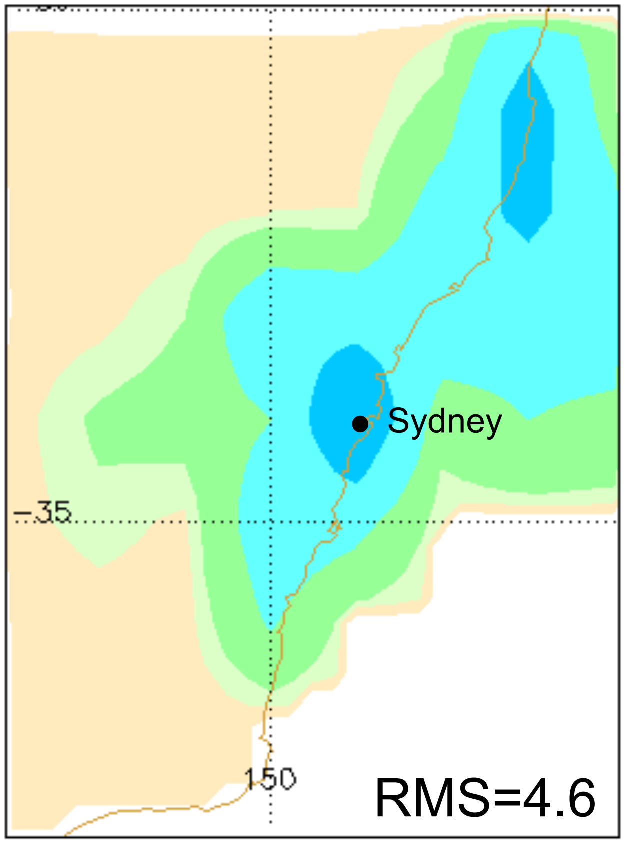
High vs. low resolution

Which rain forecast is better?

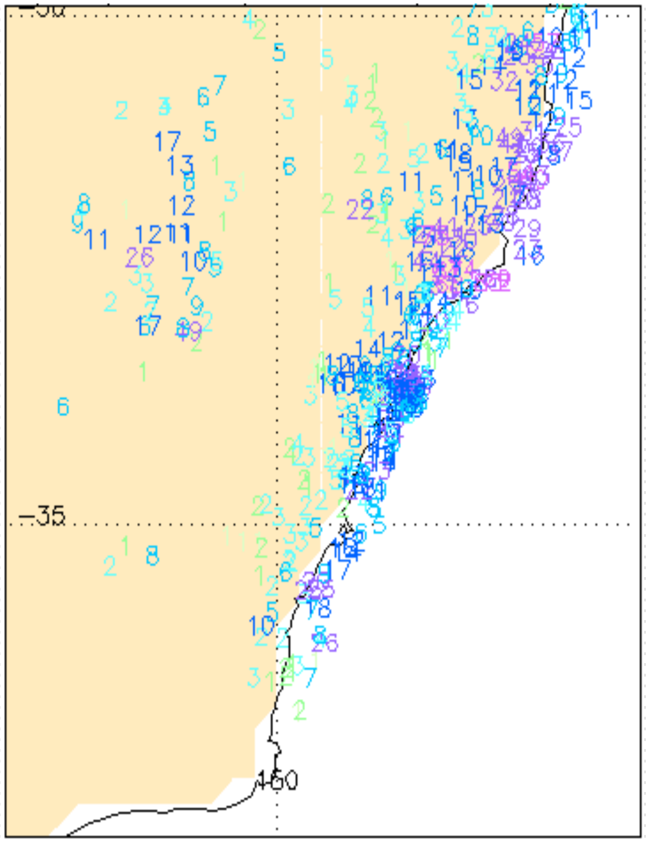
Mesoscale model (5 km) 21 Mar 2004



Global model (100 km) 21 Mar 2004



Observed 24h rain

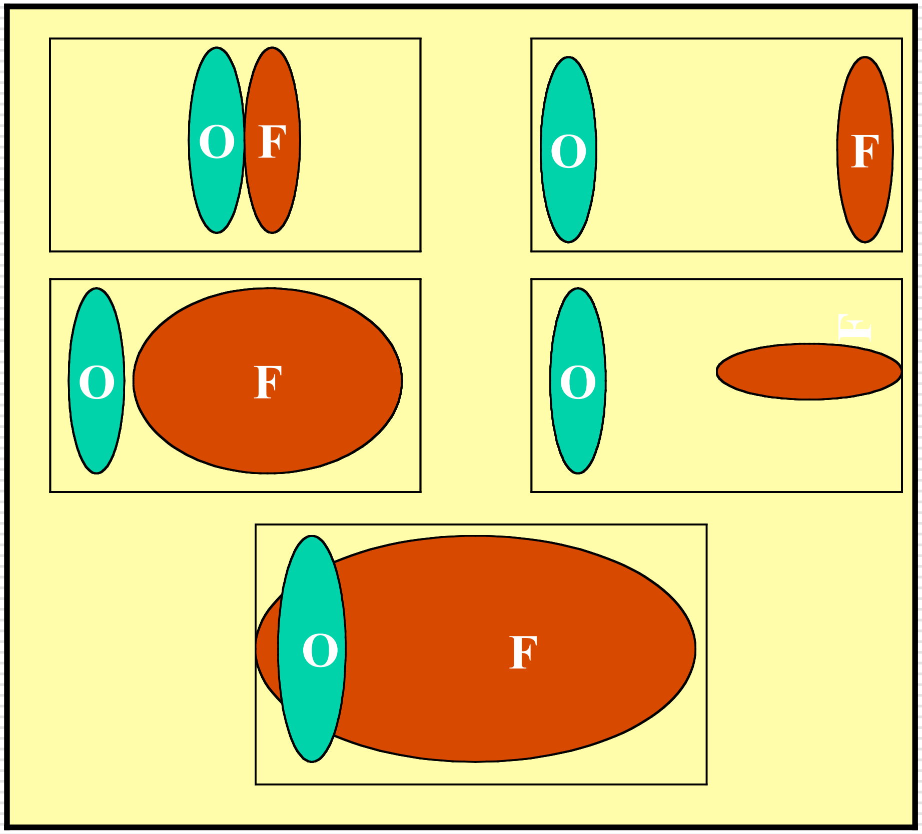


From E. Ebert

"Smooth" forecasts generally "Win" according to traditional verification approaches.

Traditional "Measures"-based approaches

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



Some problems with this approach:

(1) ***Non-diagnostic*** *– doesn't tell us what was wrong with the forecast – or what was right*

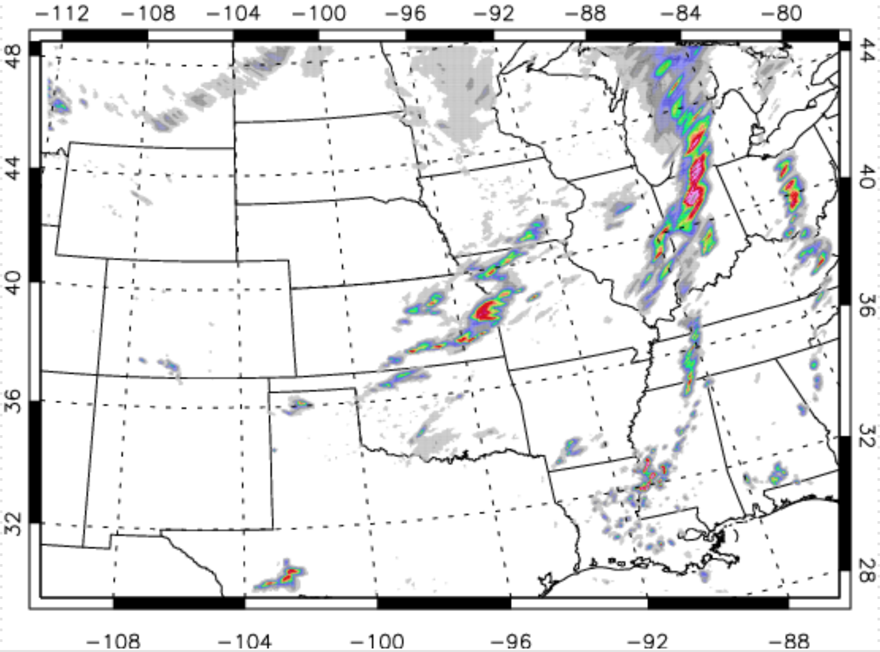
(2) ***Ultra-sensitive*** *to small errors in simulation of localized phenomena*

CSI = 0 for first 4;

CSI > 0 for the 5th

Spatial forecasts

Weather variables defined over spatial domains have **coherent structure and features**



Spatial verification techniques aim to:

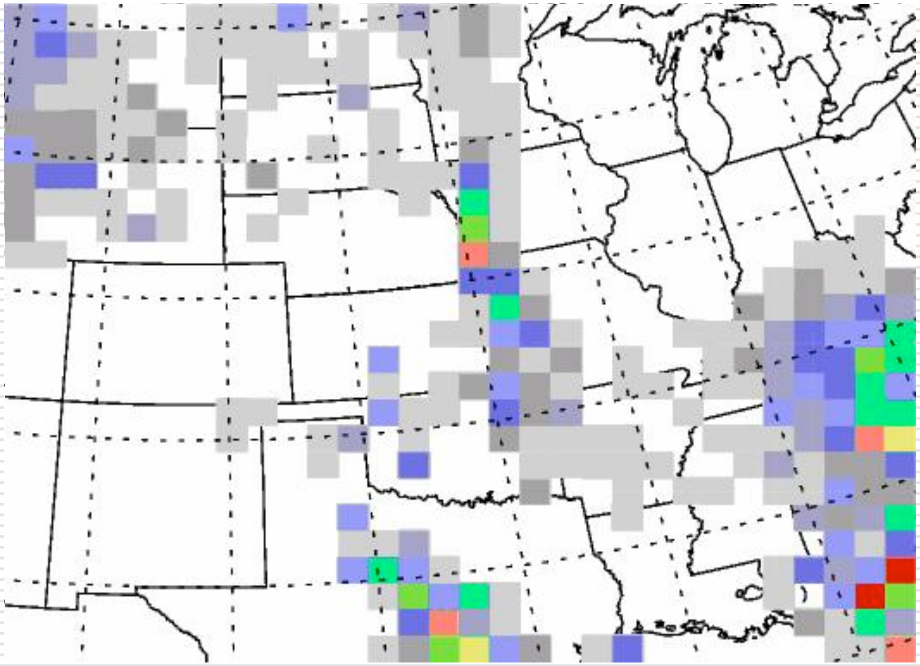
- account for uncertainties in timing and location
- account for field spatial structure
- provide information on error in physical terms
- provide information that is
  - diagnostic
  - meaningful to forecast users

Recent research on spatial verification methods

- Neighborhood verification methods
  - give credit to "close" forecasts
- Scale decomposition methods
  - measure scale-dependent error
- Feature-based methods
  - evaluate attributes of identifiable features
- Field morphing verification approaches
  - measure distortion and displacement for the whole field

Neighborhood verification

- Also called "fuzzy" verification
- Upscaling
  - put observations and/or forecast on coarser grid
  - calculate traditional metrics



Ebert (2007; Met Applications) provides a review and synthesis of these approaches

Fractions skill score (Roberts 2005; Roberts and Lean 2007)

Scale decomposition

- Errors at different scales of a single-band spatial filter (Fourier, wavelets,…)
  - Briggs and Levine, 1997
  - Casati *et al*., 2004
- Removes noise
- Examine how different scales contribute to traditional scores
- Does forecast power spectra match the observed power spectra?

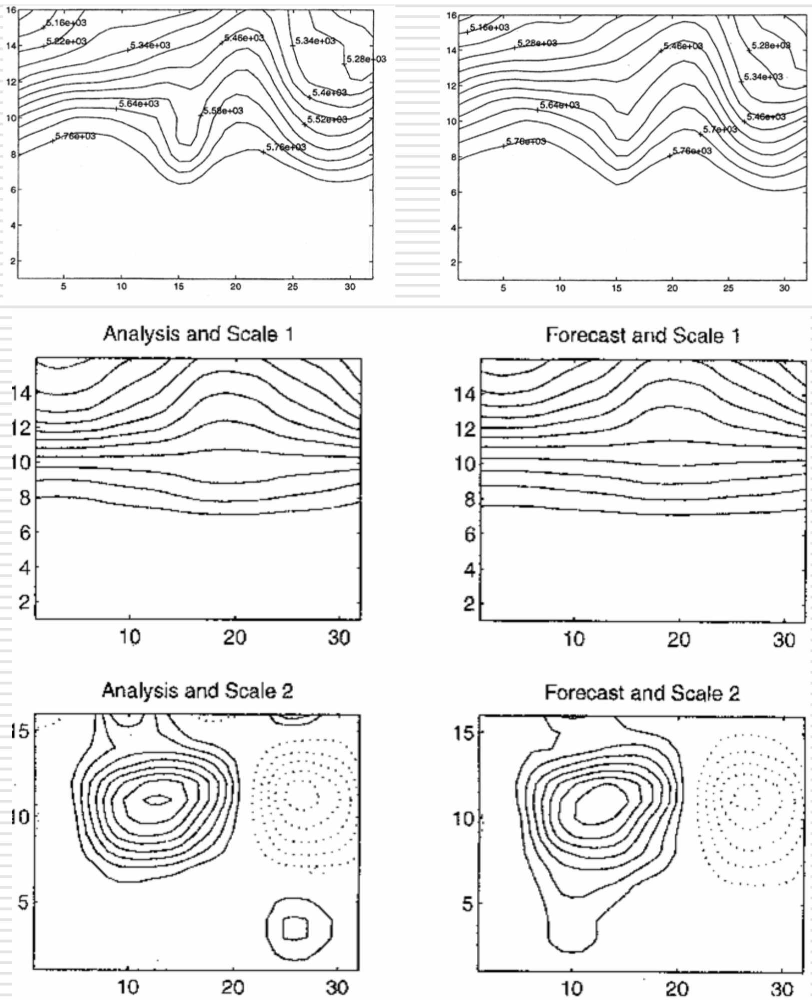
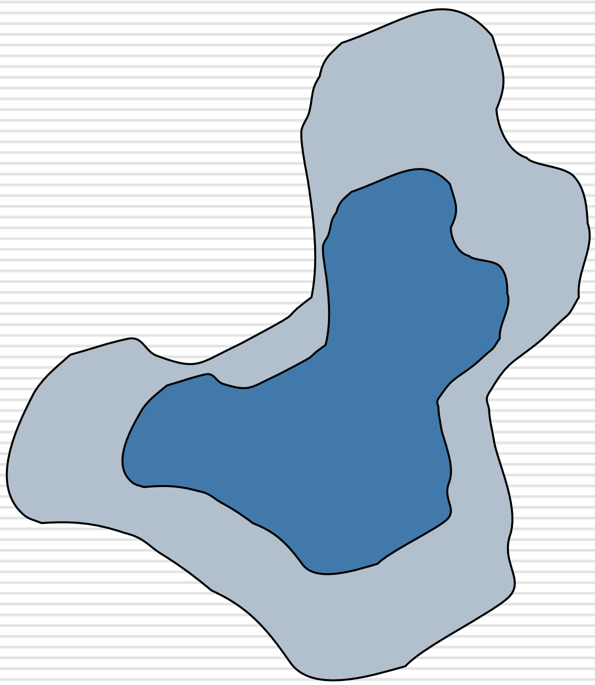


Fig. from Briggs and Levine, 1997

Feature-based verification



Error components

- displacement
- volume
- pattern

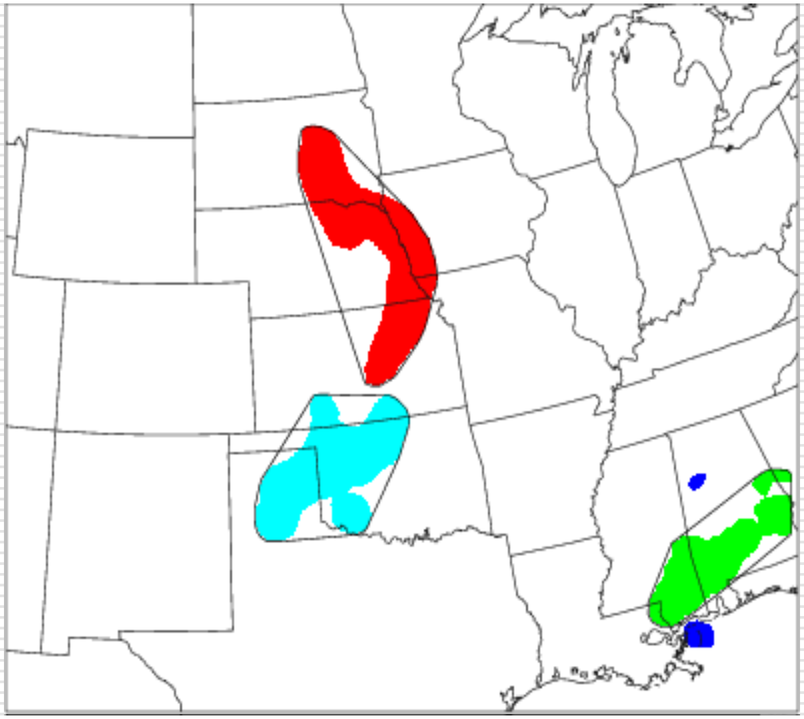
Numerous features-based methods

- Composite approach (Nachamkin)
- Contiguous rain area approach (CRA; Ebert and McBride, 2000; Gallus and others)

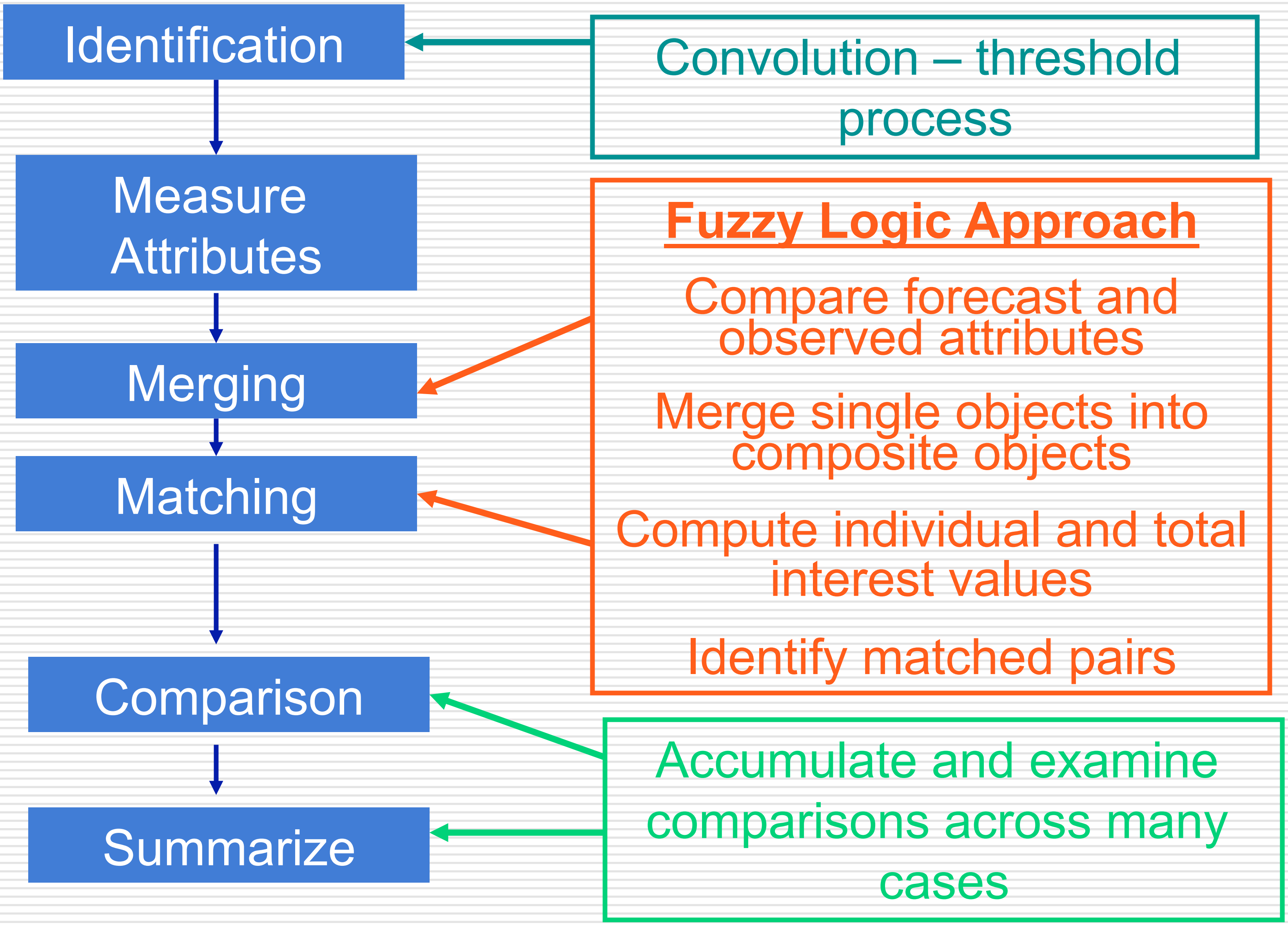
Gratuitous photo from Boulder open space

Feature- or object-based verification

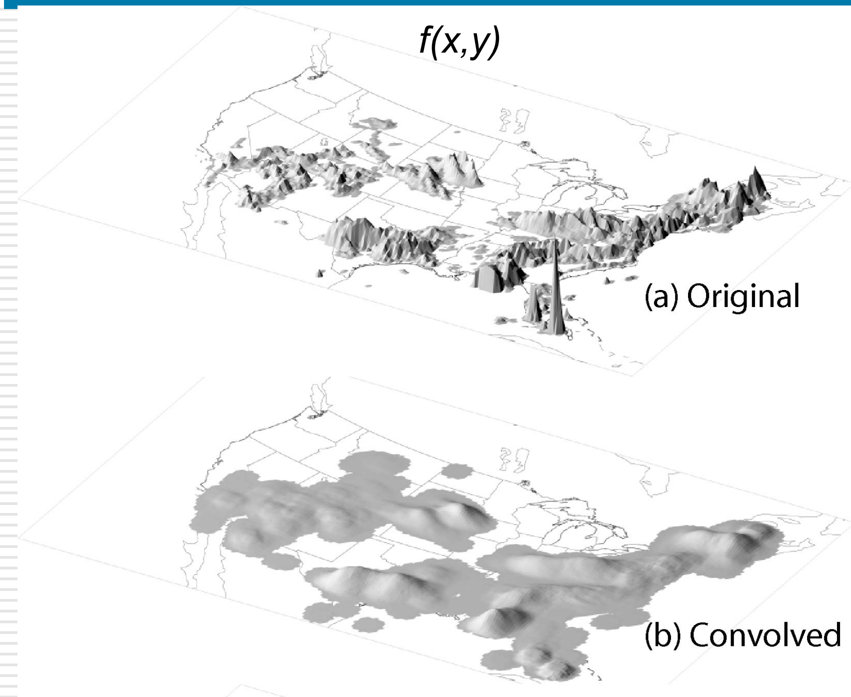
- Baldwin object-based approach
- Cluster analysis (Marzban and Sandgathe)
- SAL approach for watersheds
- Method for Object-based Diagnostic Evaluation (MODE)
- Others…



MODE: Object-based approach



Object identification

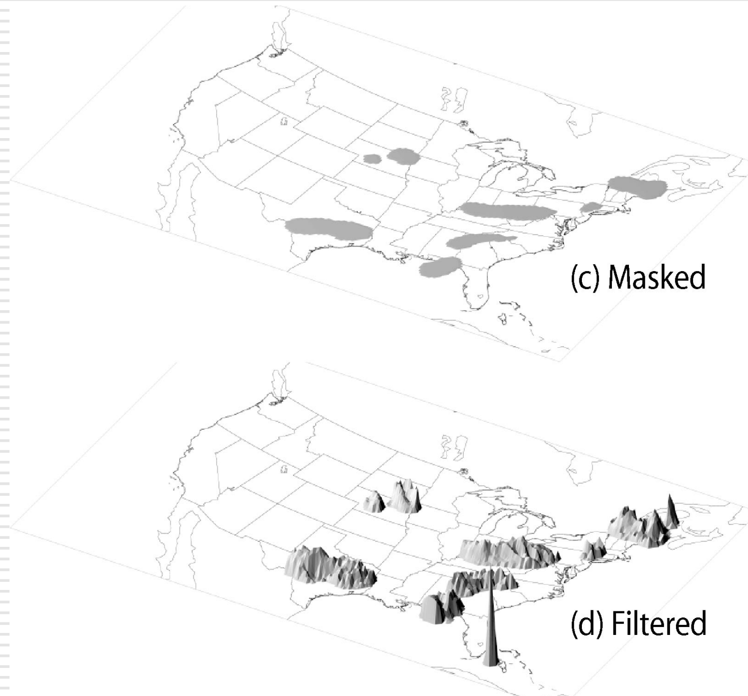


$$g(x,y) = \sum_{(u,v)\in G} \phi(u,v)\, f(x-u, y-v)$$

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

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

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



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

2 parameters:

1. Convolution radius
2. Threshold

Merging and Matching: *Fuzzy* Logic

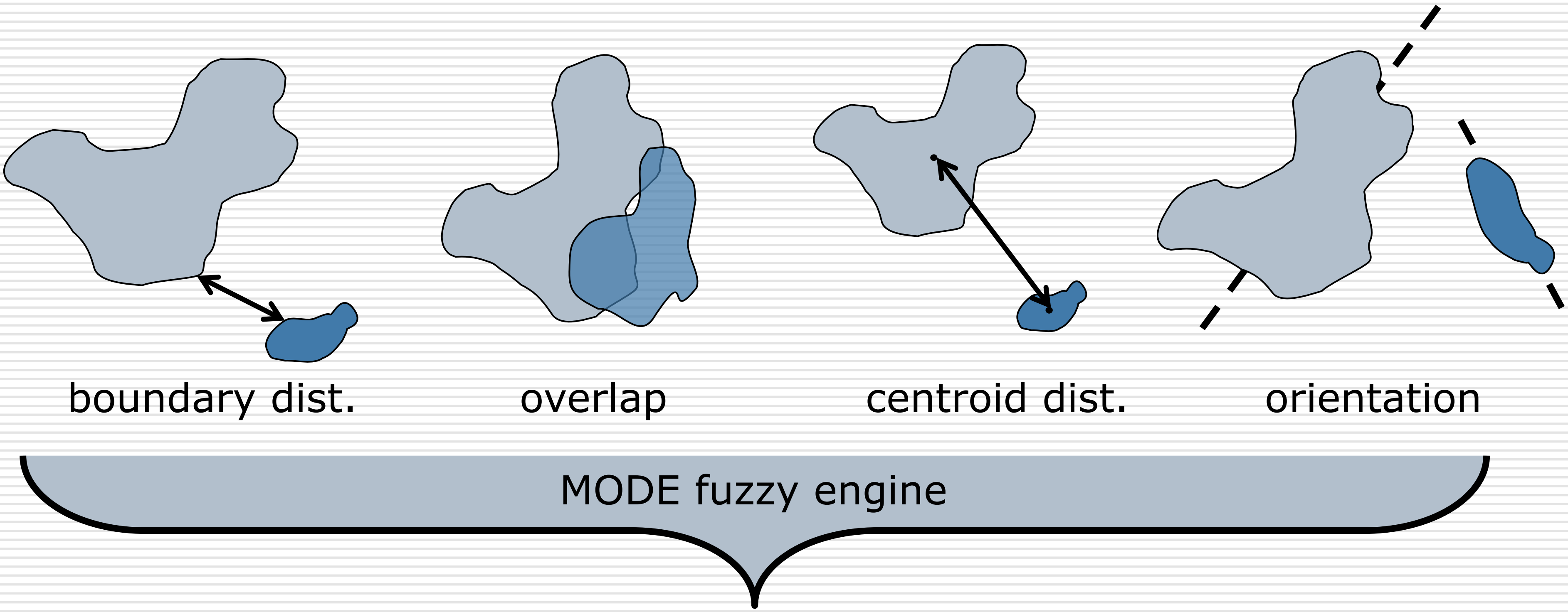
Attributes

Interest Maps

Confidence Maps

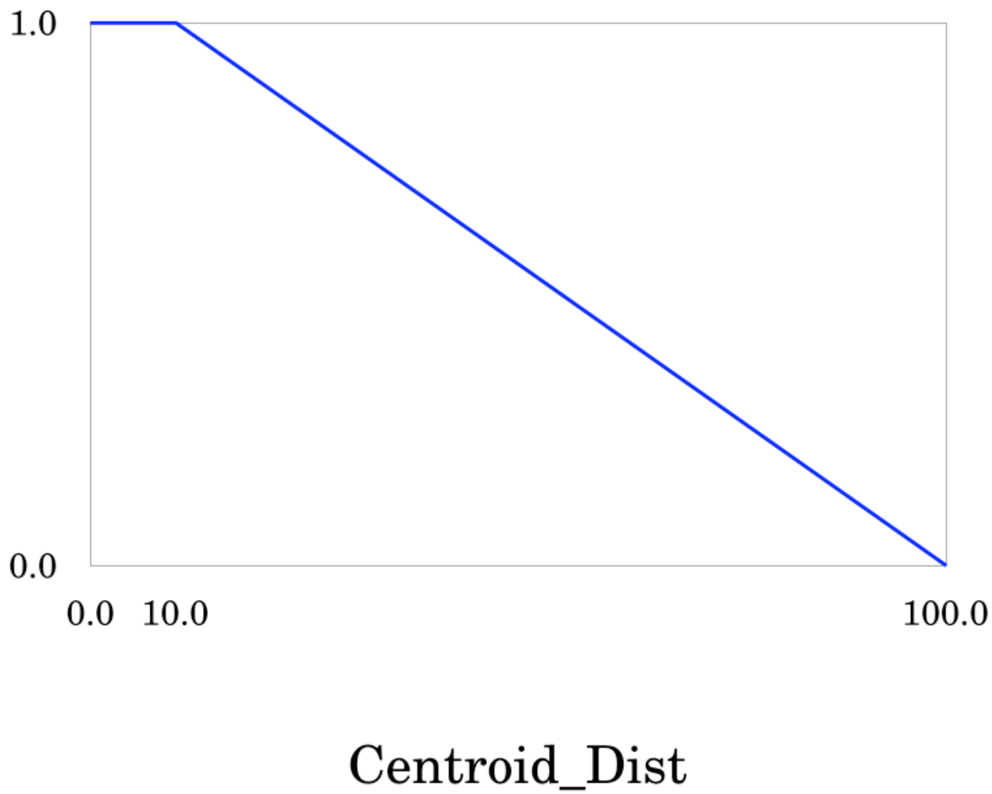
Weights

Fuzzy Logic: Attributes



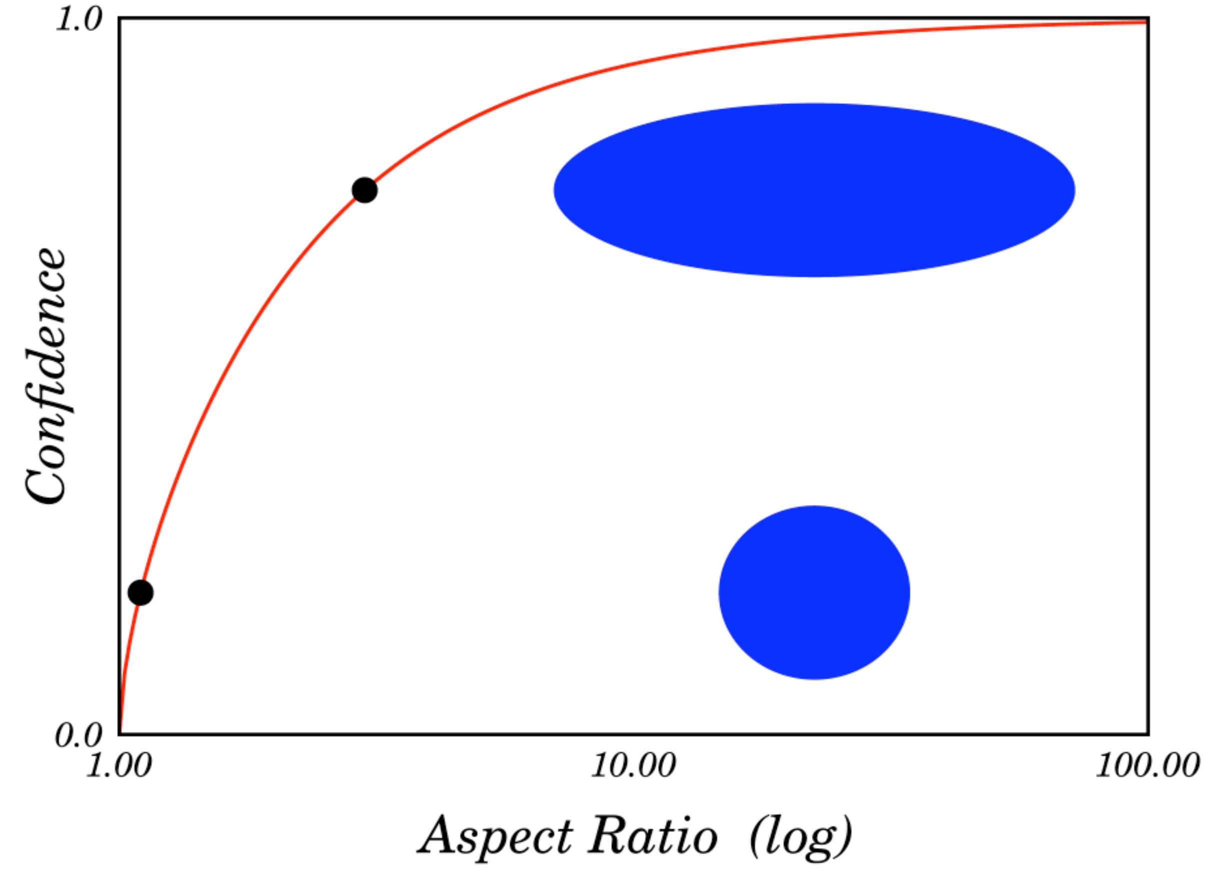
interest value for object pair
diagnostic output

Fuzzy Logic: Interest Maps



X	Y
0.0	1.0
10.0	1.0
100.0	0.0

Fuzzy Logic: Confidence Maps

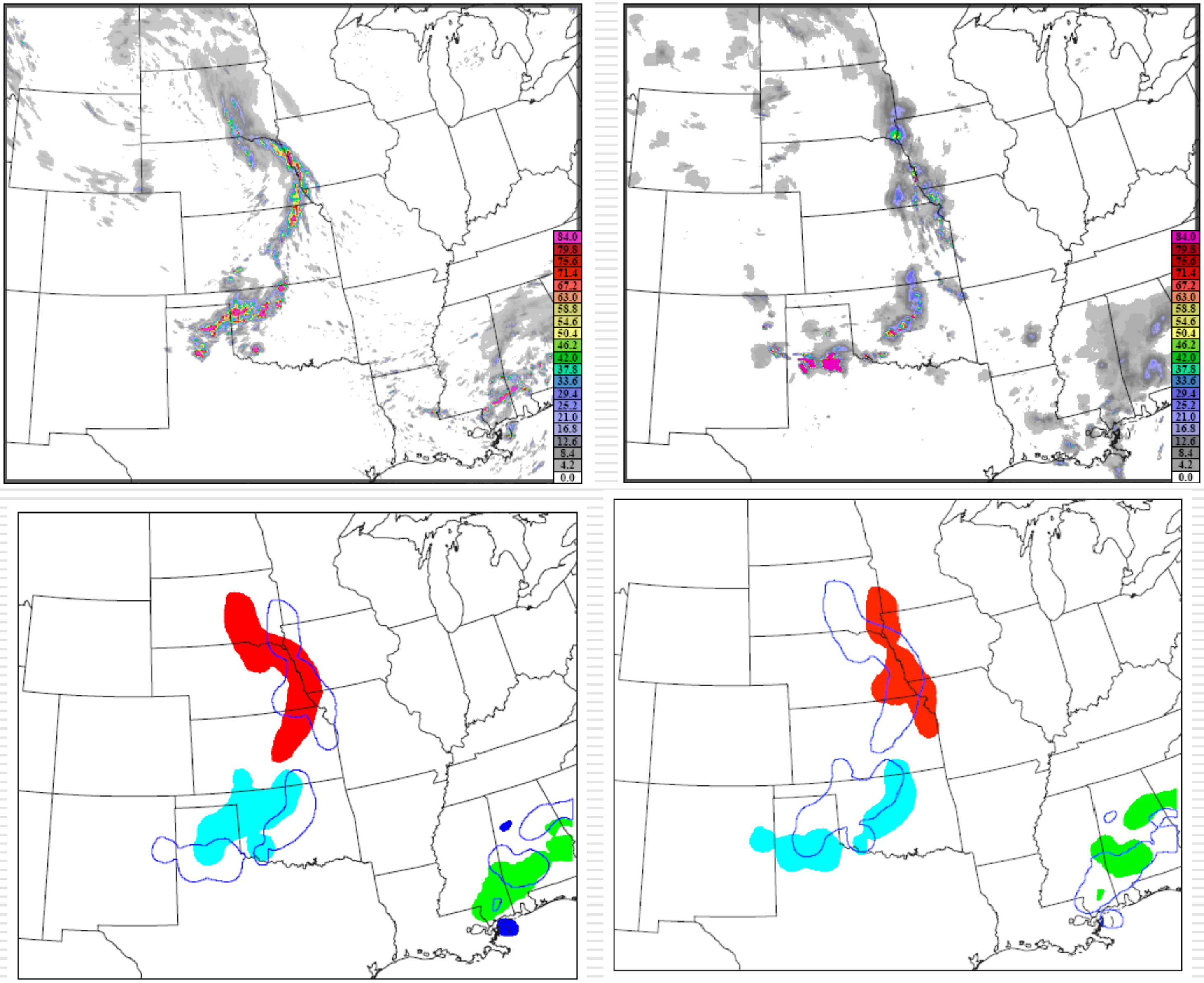


Total Interest

$$T(\alpha) = \frac{\sum_i w_i \, C_i(\alpha) \, I_i(\alpha_i)}{\sum_i w_i \, C_i(\alpha)}$$

Object-based example: 1 June 2005

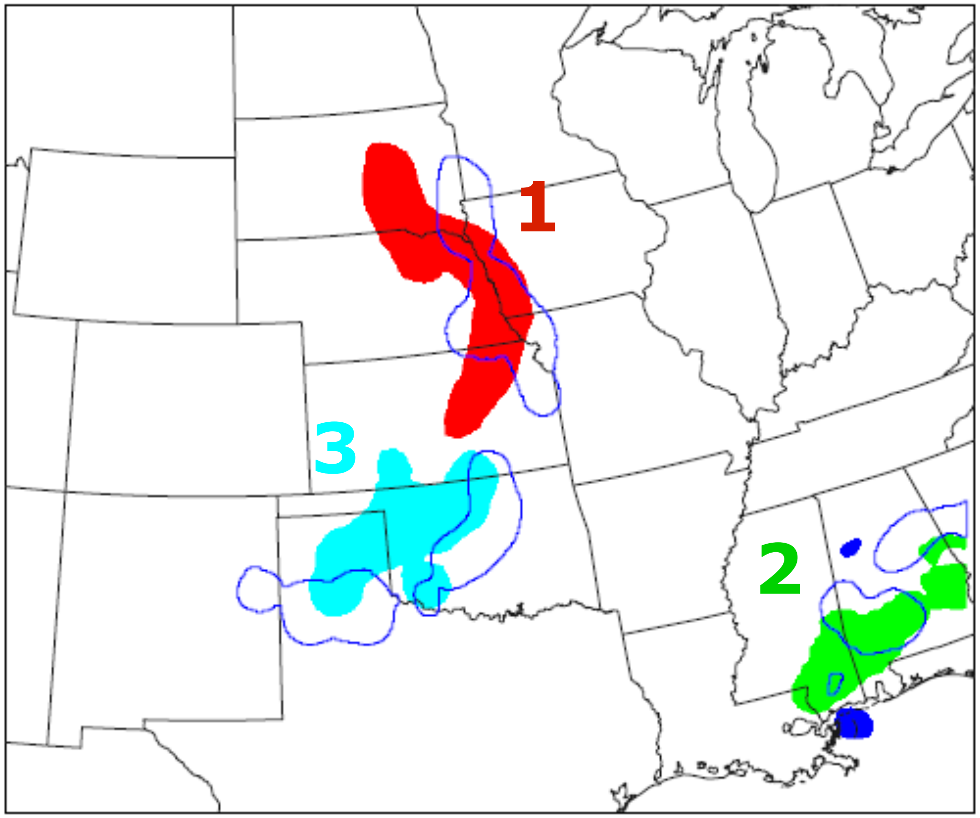
WRF ARW (24-h)



Stage II

Radius = 15 grid squares, Threshold = 0.05”

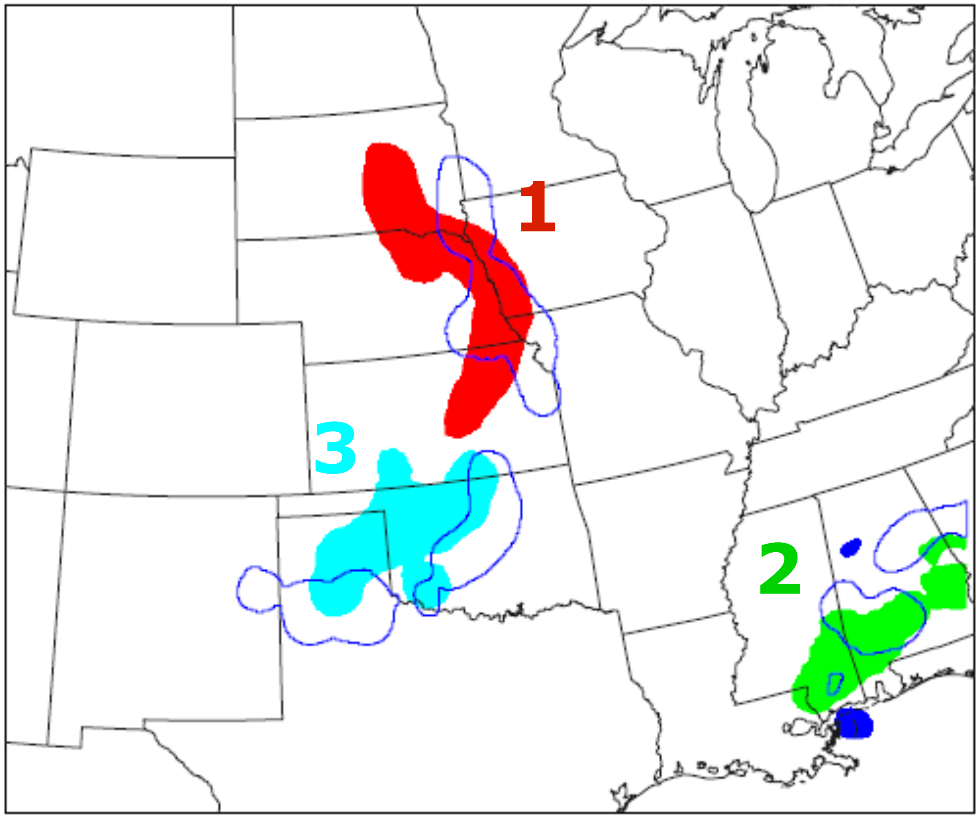
Object-based example 1 June 2006



WRF ARW-2 Objects with Stage II Objects overlaid

- Area ratios
  - (1) 1.3
  - (2) 1.2
  - (3) 1.1
  - All forecast areas were somewhat too large
- Location errors
  - (1) Too far West
  - (2) Too far South
  - (3) Too far North

Object-based example 1 June 2006



WRF ARW-2 Objects with
Stage II Objects overlaid

- MODE provides info about areas, displacement, intensity, etc.
- In contrast:

  POD = 0.40

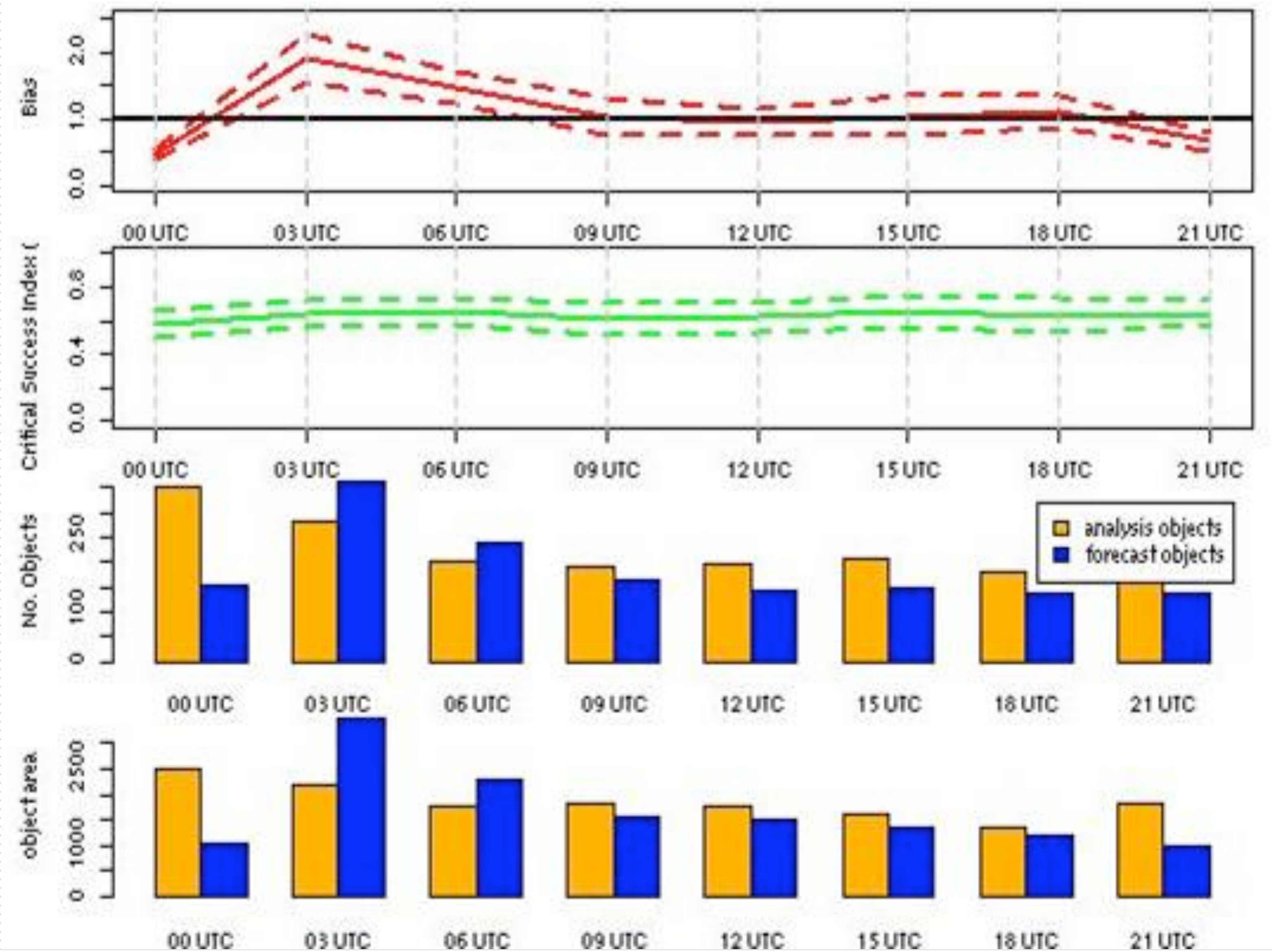
  FAR = 0.56

  CSI = 0.27

Applications of MODE

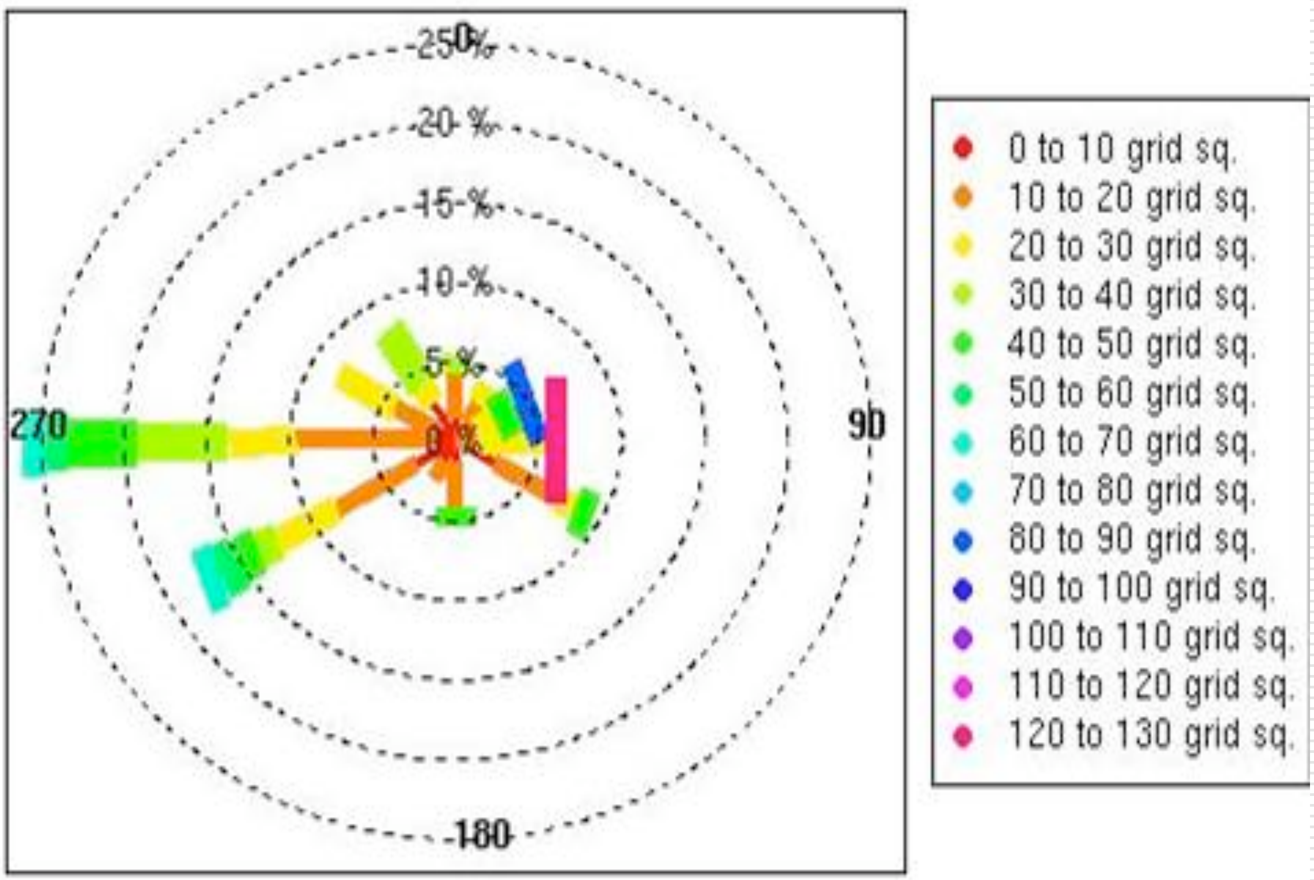
- Climatological summaries of object characteristics
- Evaluation of individual forecasting systems
  - Systematic errors
  - Matching capabilities (overall skill measure)
  - Model diagnostics
  - User-relevant information
  - Performance as a function of scale
- Comparison of forecasting systems
  - As above

Example summary statistics

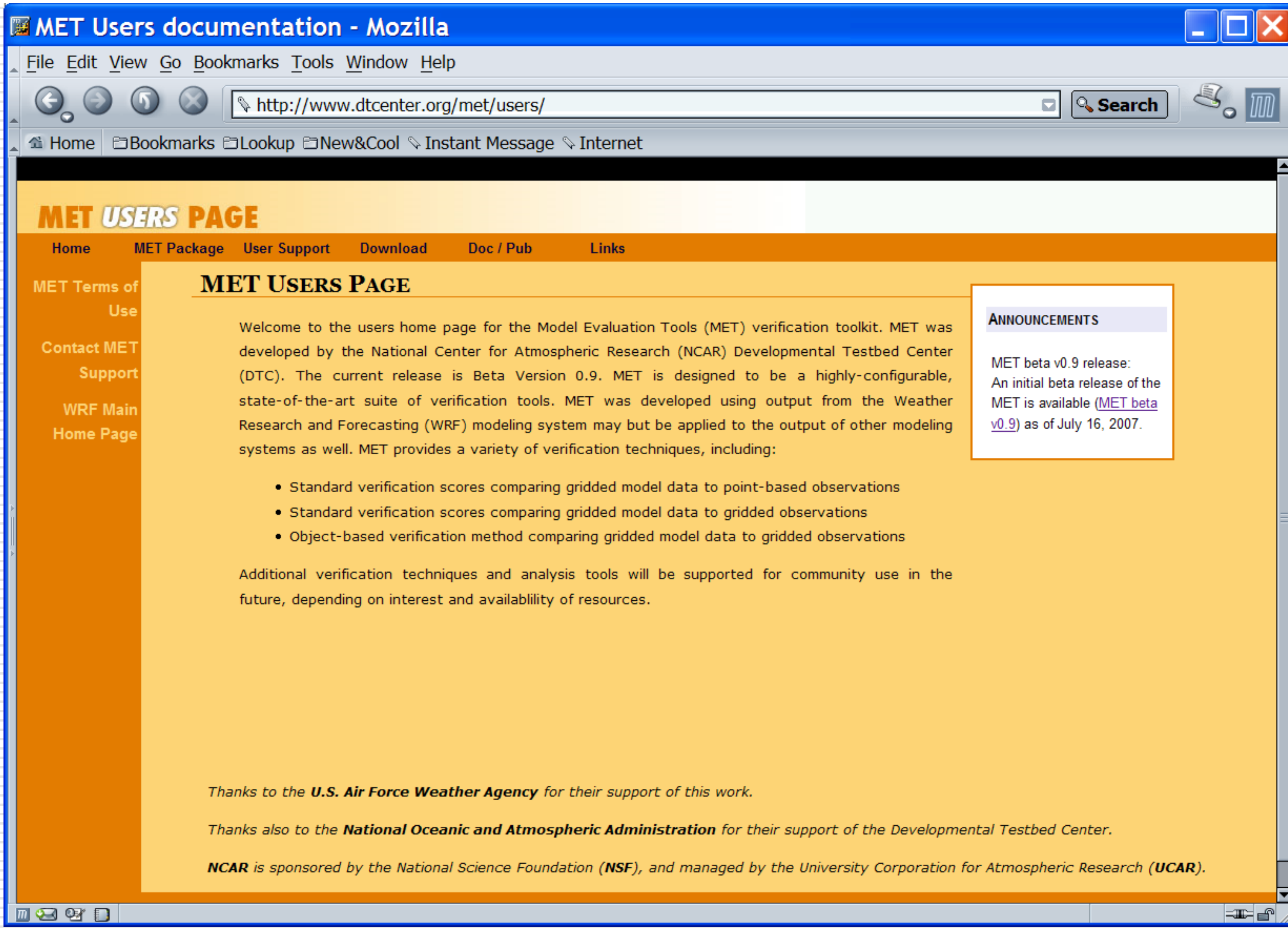


Example summary statistics

- MODE "Rose Plots"
- Displacement of matched forecast objects



MODE availability

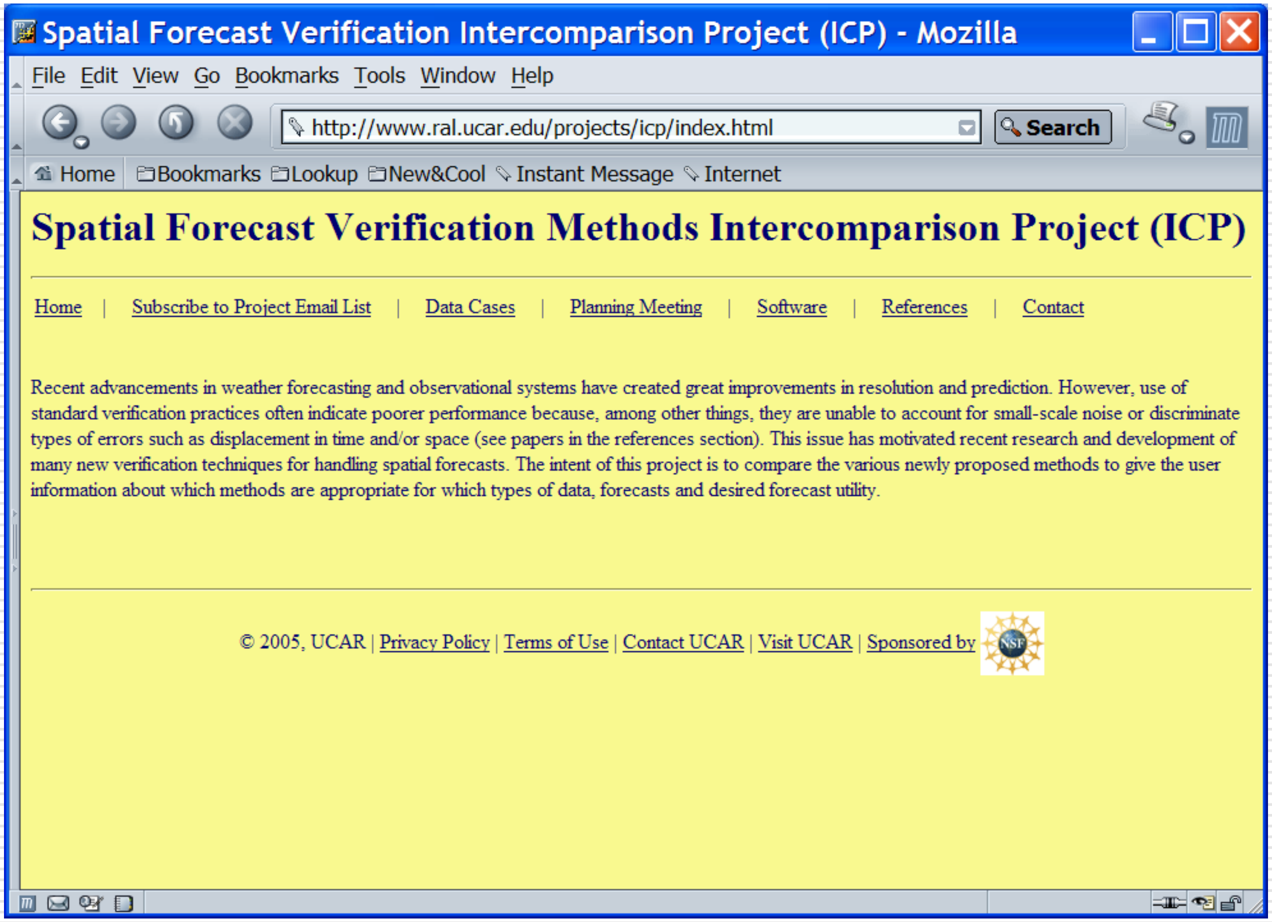


Available as part of the Model Evaluation Tools (MET)

http://www.dtcenter.org/met/users/

Intercomparison web page

- References
- Background
- Data and cases
- Software



http://www.ral.ucar.edu/projects/icp/

Acknowledgements

- **Co-authors and contributors**: David Ahijevych, Barbara G. Brown, Randy Bullock, Chris Davis, John Halley Gotway, Lacey Holland
- National Center for Atmospheric Research (NCAR)
- Boulder, Colorado
- October 2007

References: see ICP website

http://www.rap.ucar.edu/projects/icp/references.html