New Verification Approaches for Evaluating TAMDAR Impacts on Mesoscale Quantitative Precipitation Forecasts

Barbara Brown

RAL/NCAR

bgb@ucar.edu

Collaborators: C. Davis, R. Bullock, Yubao Liu, Wei Yu, RAL Verification Group

Overview

- Issue:

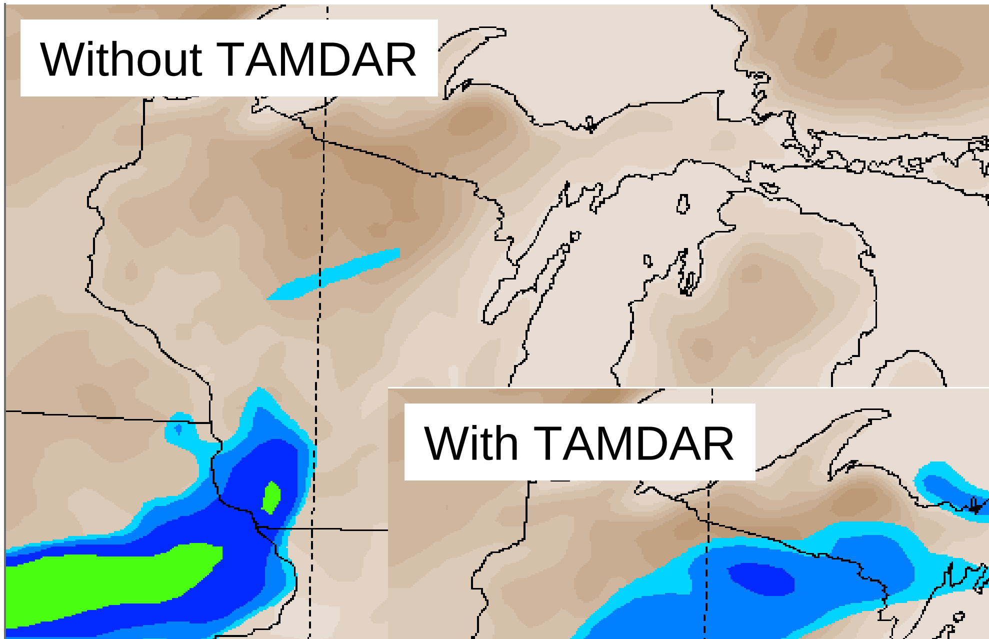
  Traditional forecast verification approaches provide limited information about
  - *Forecast quality*
  - *Differences between one forecast system and another*
- Goal:

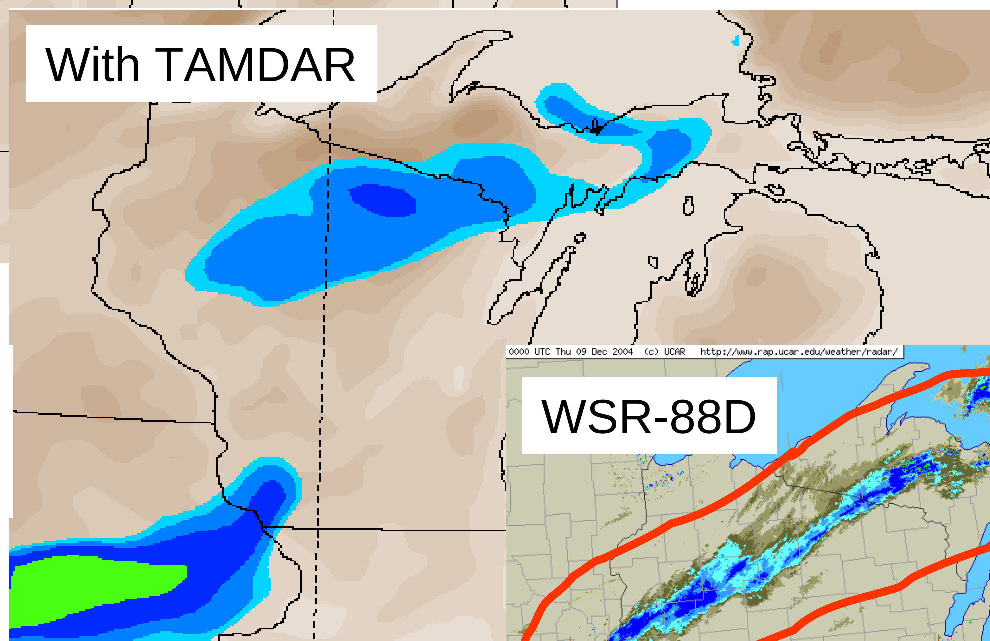
  Apply new *diagnostic* approaches that provide more meaningful information about differences between forecasts and forecasting systems.

Weak snowbands (1)

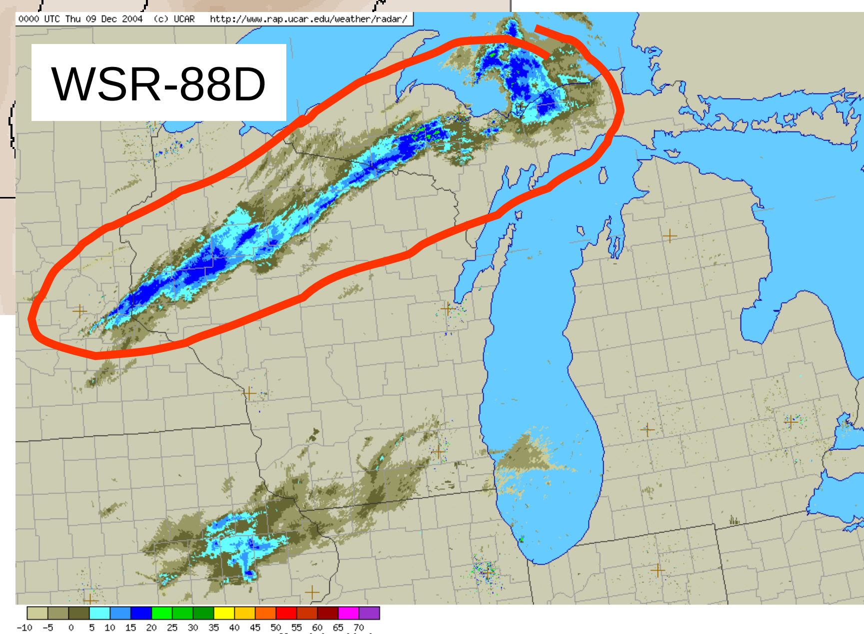
00Z, Dec. 09, 2004

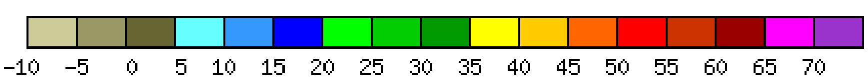
Radar reflectivity





RTFDDA
1h forecasts





Outline

- Motivation
- Object-based approach
- Examples
- Future work

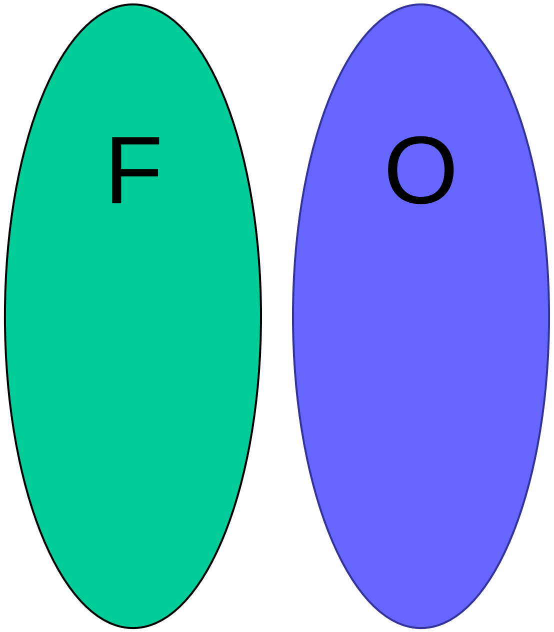
What are the purposes of verification?

- Contribute to forecast development and improvement
- Calibration (i.e., probabilities)
- Forecast comparisons
- Selection of model and transfer to operations
- Monitor forecasting capabilities
- Provide credibility
- Provide information needed by
  - *Human decision makers*
  - *Decision support systems*

What are the purposes of verification?

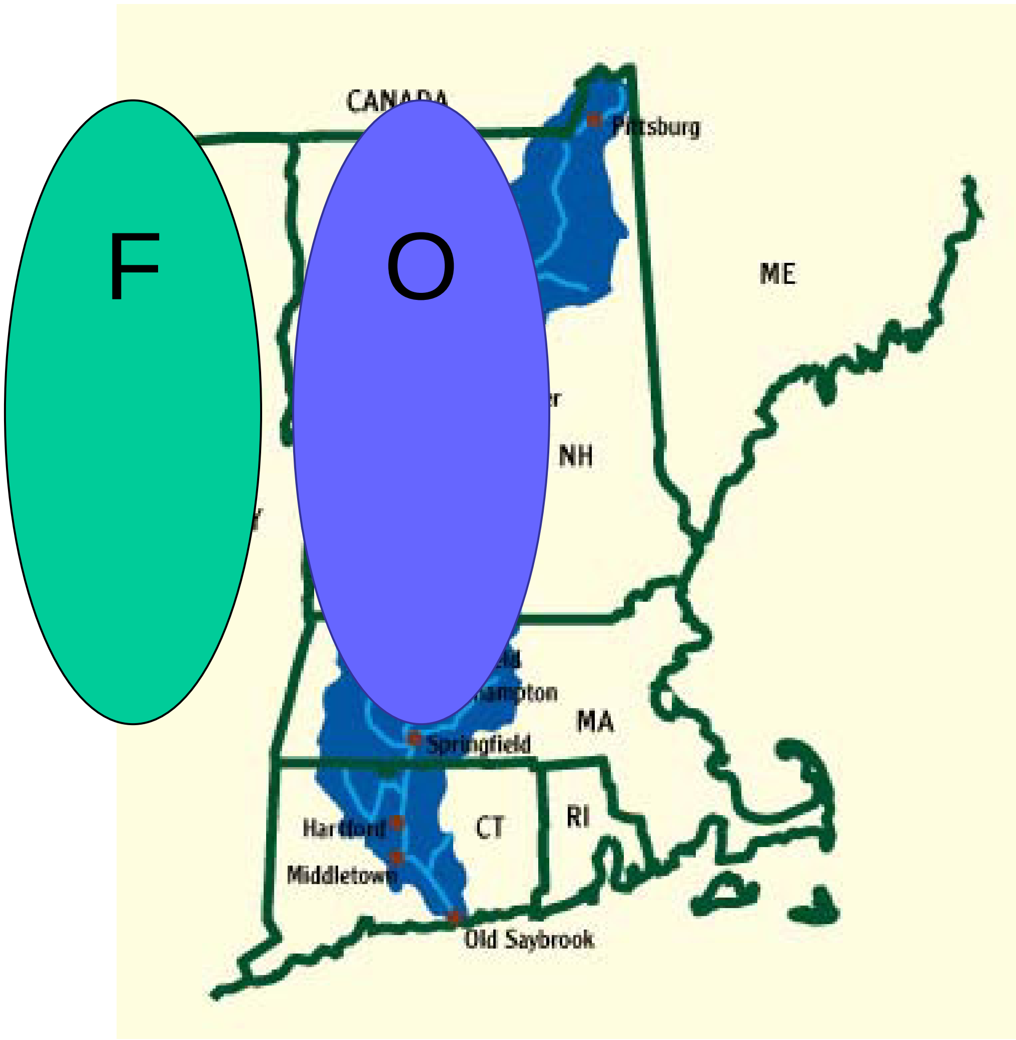
- Contribute to forecast development and improvement
- Calibration (i.e., probabilities)
- Forecast comparisons
- Selection of model and transfer to operations
- Monitor forecasting capabilities
- Provide credibility
- Provide information needed by
  - *Human decision makers*
  - *Decision support systems*

Good forecast or Bad forecast?

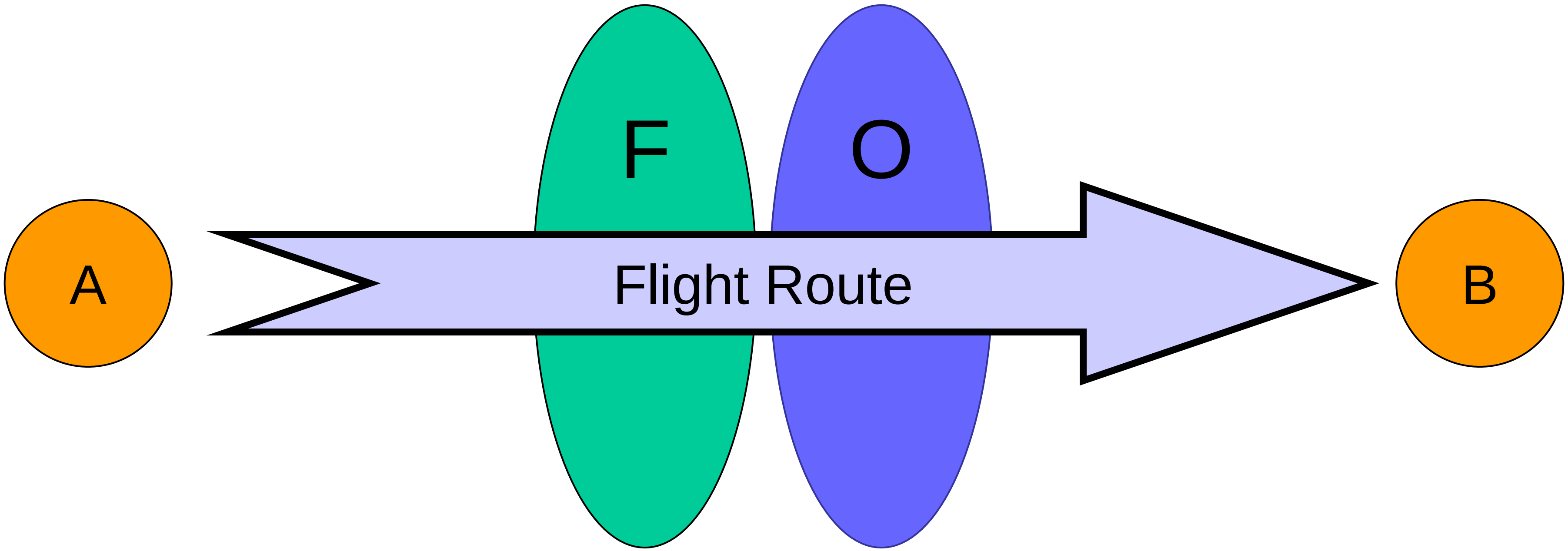


Good forecast or Bad forecast?

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



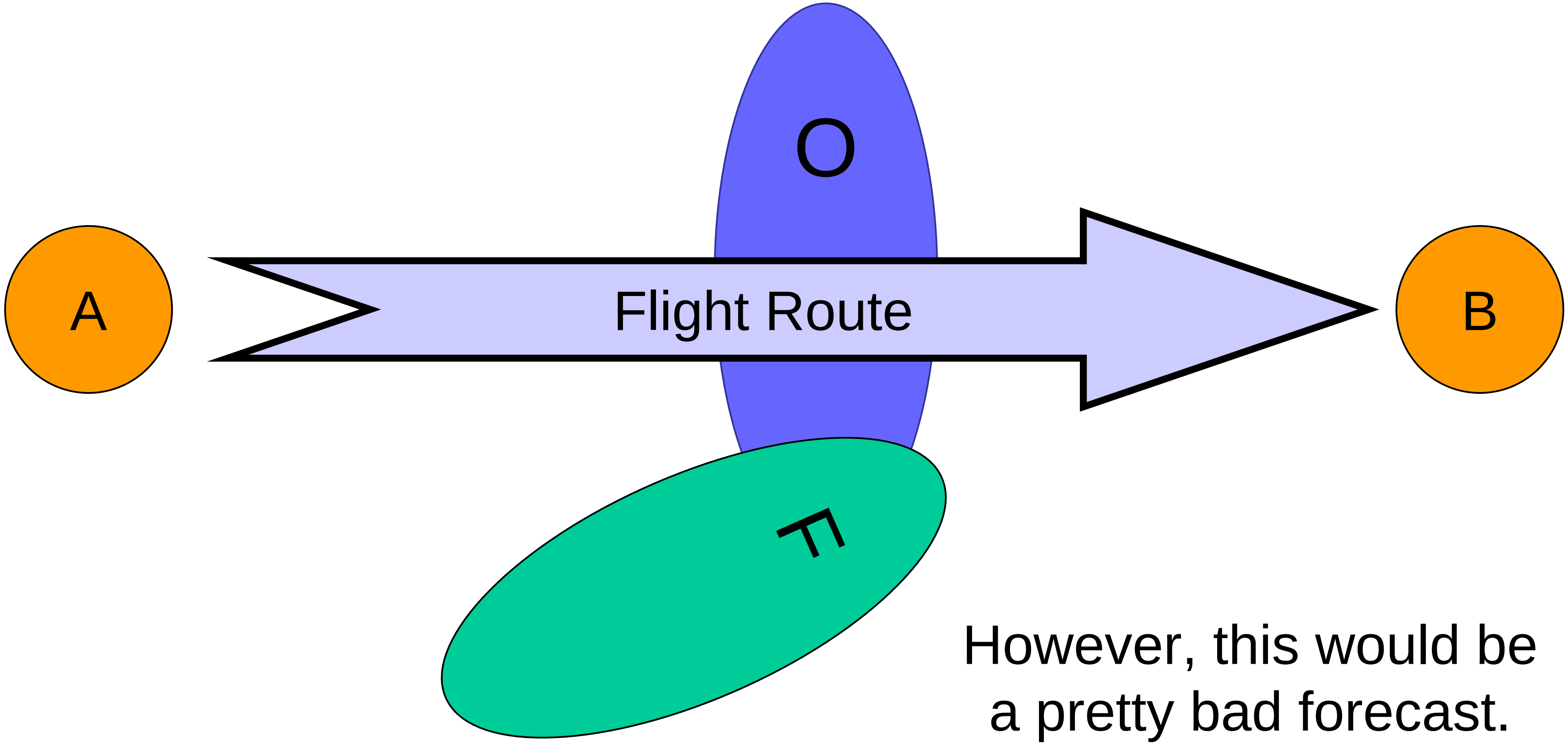
Good forecast or Bad forecast?



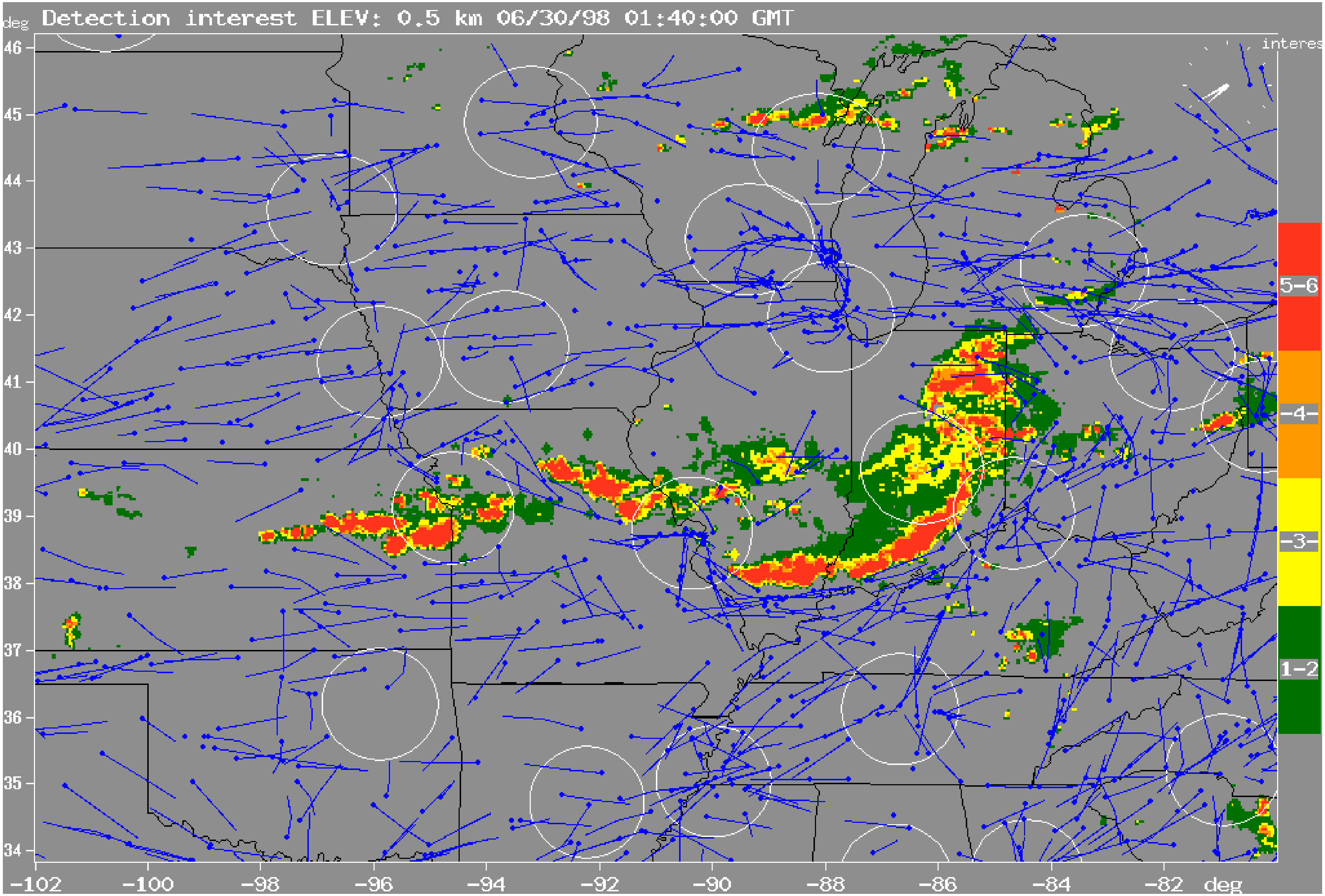
But if I'm an aviation traffic strategic planner…

It might be a pretty good forecast

Good forecast or Bad forecast?



However, this would be a pretty bad forecast.



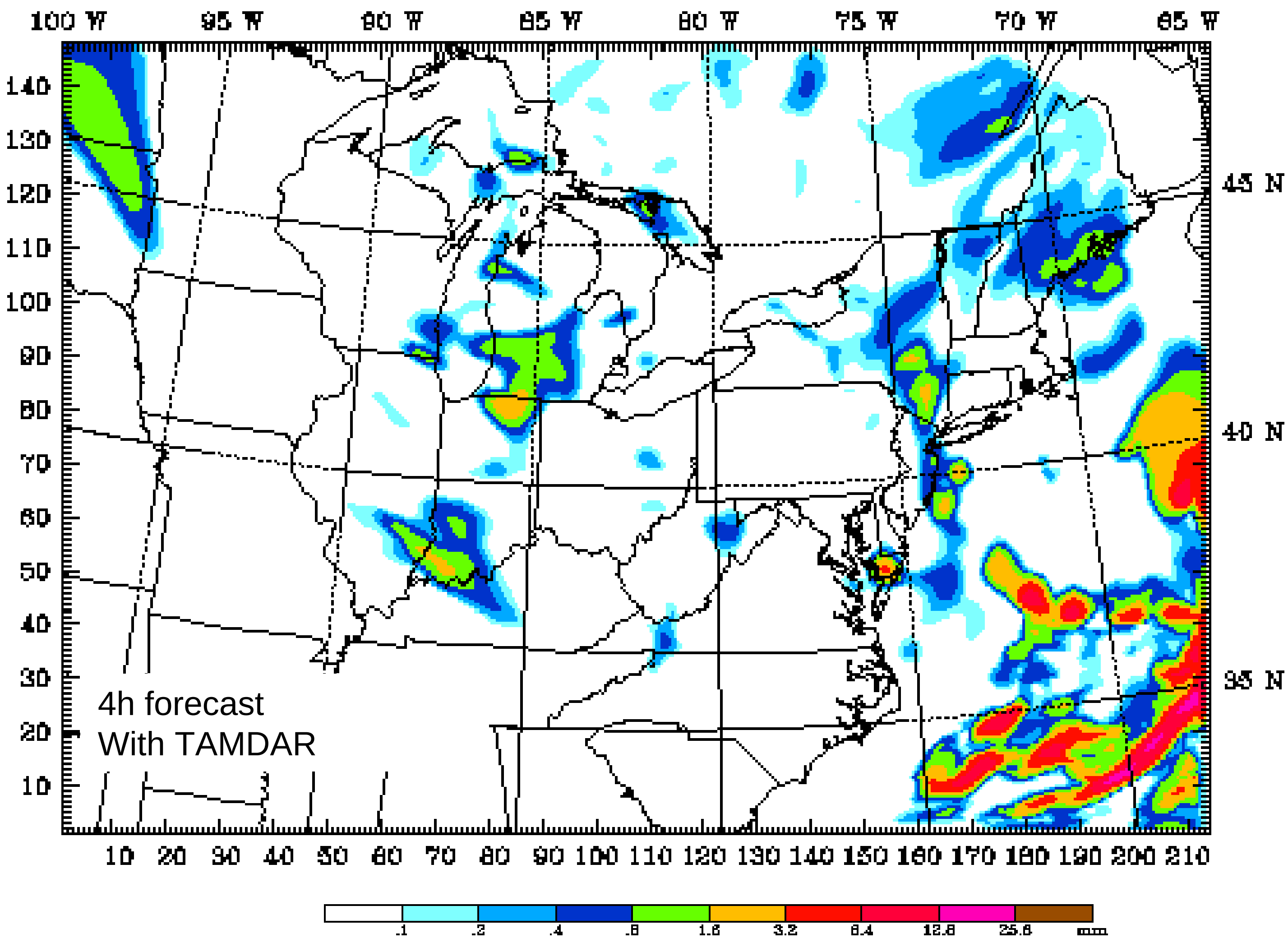
deg
Detection interest ELEV: 0.5 km 06/30/98 01:40:00 GMT
interes
46
45
44
43
42
41
40
39
38
37
36
35
34
5-6
4
3
1-2
-102
-100
-98
-96
-94
-92
-90
-88
-86
-84
-82
deg

Focus: Spatial forecasts

- Precipitation
- Convection
- (Extensions to other variables possible: e.g., clouds, icing)
- Assume forecast and observations can be represented on a grid

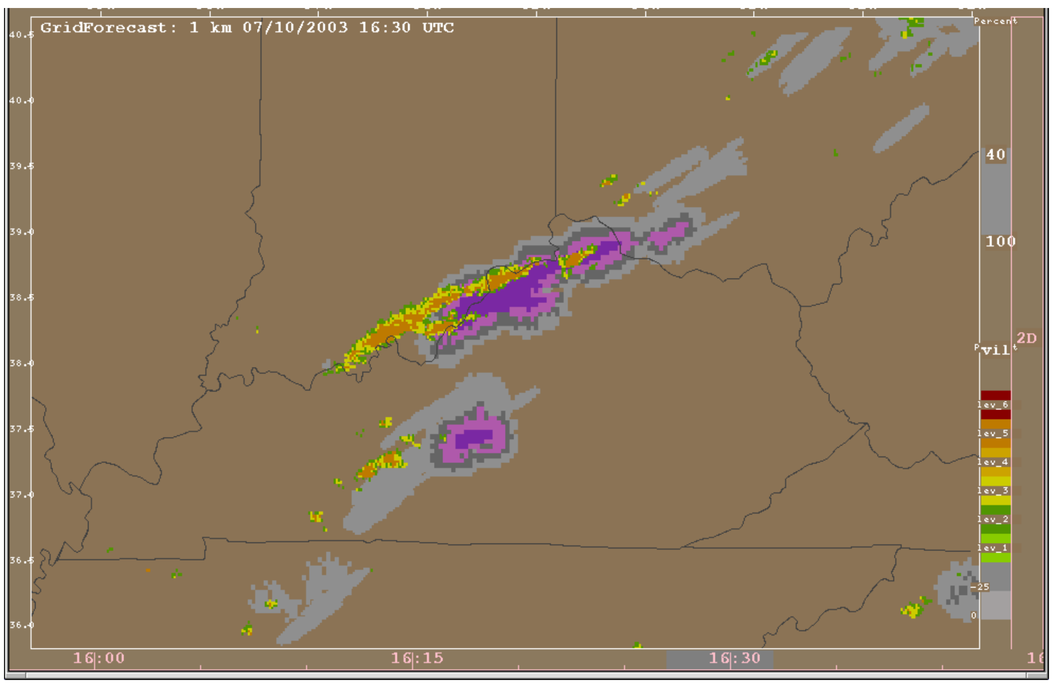
Example: Precipitation

06Z,
12Mar05,
1-h accum.



Example: Convection (extrapolated)

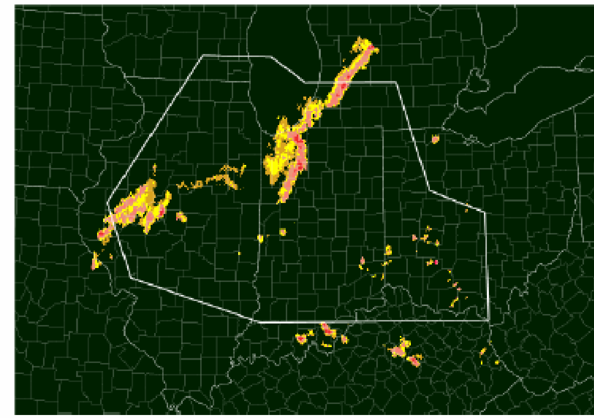
National Convective Weather Forecast (NCWF)



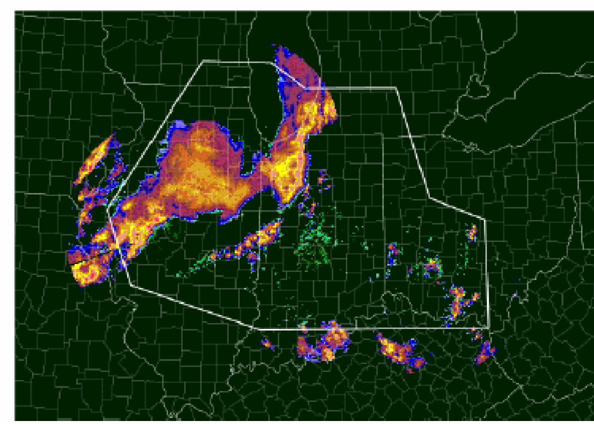
Mueller et al., NCAR

Autonowcaster

Fcst

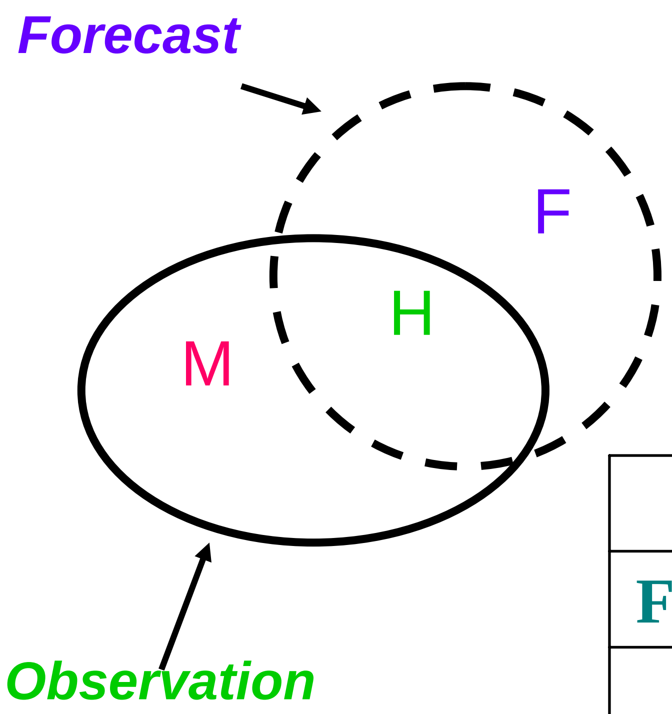


Obs



Roberts et al., NCAR

"Traditional" Verification Approach (Yes/No forecasts)



H = Hits

M = Misses

F = False Alarms

	Observations	
Forecasts	**x=1**	**x=0**
f=1	**H** = p_{11}	**F** = p_{10}
f=0	**M** = p_{01}	**CR** = p_{00}

Verification Contingency Table, Example Summary Measures, and Scores

	Observations	
Forecasts	**x=1**	**x=0**
f=1	**H** = p_{11}	**F** = p_{10}
f=0	**M** = p_{01}	**CR** = p_{00}

POD = H / (H + M)

= Prob of Detection

= Pr(f=1 | x=1)

= proportion of "Yes" area correctly forecast to be "Yes"

POFD = CR / (F+CR)

= Prob of False Detection

= Pr(f=0 | x=0)

= proportion of "No" area that was correctly forecast to be "No"

FAR = F / (H + F)

= False Alarm Ratio

= Pr(x=0 | f=1)

= proportion of "Yes" forecast area that was incorrect

Bias = (F + H) / (M + H)

= Pr(f=1) / Pr(x=1)

= Ratio of area of "Yes" forecast to "Yes" observed

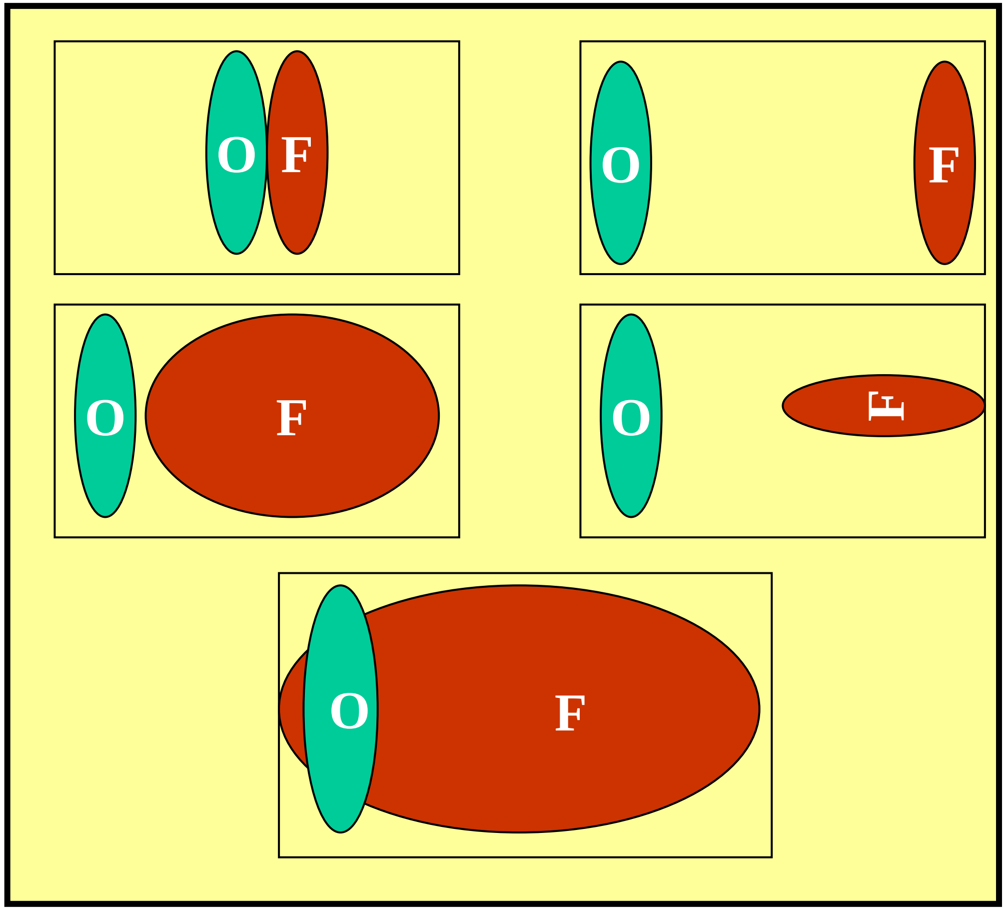
Critical Success Index (CSI) = "Threat Score"

= H / (H + M + F)

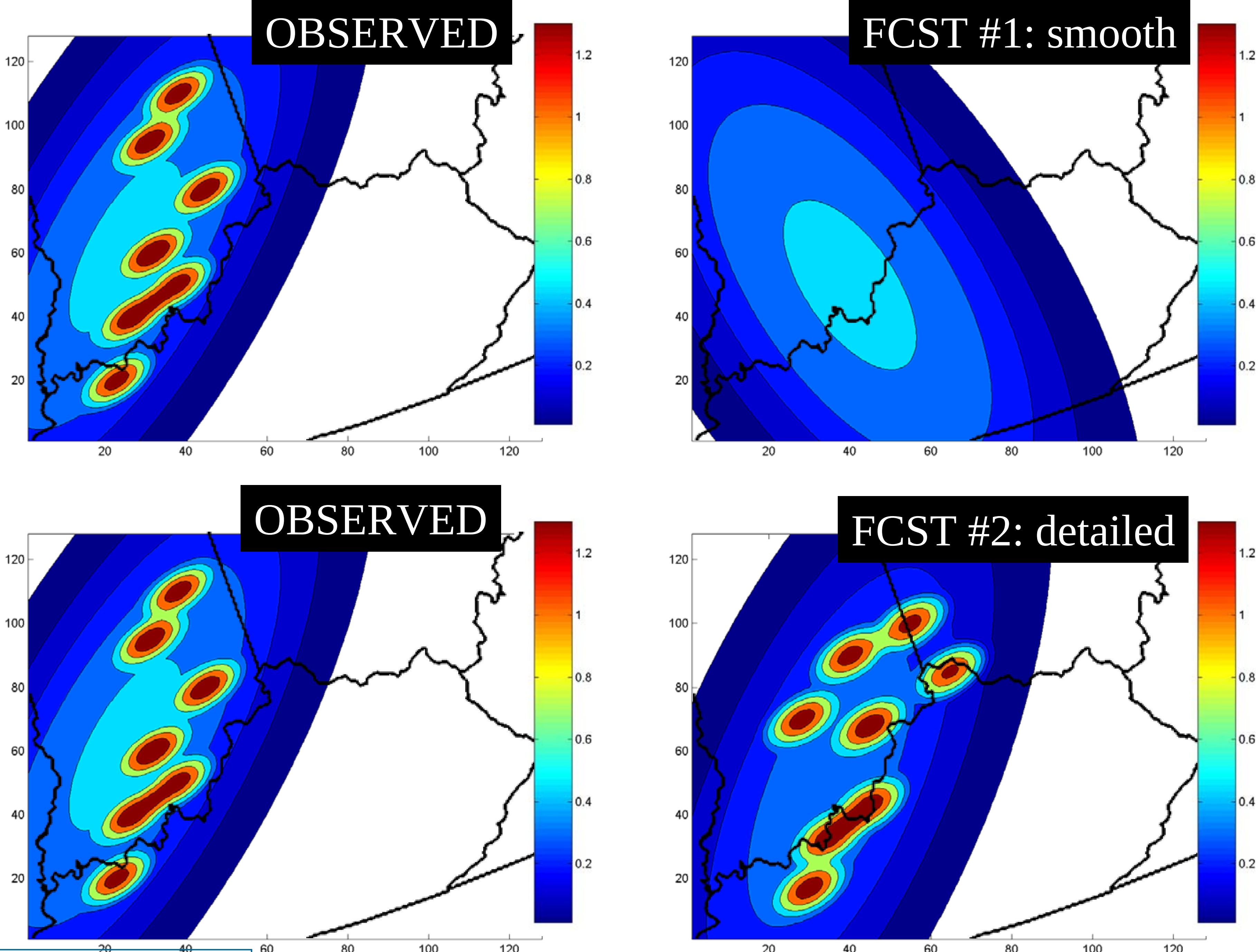
= $p_{11} / (p_{11} + p_{01} + p_{10})$

Other skill indices (compare accuracy of forecast to some non-intelligent standard)

Example



- First four forecasts have POD=0; FAR=1; CSI=0
  - *i.e., all are equally "BAD"*
- Fifth forecast has POD>0, FAR<1, CSI>1
- Traditional verification approach identifies "worst" forecast as the "best"



From Baldwin 2002

“Measures-oriented” approach to verifying these forecasts

From Baldwin 2002

Verification Measure	Forecast #1 (smooth)	Forecast #2 (detailed)
Mean absolute error	*0.157*	0.159
RMS error	*0.254*	0.309
Bias	0.98	0.98
Threat score (>0.45)	*0.214*	0.161
Equitable threat score (>0.45)	*0.170*	0.102

The goal: Diagnostic evaluation approaches

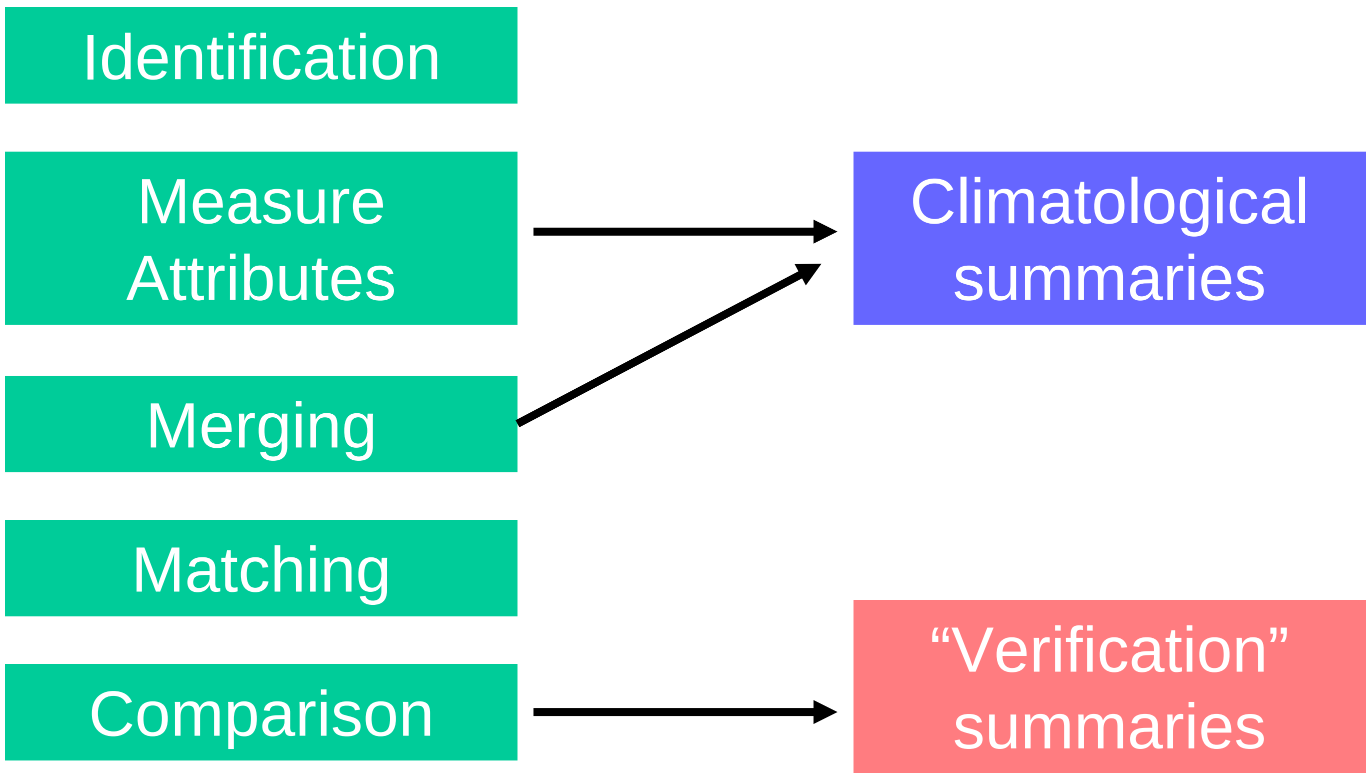
- Identify and evaluate meaningful attributes of the forecasts
  - *Example questions: What is the typical location error? Size error? Intensity error?*
- Provide detailed information about forecast quality

  Examples:
  - *What went wrong? What went right?*
  - *How can the forecast be improved?*
  - *How much uncertainty is there in particular attributes?*
  - *In what respects do 2 forecasts differ from each other, and in what ways is one better than the other?*

Alternative diagnostic approaches

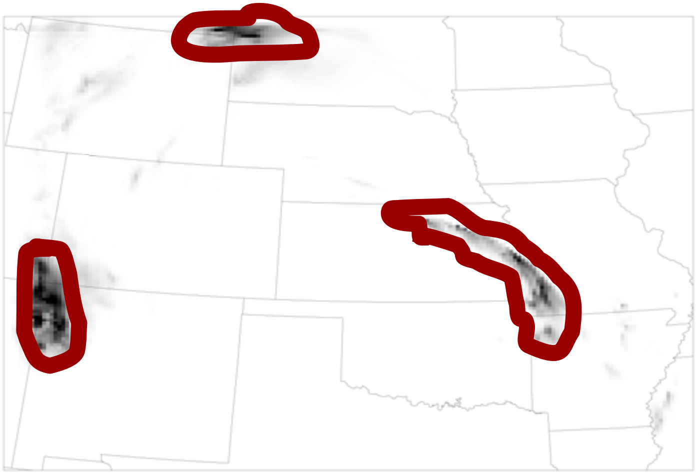
- Practically perfect approach
- Scale-separation approaches
- Composite approaches
- Entity-based verification (Ebert and McBride)
- Object-based verification
  - *Directly aims to meet the objectives we've defined*

Object-based approach

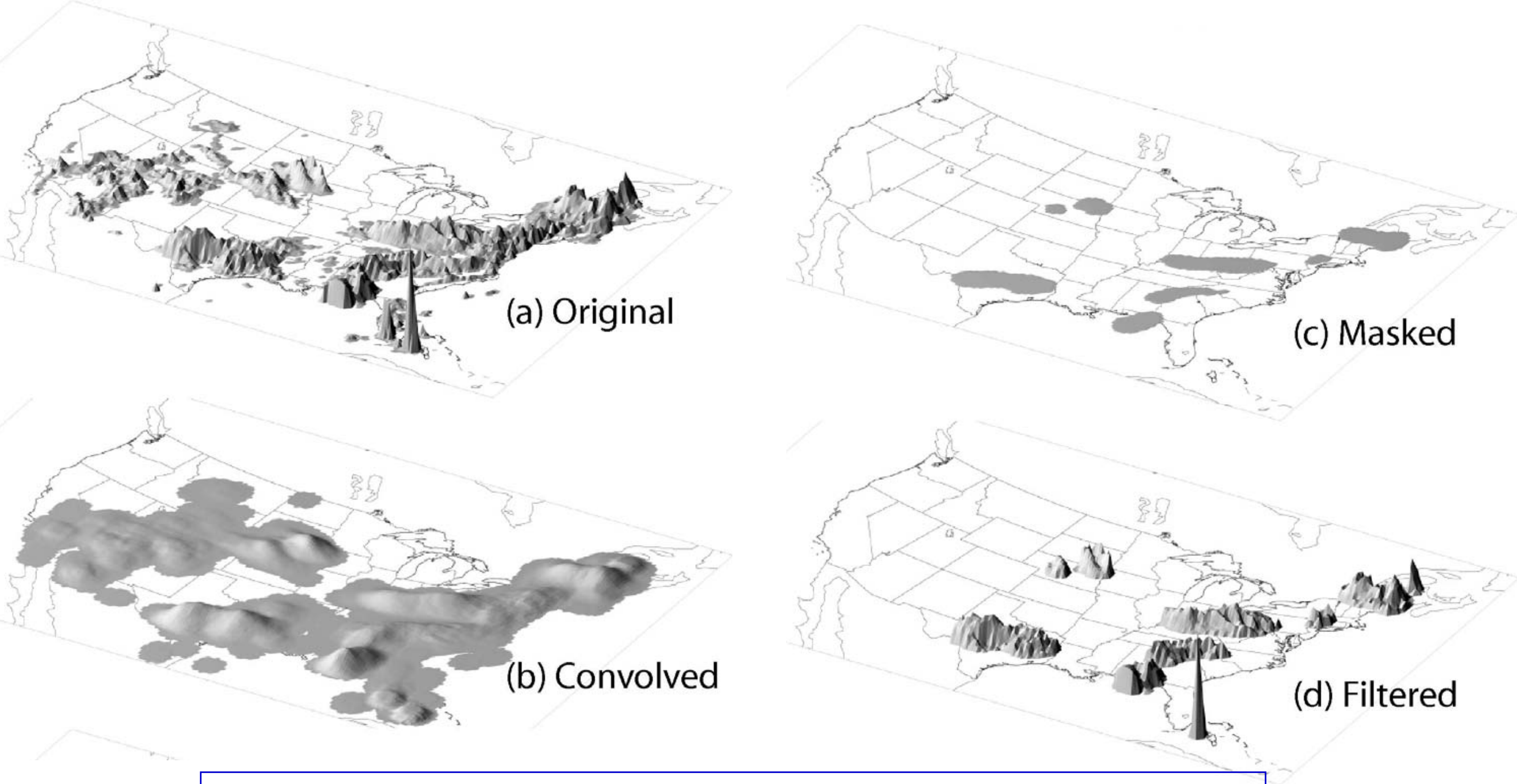


Basis of object-based approach

Objectively identify meaningful forecast and observed objects



Object identification



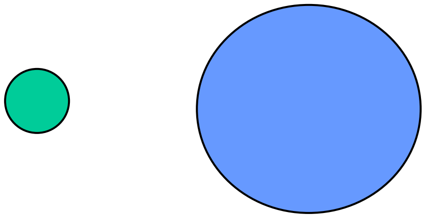
Parameters: (i) Convolution radius (Step b);

(b) Threshold (Step c)

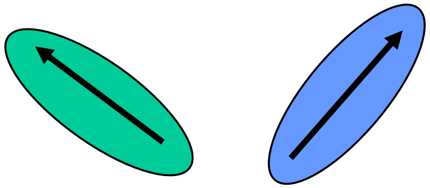
Merging and Matching: Fuzzy logic

Identify and measure meaningful attributes describing relevant characteristics of objects

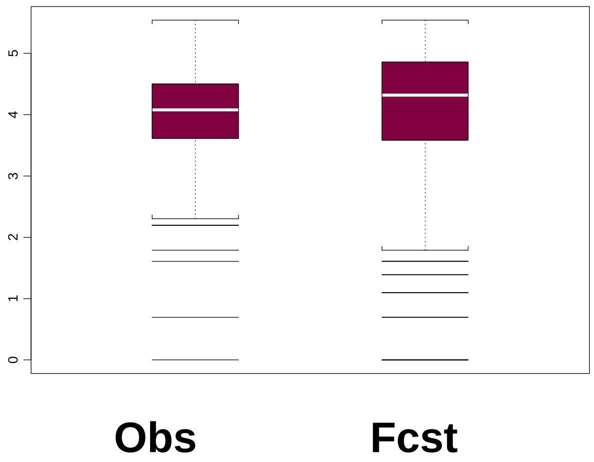
Size



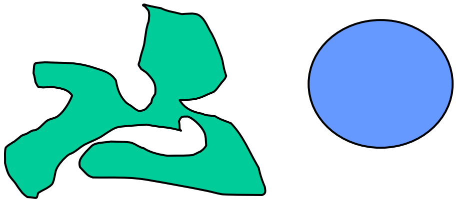
Orientation



Intensity



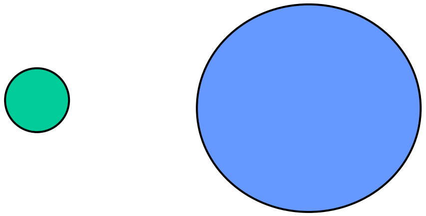
“Ugliness”



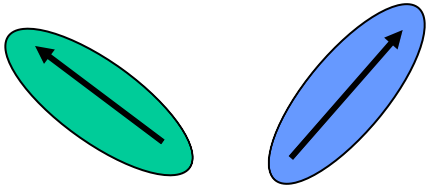
Merging and Matching: Fuzzy logic

Compare attributes for pairs of forecast and observed objects

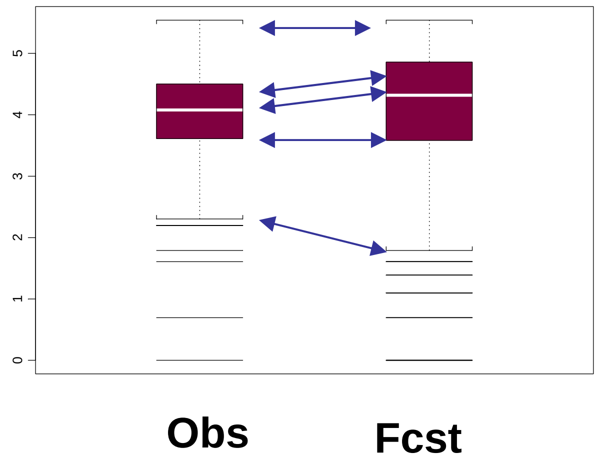
Size



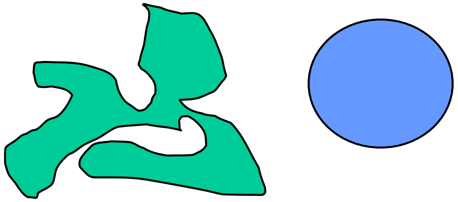
Orientation



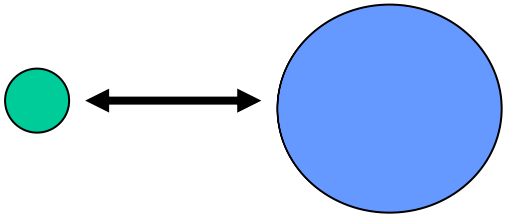
Intensity



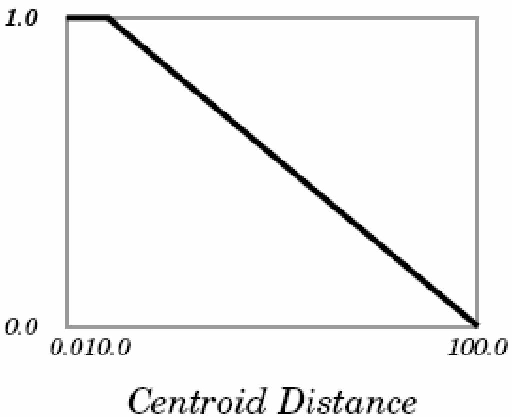
“Ugliness”

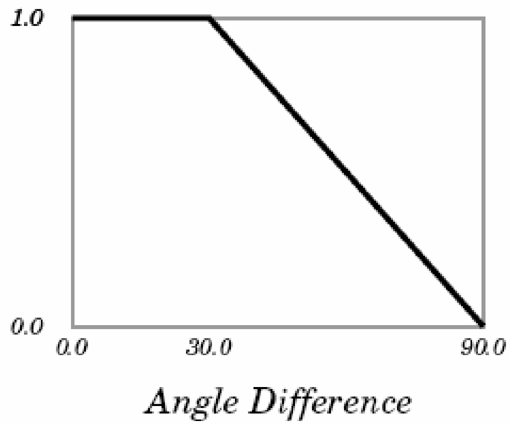


Displace-
ment

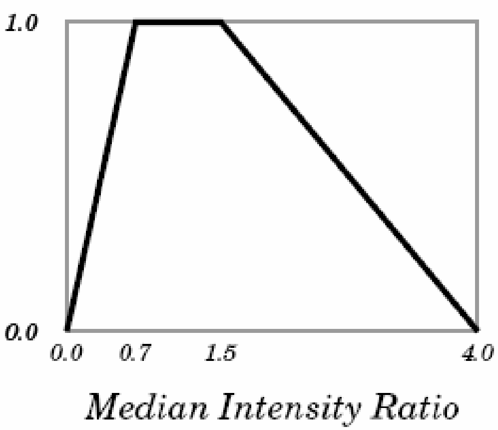


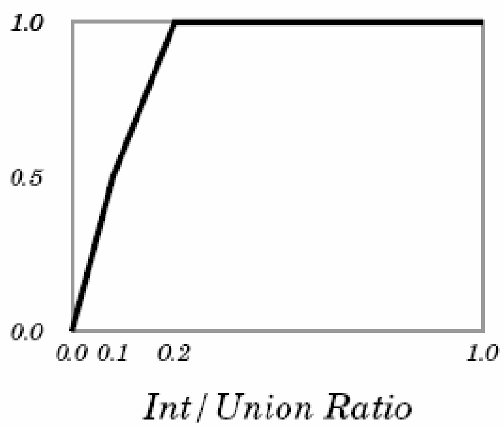
Merging and Matching: Fuzzy logic

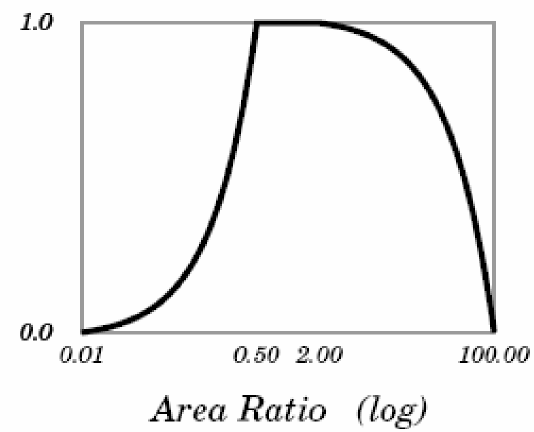




Compute interest values for pairs of forecast and observed objects







Compute total interest

Compute "Total Interest" for all pairs of forecast and observed objects.

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

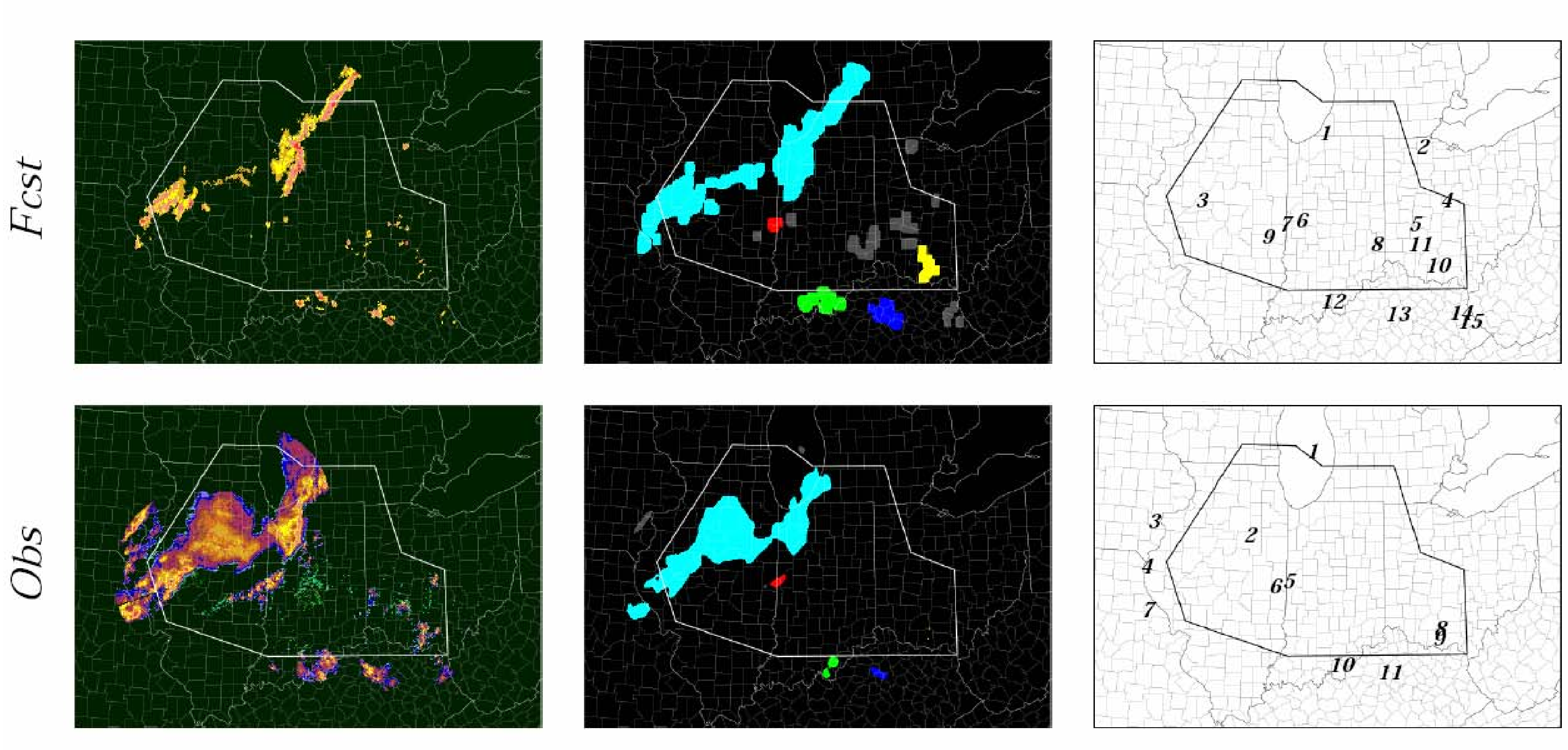
<u>Initial weights</u>

- Centroid distance: 1
- Angle difference: 0.1
- Median intensity ratio: 0.1
- Area ratio: 0.1
- Intersection/Union: 0.1

Apply threshold to Total Interest to determine merges/matches.

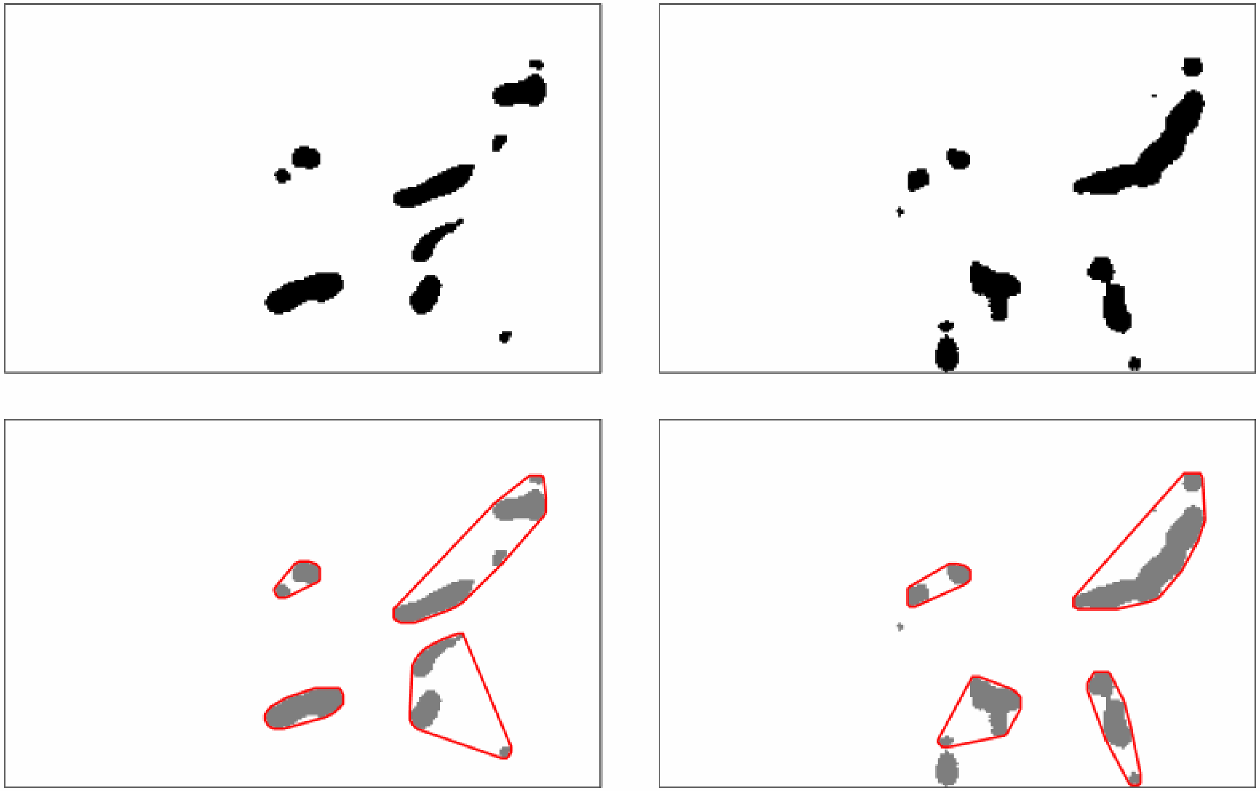
Initial threshold: 0.7

Example: Convective Nowcasts



Example: Precipitation

Forecasts | **Observations**

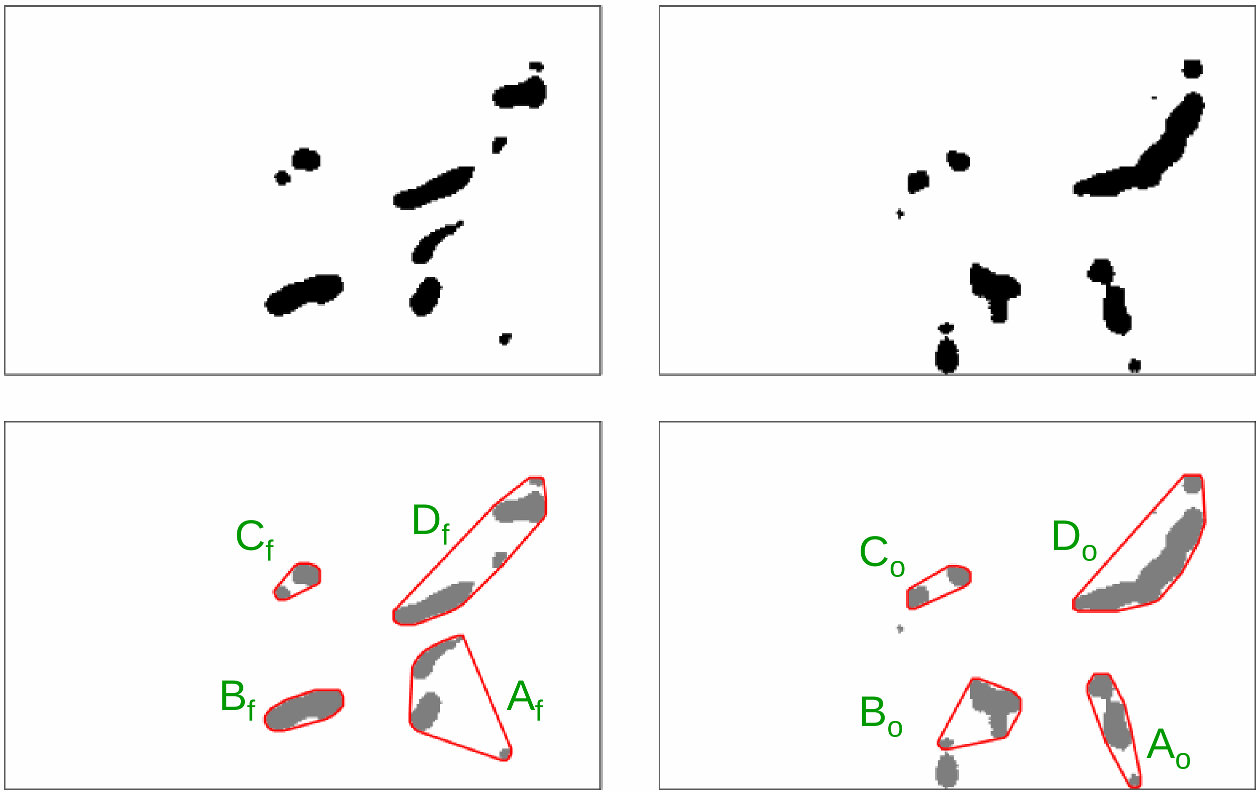


Objects identified

Objects merged

Object-based approach

Forecasts | **Observations**

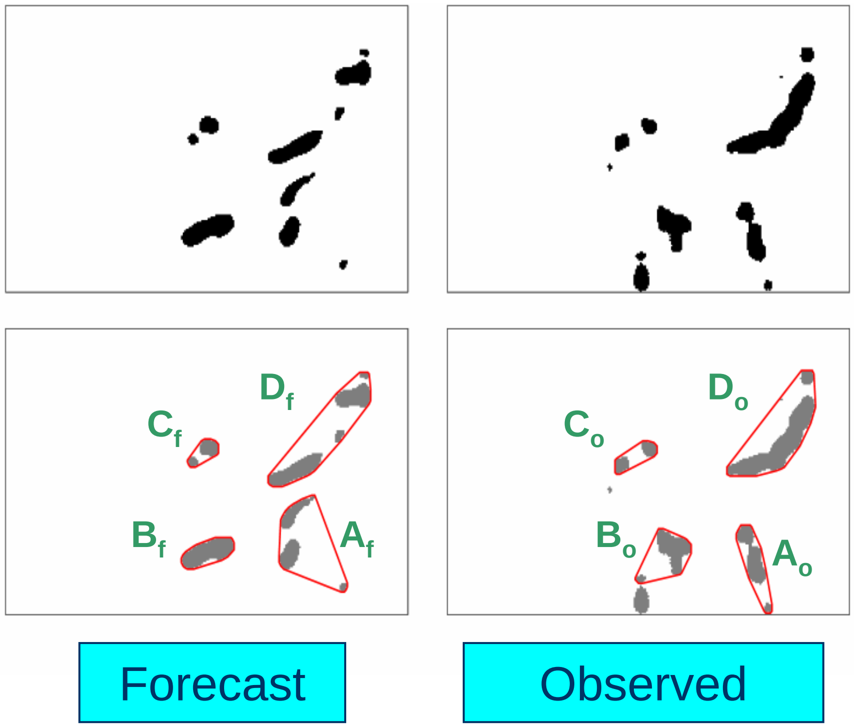


Objects identified

Objects merged

Objects matched

Gridded forecast example: Summary

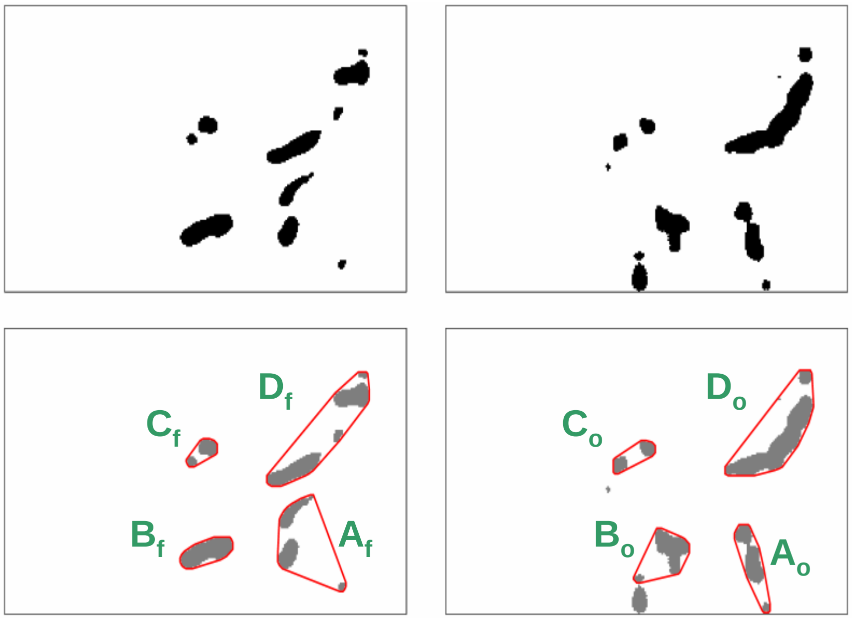


POD = 0.27

FAR = 0.75

CSI = 0.34

Gridded forecast example: Summary



Forecast | Observed

POD = 0.27

FAR = 0.75

CSI = 0.34

Locations:

Forecast objects are

- Too far North (except B)
- Too far West (except C)

Precipitation intensity:

- Median intensity is too large
- Extreme (0.90th) intensity is too small

Size:

- Forecasts C and D are too small
- Forecast B is somewhat too large

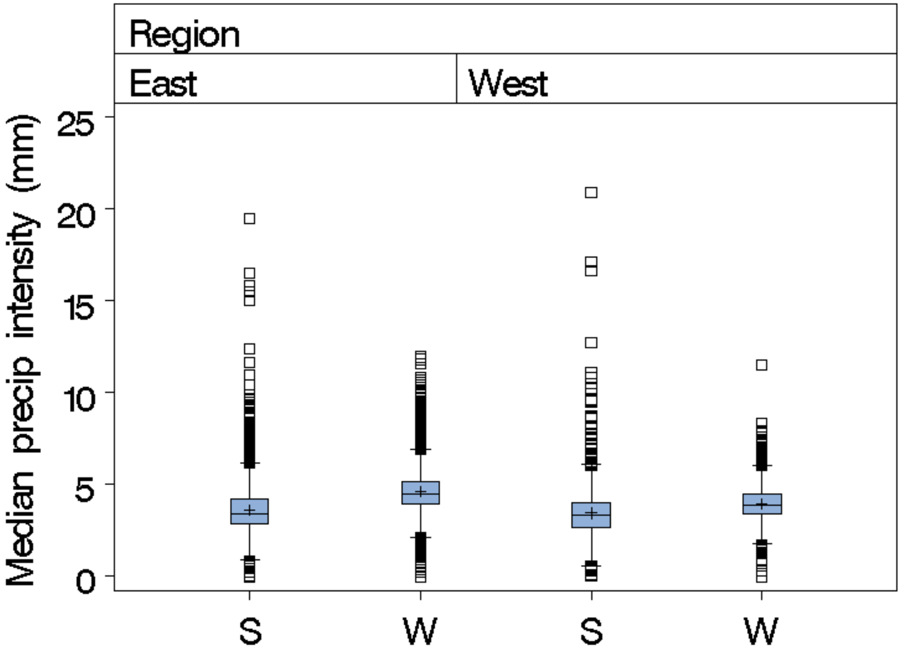
Matching:

- Two small observed objects were not matched

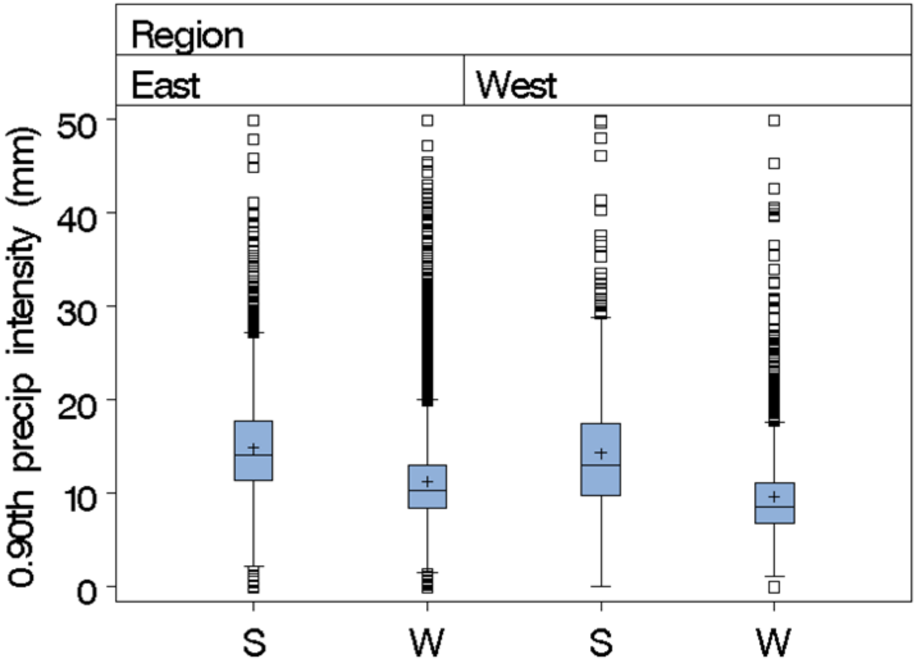
Example: Summarize across many forecasts

Does precipitation intensity vary between Forecast and Observed objects?

Median



0.90th Quantile



S = Observed; W = Forecast

AIRDAT/TAMDAR Application

- Apply object-based approach to RTFDDA precipitation forecasts *with* and *without* TAMDAR observations (Yubao Liu and Wei Yu)
- Establish stable set of observed objects to allow meaningful comparisons
- Summarize “climatological” differences as well as forecast-observation differences

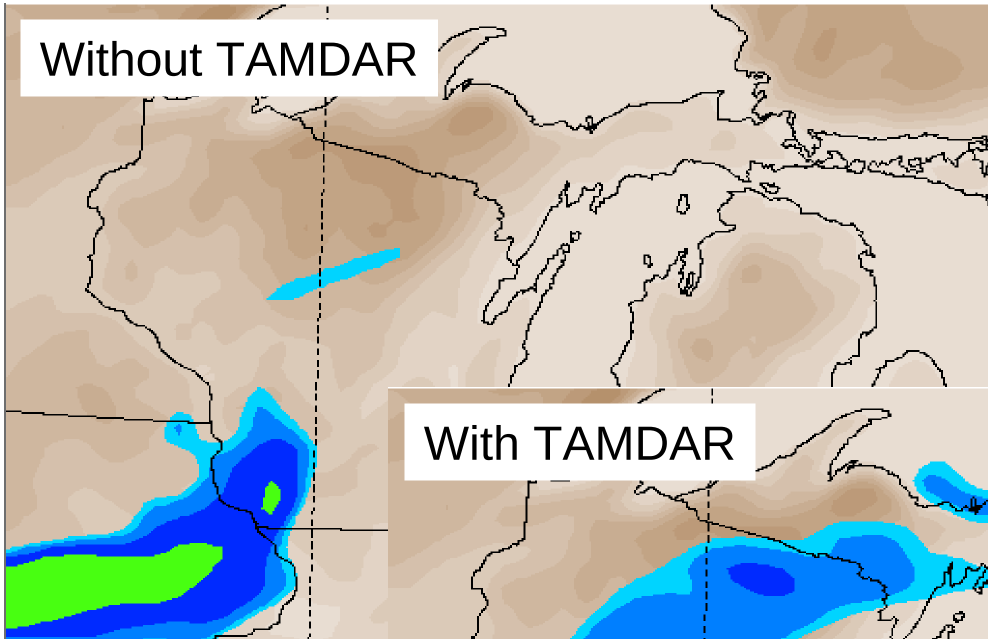
Weak snowbands (1)

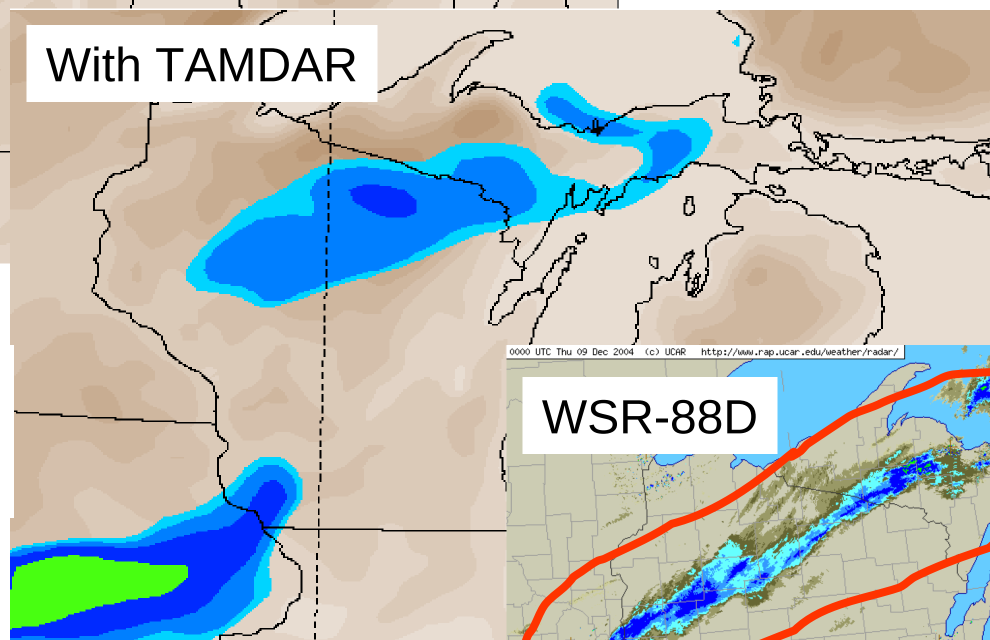
00Z, Dec. 09, 2004

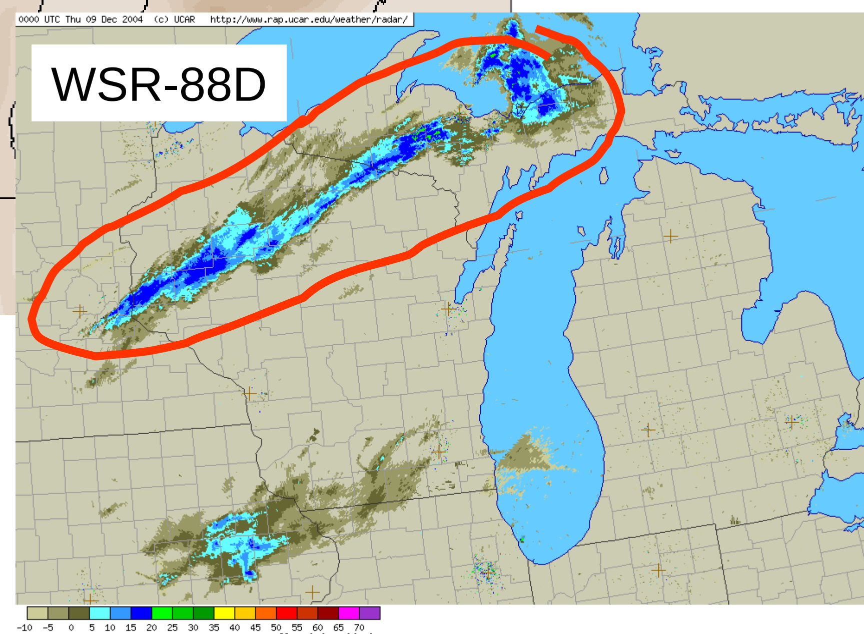
Radar reflectivity

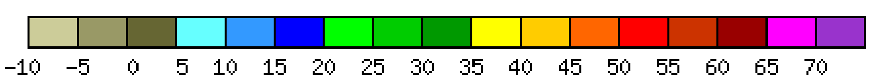
RTFDDA

1h forecasts









Extensions

- Include temporal dimension
- Additional application areas (clouds, icing, etc.)
- Ensemble forecasts
- Incorporate scaling approaches