Object identification techniques for object-oriented verification

Mike Baldwin

CIMMS/University of Oklahoma

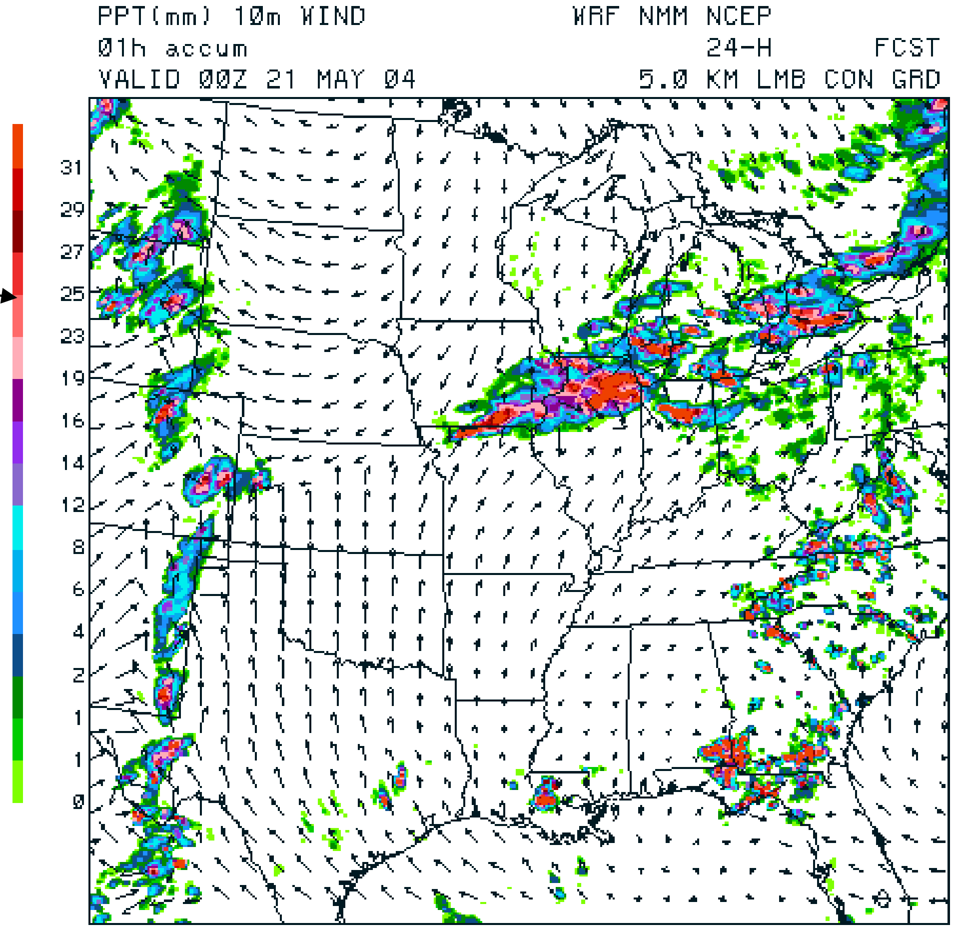
Baldwin's presentation on object-oriented verification

Harold Brooks

NSSL

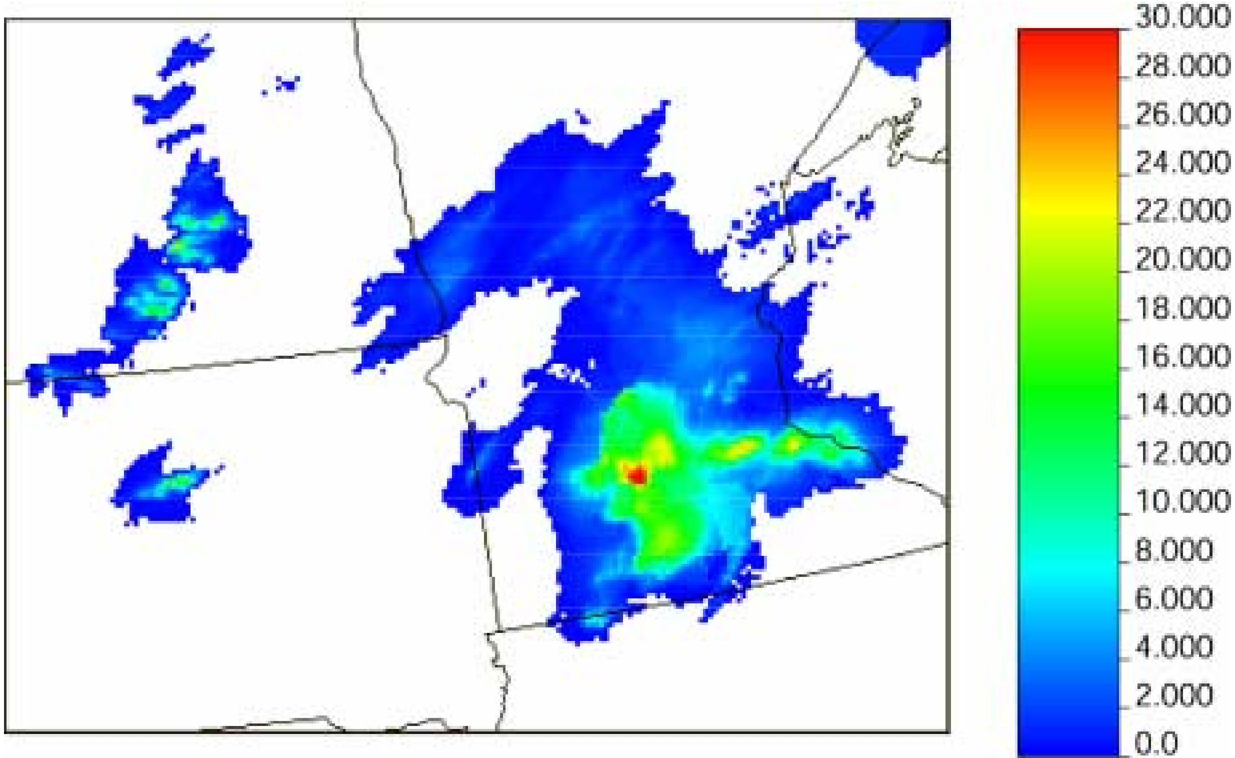
Issues

- Object identification – how many objects do *you* see?
- How to characterize and measure differences between objects?
- Dealing with different numbers of observed and forecast objects



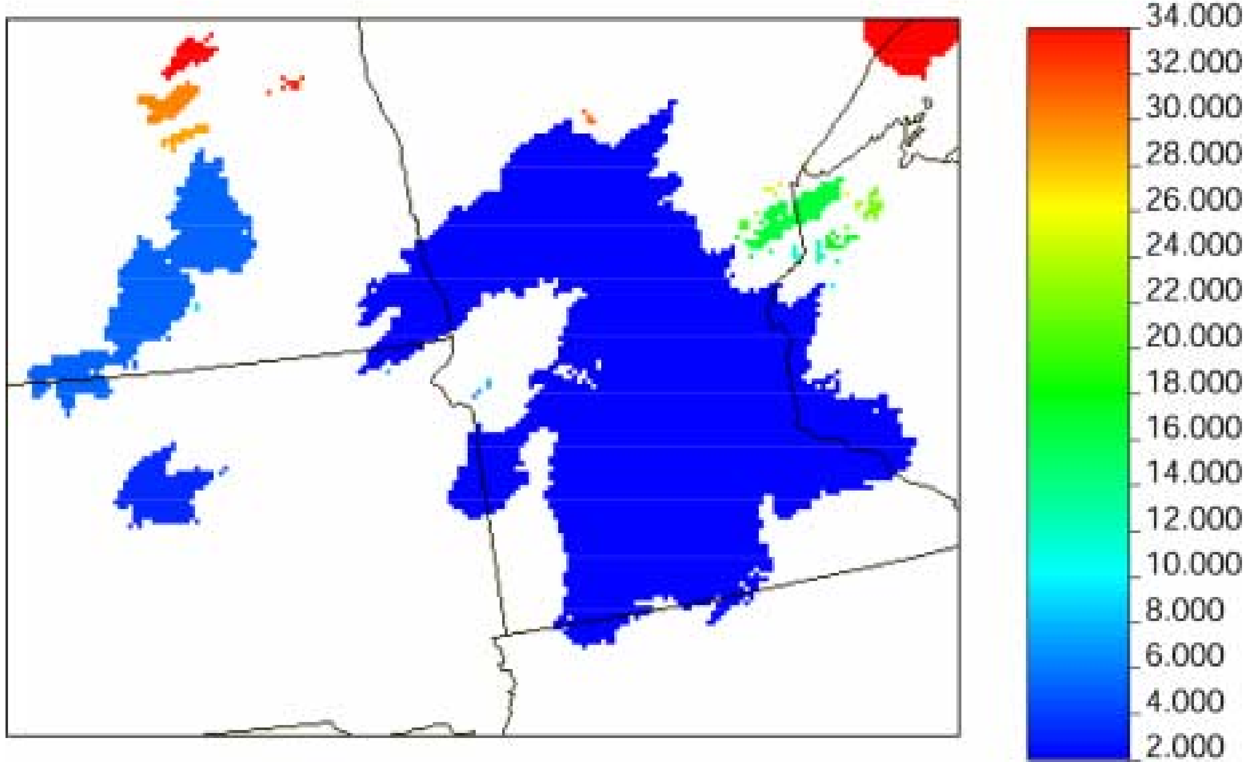
Automated rainfall object identification

- Contiguous regions of measurable rainfall (similar to Ebert and McBride 2000)



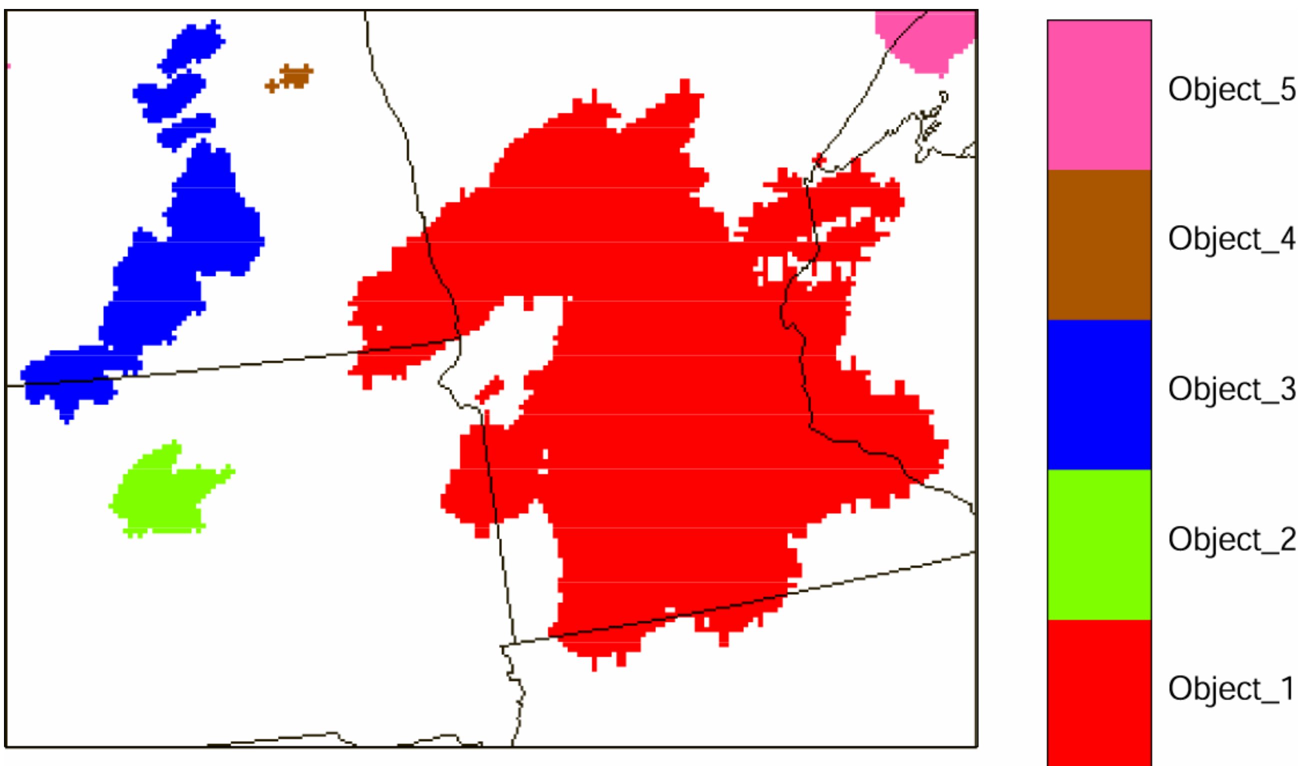
Connected component labeling

- *Pure* contiguous rainfall areas result in 34 unique “objects” in this example



Expand areas by 15%, connect regions that are within 20km

- Results in 5 objects

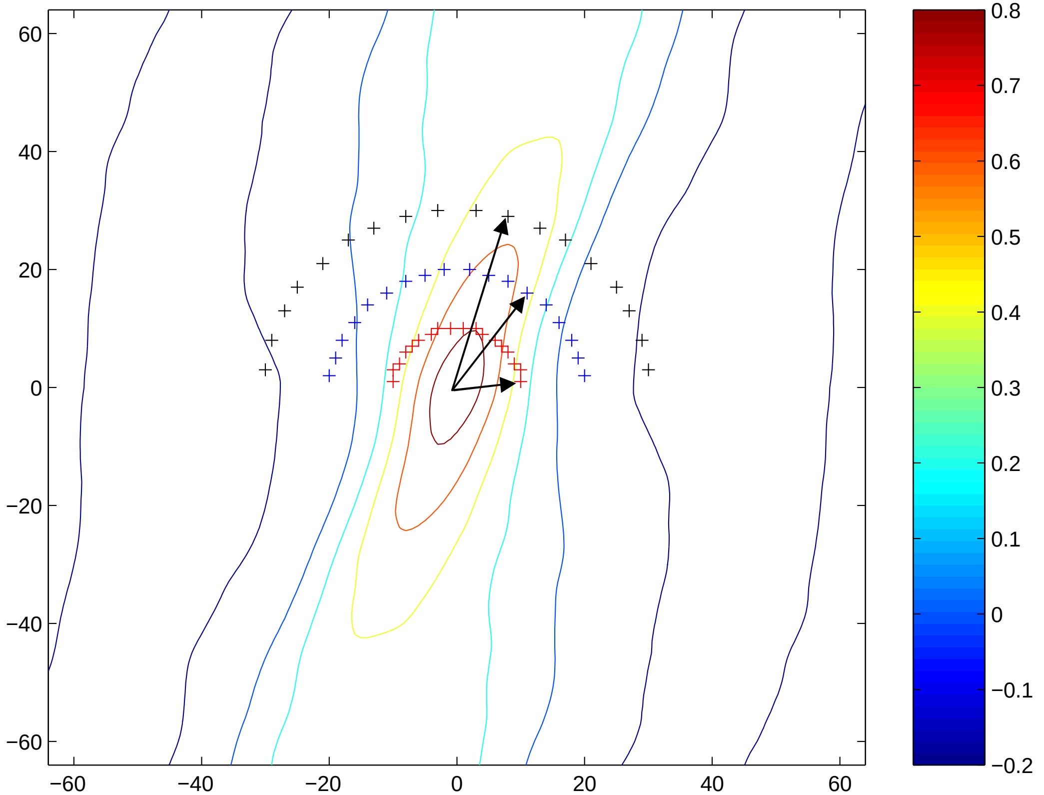


Useful characterization

- Attributes related to rainfall intensity and auto-correlation ellipticity were able to produce groups of stratiform, cellular, linear rainfall systems in cluster analysis experiments
- However, autocorrelation calculation is SLOW

New auto-correlation attributes

- Replaced ellipticity of AC contours with max-min correlation at specific lags (50, 100, 150km, every 10°)



Attributes

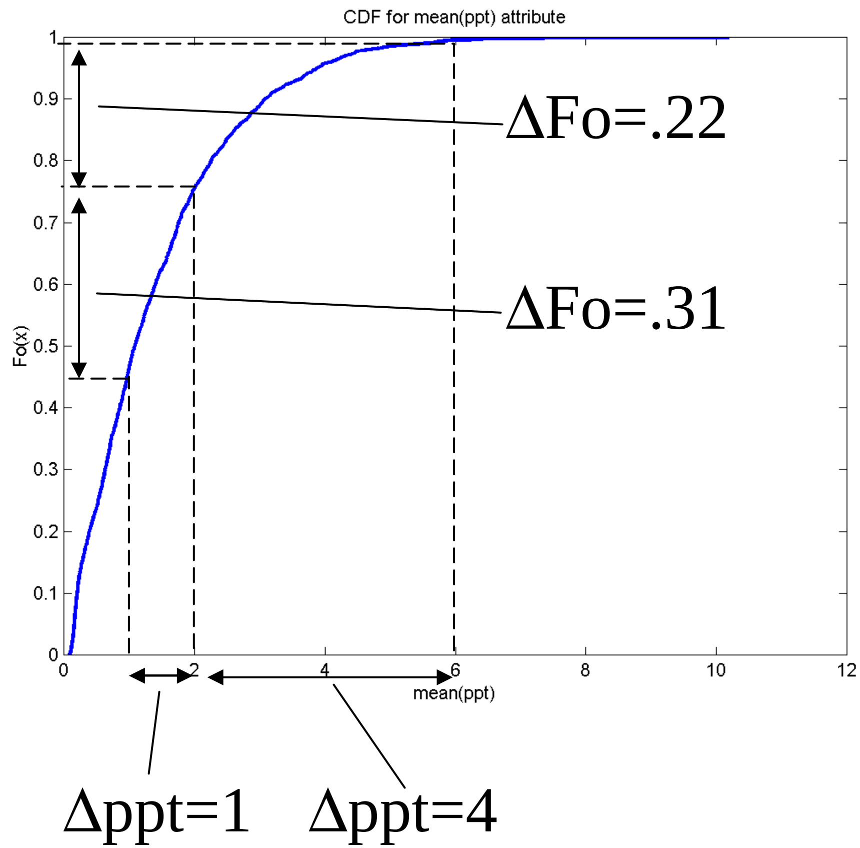
- Area (km^2), lat, lon
- Mean, std dev (σ) of precip (mm) within object
- Difference between max & min correlation at 50, 100, 150km lags (Δcorr)
- Orientation angle (θ) of max correlation at 50, 100, 150km lags (E-W = 0°, N-S=90°)
- Each object is characterized by 11 attributes, with a wide variety of units, ranges of values, etc.

How to measure “distance” between objects

- How to weigh different attributes?
  - Is 250km spatial distance same as 5mm precipitation distance?
- Do attribute distributions matter?
  - Is 55mm-50mm same as 6mm-1mm?
- How to standardize attributes?
  - X'=(x-min)/(max-min)
  - X'=(x-mean)/σ
  - LEPS

Decided to use LEPS

- Distance = 1 equates to difference between largest and smallest object for a particular attribute
- Linear for uniform dist (lat, lon, θ)
- Have to be careful with Δθ
- L1-norm: $d(x,y) = \sum_{i=1}^{n} |x_i - y_i|$



NSSL/SPC Spring Program 2004

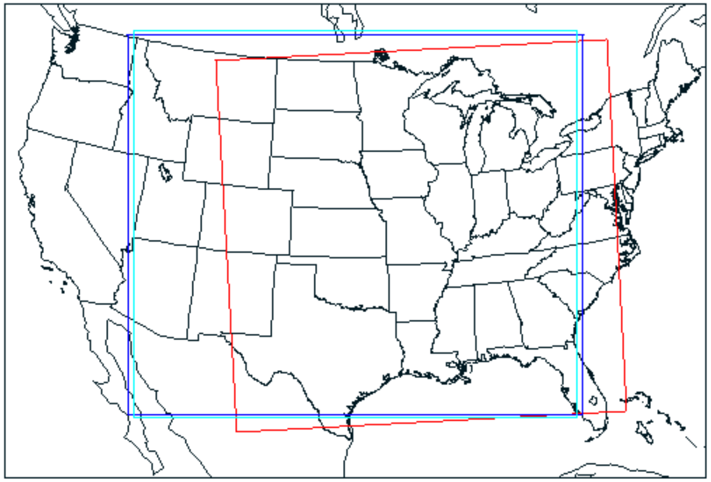
	WRF-NMM	WRF-NCAR	WRF-CAPS
Horz/ vert grid	4.5km/ 35 lvls	4.0km/ 35 lvls	4.0km/ 51 lvls
Physics	MYJ PBL Ferrier micro, GFDL rad	YSU PBL, Lin et al. micro, Dudhia-RRTM rad	YSU PBL, Lin et al. micro, Dudhia-RRTM rad
Init cond	Eta (interp 40 km)	Eta (interp 40km grid)	Eta + ADAS + Level II

Observed ppt = Stage II (radar-only) 4km 1h accum

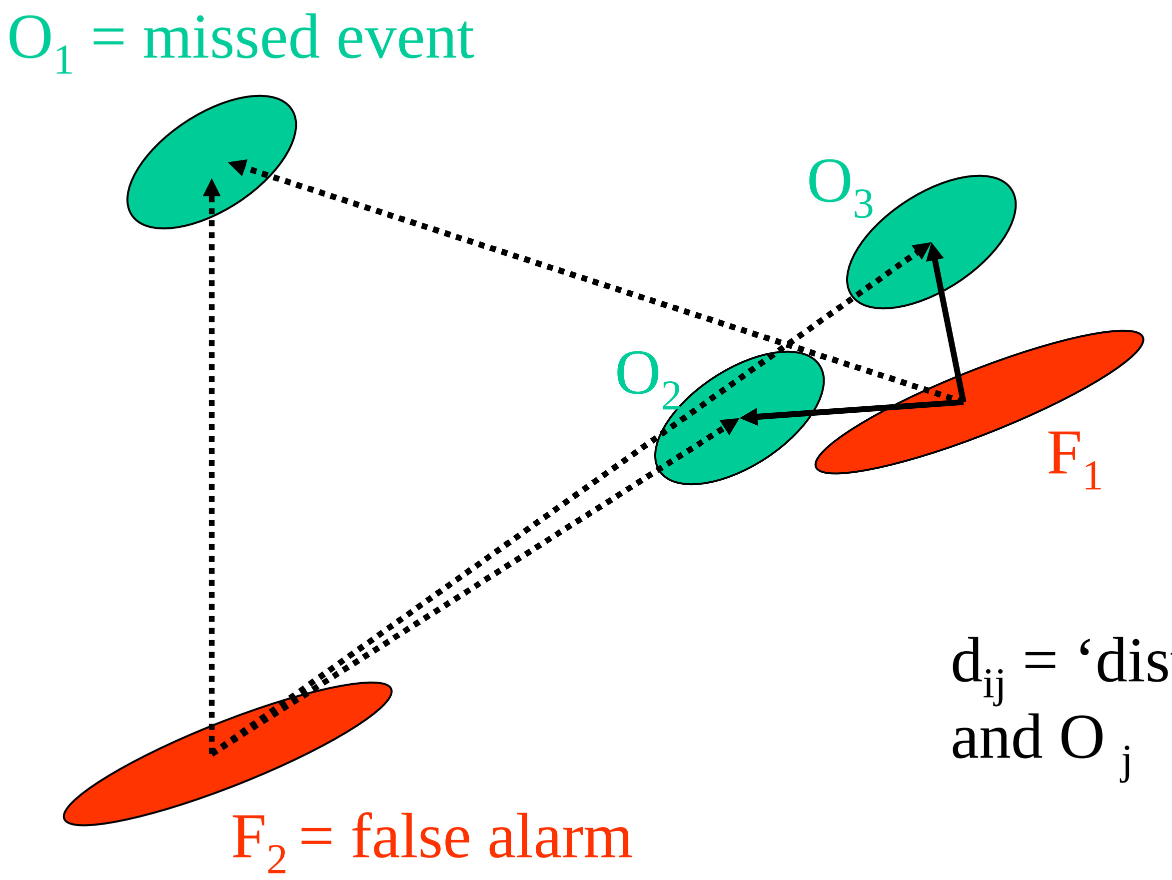
- Comparison for ~1 month (May 10 – Jun 4)

Object ID and characterization

- Remapped each model to same grid as Stage II, common domain for all
- Run object ID, get attributes
- Create database of objects meso-α scale and larger [~ (200 km)^2]

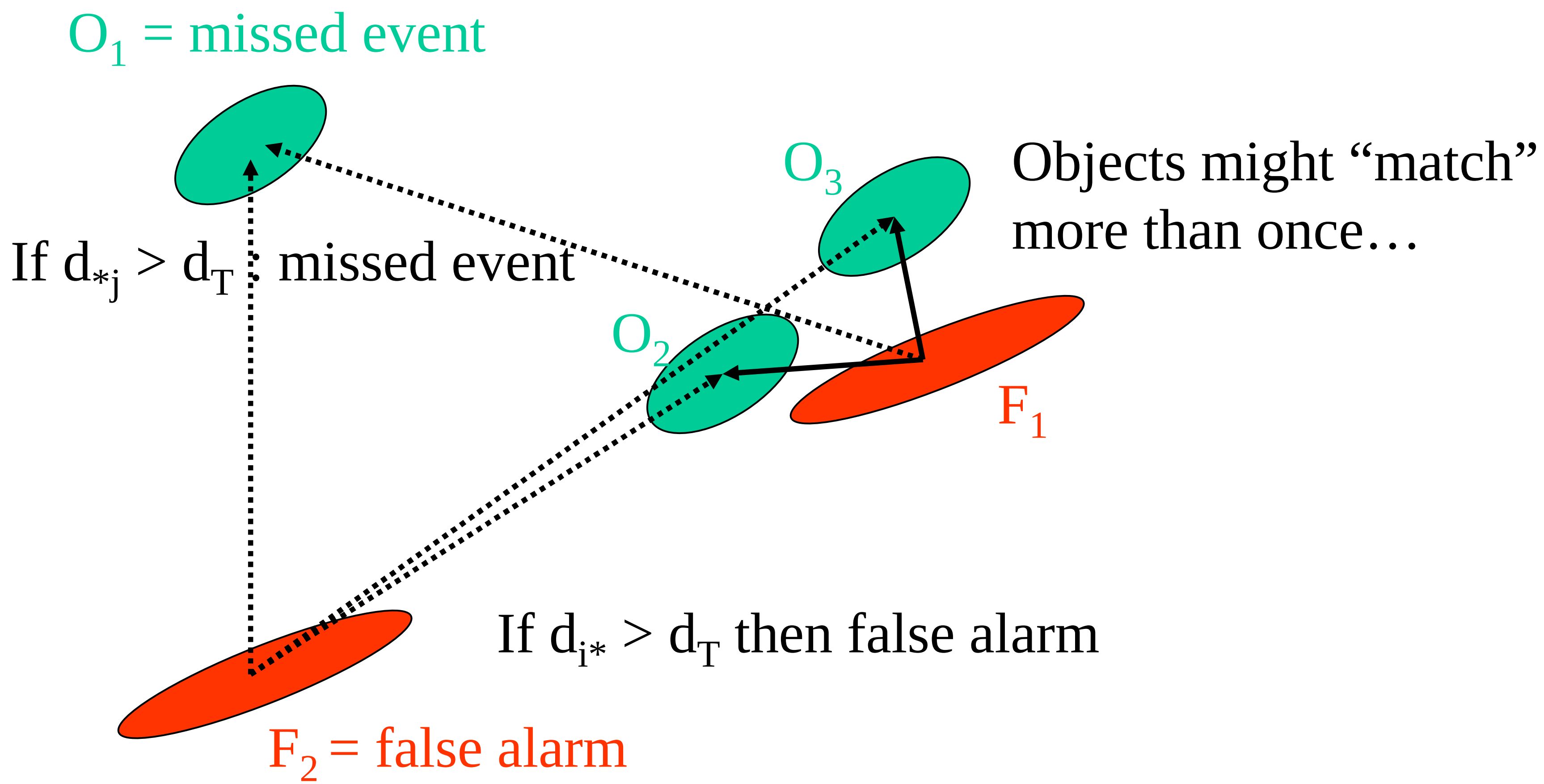


How to match observed and forecast objects?



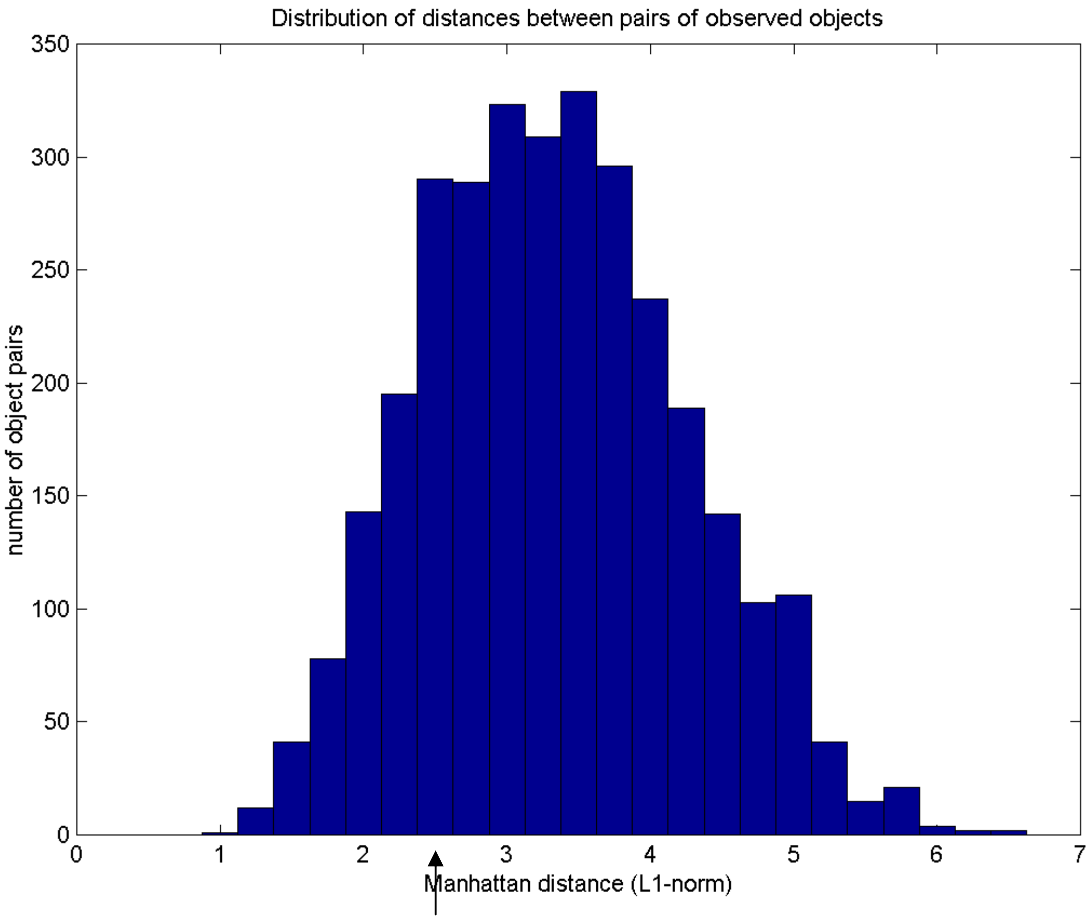
d_{ij} = ‘distance’ between F $_i$ and O $_j$

How to match observed and forecast objects?



Estimate of d_T threshold

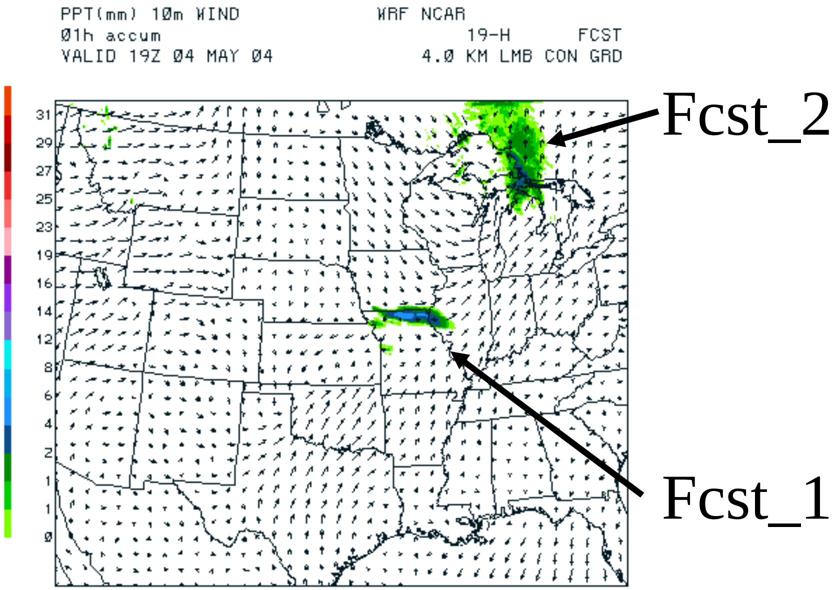
- Compute distance between each observed object and all others at the same time
- d_T = 25th percentile = 2.5
- Forecasts have similar distributions



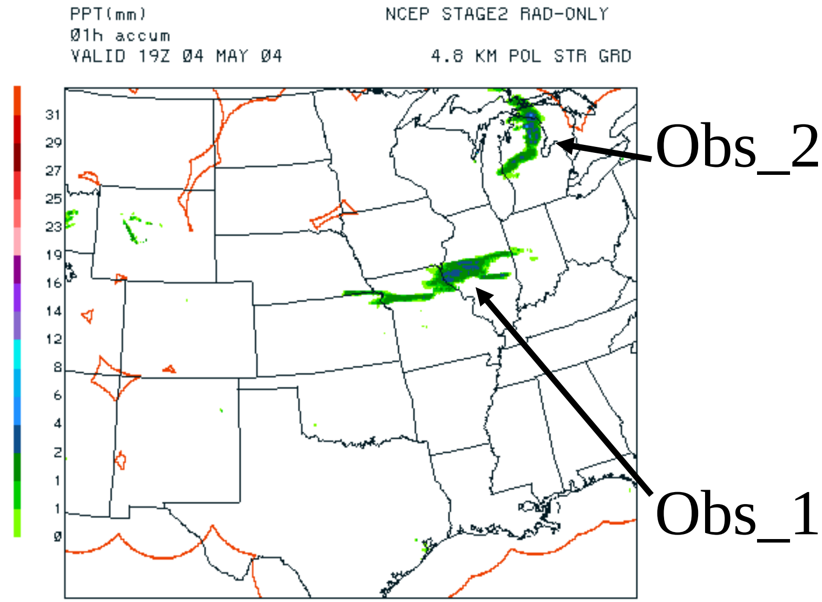
25th %-ile

Example of object verf

NCAR WRF 4km



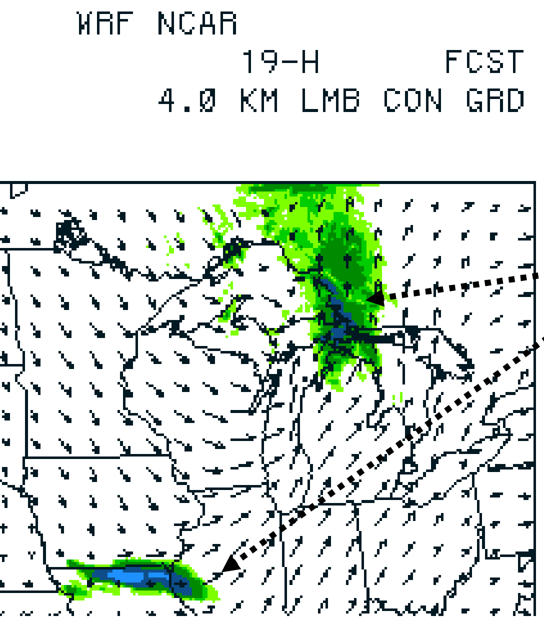
Stage II radar ppt



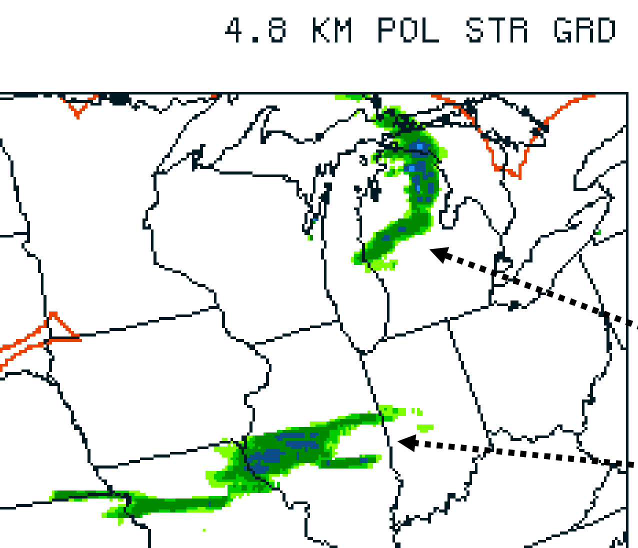
Object identification procedure identifies 2 forecast objects and 2 observed objects

Attributes

NCAR WRF 4km



Stage II radar ppt



	Fcst_1	Fcst_2	Obs_1	Obs_2
	Area=70000 km^2	Area=285000	Area=135000	Area=70000 km^2
	mean(ppt)=0.97	mean(ppt)=0.32	mean(ppt)=0.45	mean(ppt)=0.60
	σ (ppt)= 1.26	σ (ppt)= 0.44	σ (ppt)= 0.57	σ (ppt)= 0.67
	Δcorr(50)=1.17	Δcorr(50)=0.27	Δcorr(50)=0.37	Δcorr(50)=0.36
	Δcorr(100)=0.99	Δcorr(100)=0.42	Δcorr(100)=0.54	Δcorr(100)=0.52
	Δcorr(150)=0.84	Δcorr(150)=0.48	Δcorr(150)=0.58	Δcorr(150)=0.49
	θ(50)=173°	θ(50)=95°	θ(50)=171°	θ(50)=85°
	θ(100)=173°	θ(100)=85°	θ(100)=11°	θ(100)=75°
	θ(150)=173°	θ(150)=85°	θ(150)=11°	θ(150)=65°
	lat = 40.2°N	lat = 47.3°N	lat = 39.9°N	lat = 44.9°N
	lon = 92.5°W	lon = 84.7°W	lon = 91.2°W	lon =84.5°W

Obs_2

Obs_1

Distances between objects

- After transforming raw attributes to probability space (observed CDF: LEPS)
- Using L1-norm (Manhattan distance)

Fcst_1, Obs_1 : 1.48 [match]

Fcst_2, Obs_1 : 2.74

Fcst_1, Obs_2 : 2.75

Fcst_2, Obs_2 : 1.39 [match]

Obs_1, Obs_2 : 2.18
Fcst_1, Fcst_2 : 3.81

Average distances for matching fcst and obs objects

- 1-30h fcsts, 10 May – 03 June 2004
- Eta (12km) = 2.12
- WRF-CAPS = 1.97
- WRF-NCAR = 1.98
- WRF-NMM = 2.02

With set of matching obs and fcsts

- Nachamkin (2004) compositing ideas
  - errors given fcst event
  - errors given obs event
- Distributions of errors for specific attributes
- Use classification to stratify errors by convective mode