NCAR

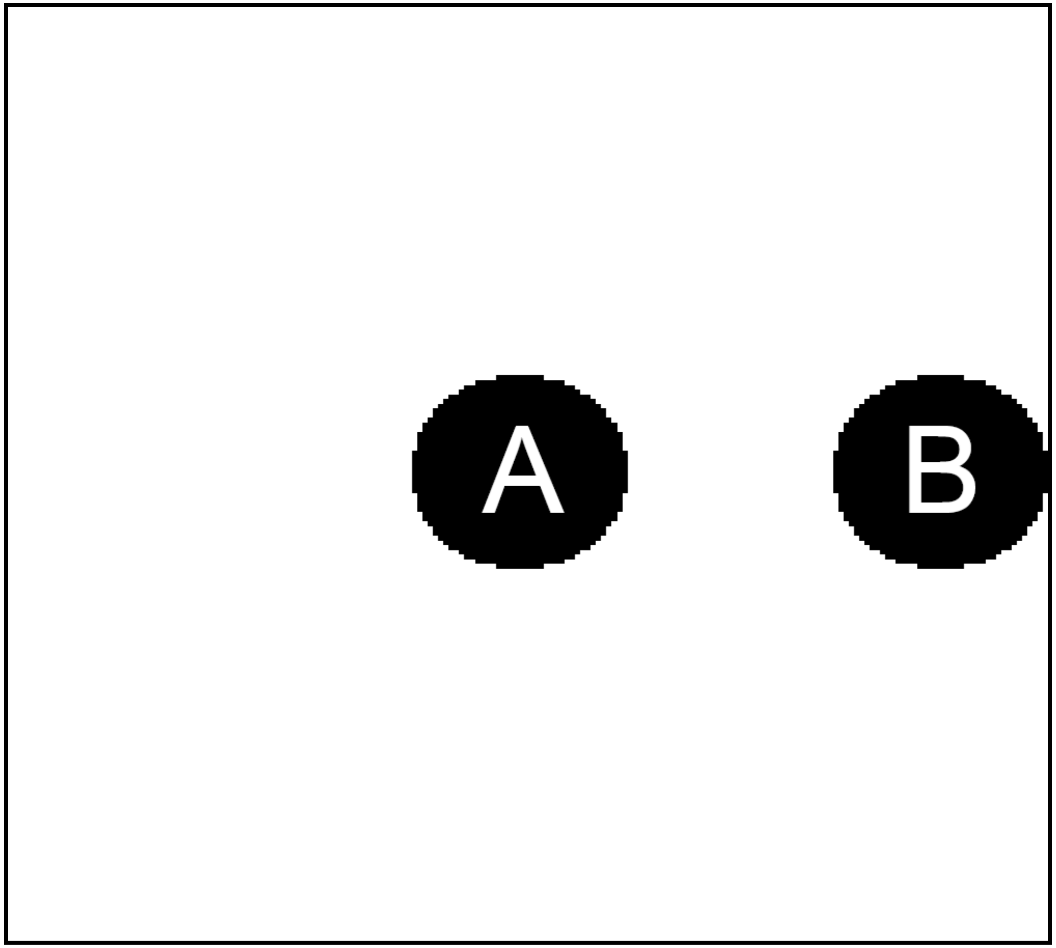
Comparing Distance Methods for Spatial Verification

Eric Gilleland and Barbara G. Brown
Weather Systems Assessment Program
Research Applications Laboratory

7th International Verification Methods Workshop

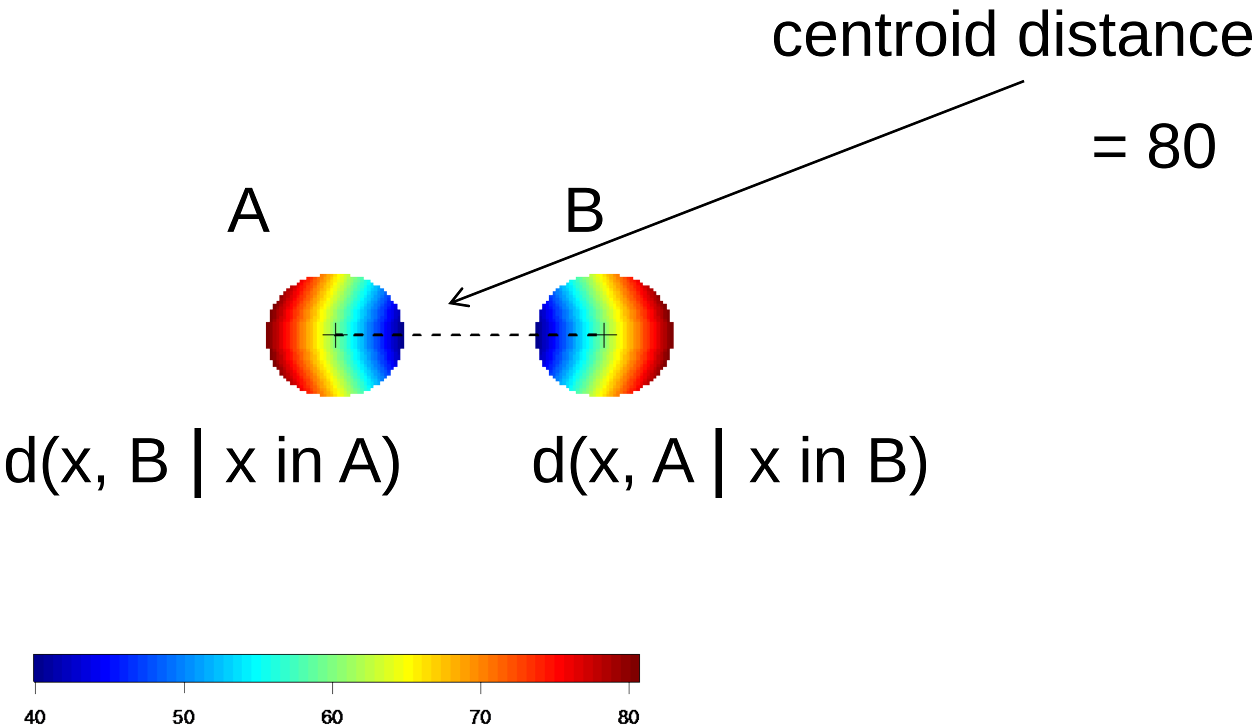
10 May 2017

National Center for Atmospheric Research


A
B

Mean Error Distance

MED(A, B) is the average distance from points in the set B to points in the set A



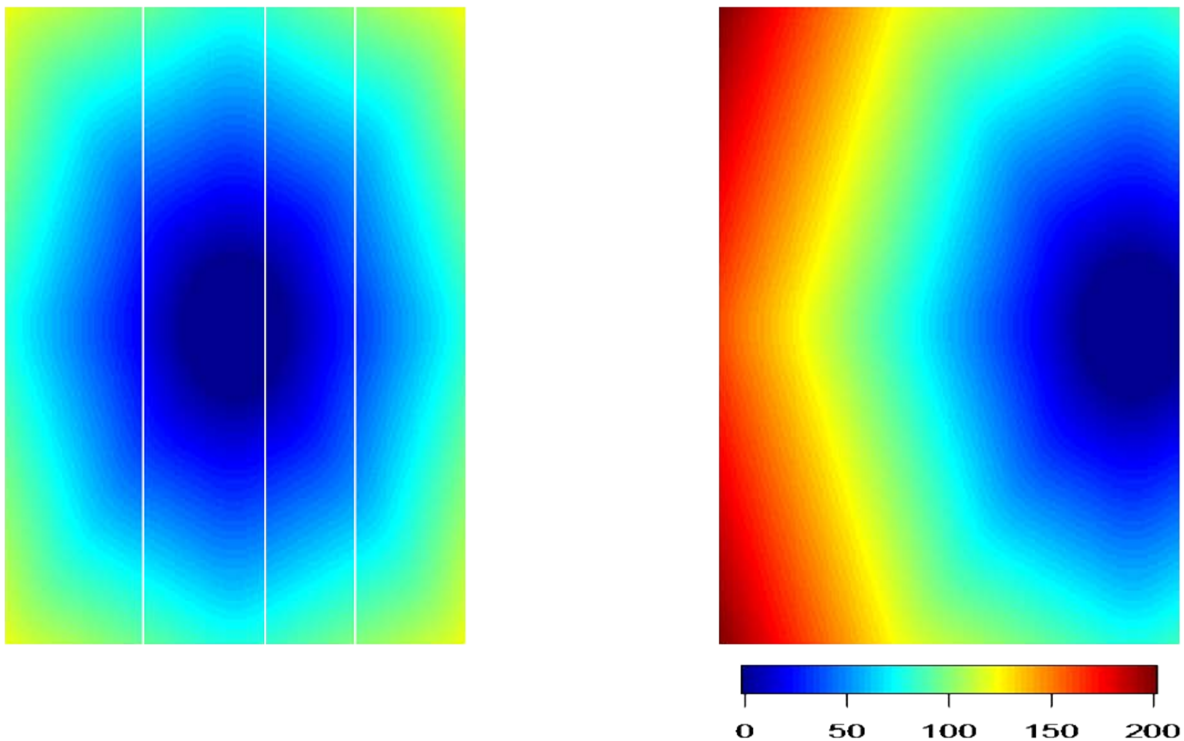
$MED(A,B) = \Sigma_x \, d(x, A \mid x \text{ in } B) \, / \, N_B$ $MED(B,A) = \Sigma_x \, d(x, B \mid x \text{ in } A) \, / \, N_A$

N_B is the number of points in the set B

Baddeley's Δ Metric

NCAR

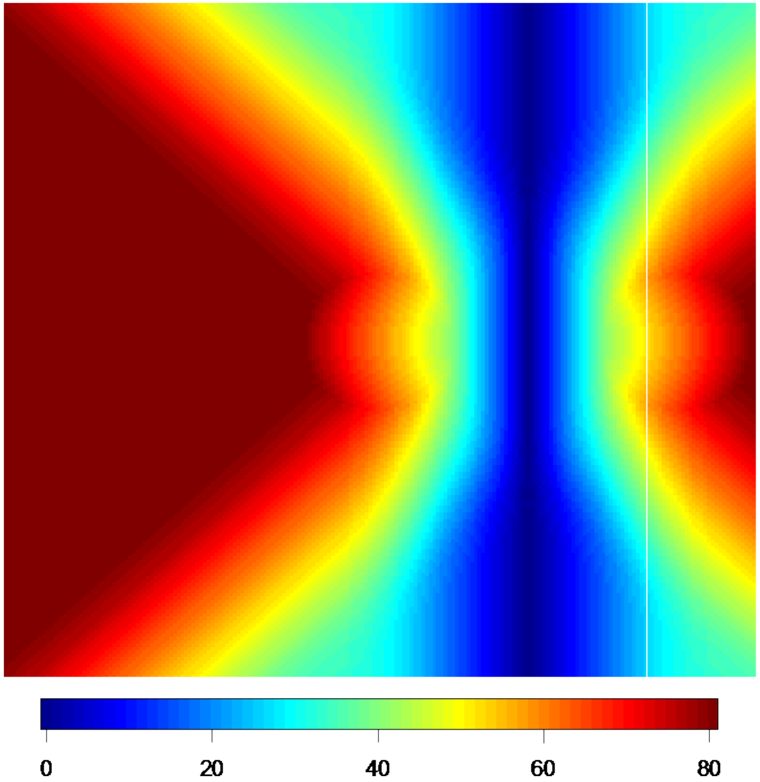
d(x, A) d(x, B)



Distance maps for A and B. Note dependence on location within the domain.

Baddeley's Δ Metric

$T = | d(x, A) - d(x, B) |$



- p = 1 gives the arithmetic average of T
- p = 2 is the usual choice
- p = ∞ gives the max of T (Hausdorff distance)

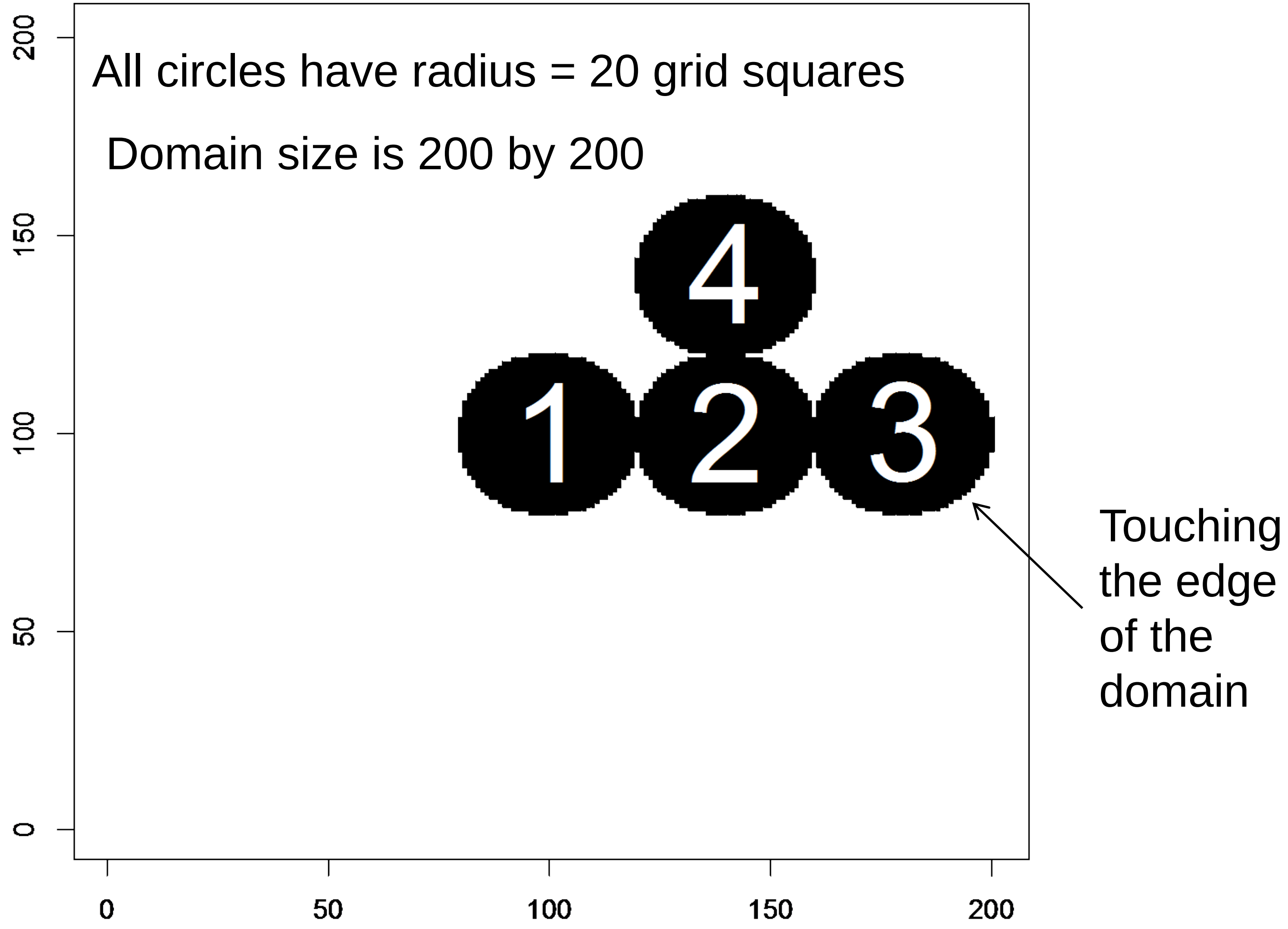
Δ is the L_p norm of T

d(x, A) and d(x, B) are first transformed by a function ω. Usually,
ω(x) = max(x, constant), but all results here use ∞ for the constant term.

$$\Delta(A, B) = \Delta(B, A) = \left[\Sigma_{x \text{ in Domain}} | d(x, A) - d(x, B) |^p \right]^{1/p} / N$$

N is the size of the domain

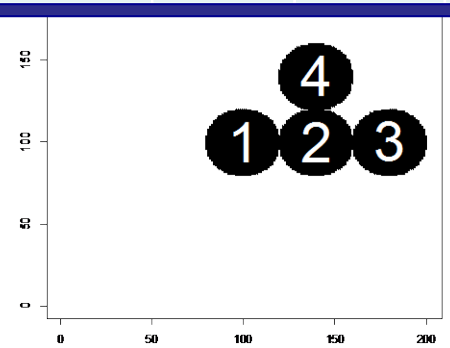
Contrived Examples: Circles



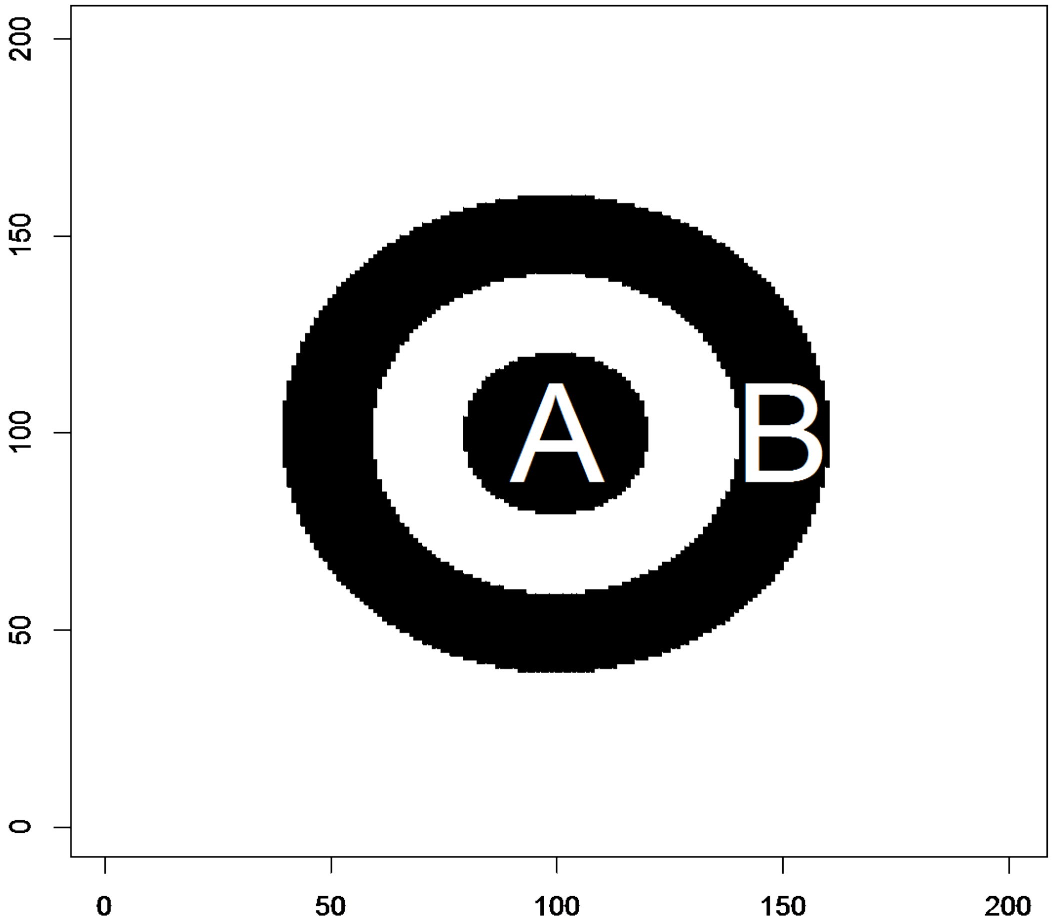
Contrived Examples: Circles

A	B	MED(A, B)	rank	MED(B, A)	rank	Δ(A, B)	rank	cent dist.	rank
1	2	22	2	22	1	29	2	40	2
1	3	62	4	62	3	57	6	80	4
1	4	38	3	38	2	41	5	57	3
2	3	22	2	22	1	31	3	40	2
2	4	22	2	22	1	28	1	40	2
2	1, 3, 4	11	1	22	1	29	2	13	1
3	4	38	3	38	2	38	4	57	3

If comparisons are made after centering the two binary fields on a new, square grid (201 by 201), then Δ is 28.84 for 1 vs 2, 2 vs 3 and 2 vs 4



Circle and a Ring

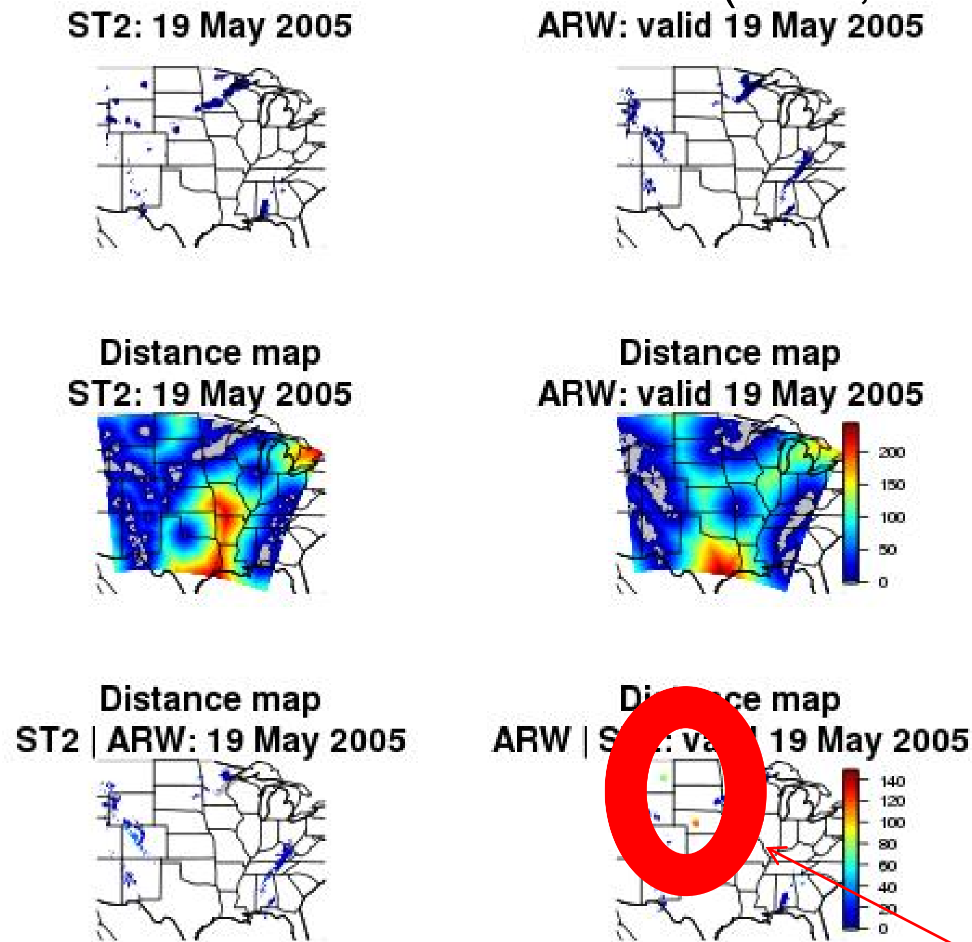


MED(A, B) = 32
MED(B, A) = 28
Δ(A, B) = 38

centroid distance = 0

Mean Error Distance

MED(ST2, ARW) ≈ 15.42 is much smaller than MED(ARW, ST2) ≈ 66.16



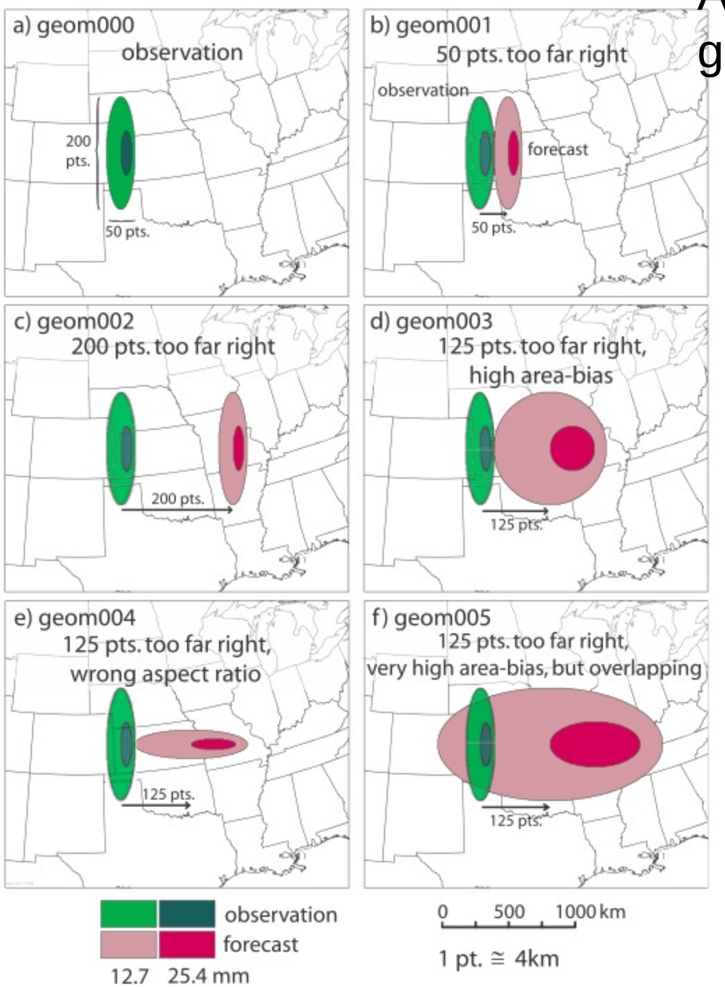
High sensitivity to small changes in the field!

Good or bad quality depending on user need.

Fig. 2 from G. (2016 submitted to WAF, available at: http://www.ral.ucar.edu/staff/ericg/Gilleland2016.pdf)

Missed Areas

Geometric ICP Cases



Avg. Distance from green to pink

Avg. Distance from pink to green

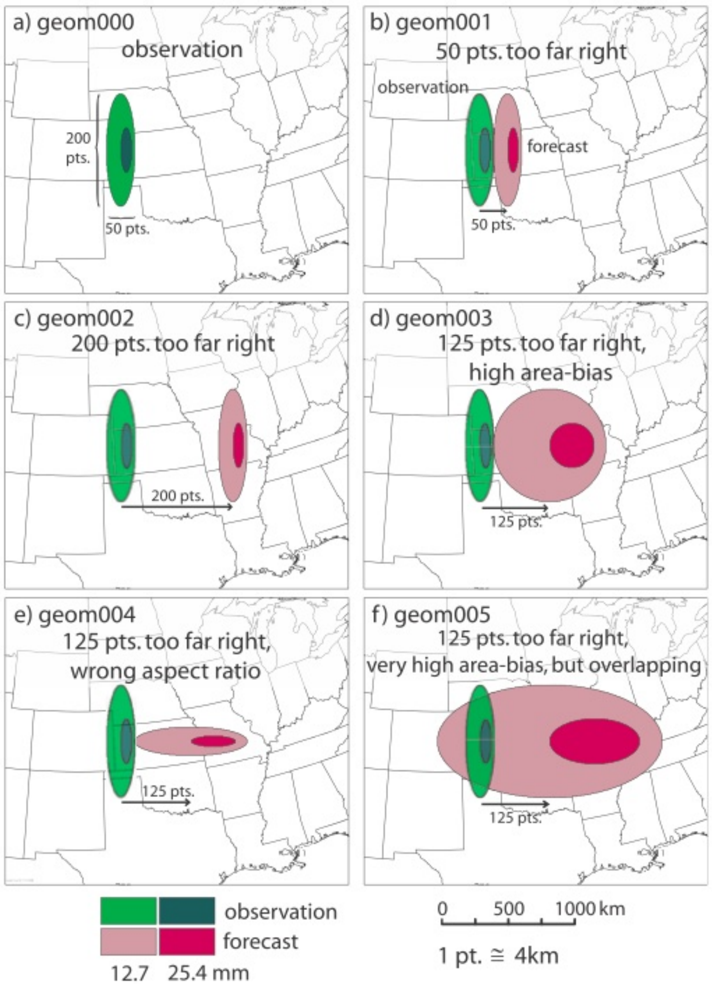
Case	MED(A, Obs)	rank	MED(Obs, A)	rank
1	29	2	29	1
2	180	5	180	5
3	36	3	104	3
4	52	4	101	2
5	1	1	114	4

Values rounded to zero decimal places

Table from part of Table 1 in G. (2016, submitted to WAF)
Fig. 1 from Ahijevych *et al*. (2009, *WAF*, **24**, 1485 – 1497)

Geometric ICP Cases

NCAR



Case	MED(A, Obs)	rank	MED(Obs, A)	rank
1	29	2	29	1
2	180	5	180	5
3	36	3	104	3
4	52	4	101	2
5	1	1	114	4

Case	Δ(A, Obs)	rank
1	45	1
2	167	5
3	119	3
4	106	2
5	143	4

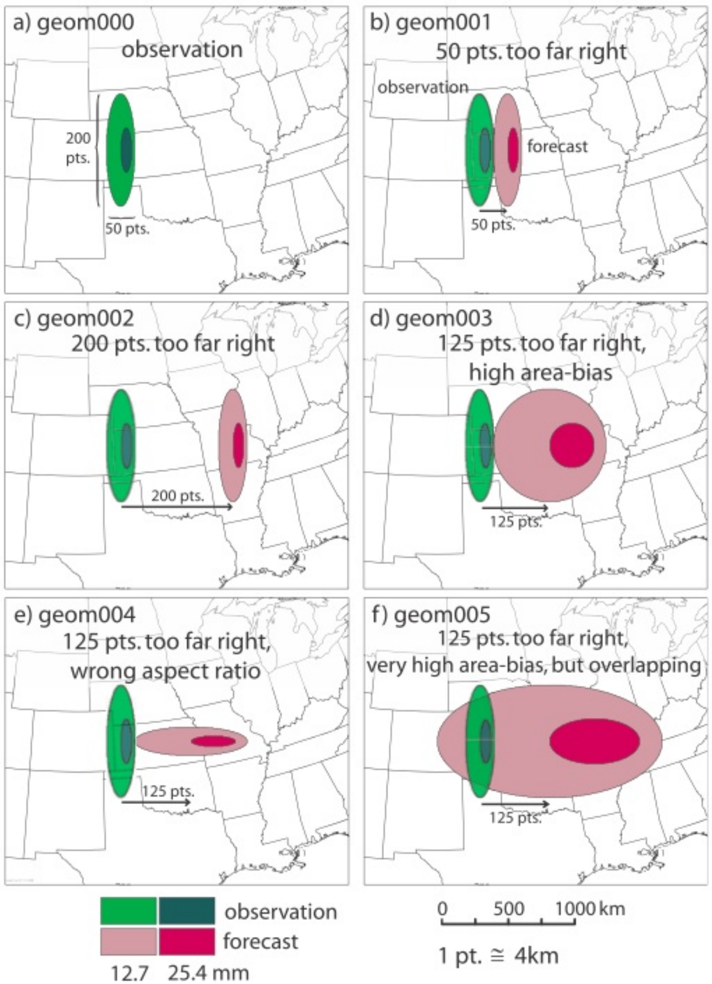
Values rounded to zero decimal places

Table from part of Table 1 in G. (WAF, 2017)

Fig. 1 from Ahijevych *et al*. (2009, *WAF*, **24**, 1485 – 1497)

Geometric ICP Cases

NCAR



Case	Δ(A, Obs)	rank
1	45	1
2	167	5
3	119	3
4	106	2
5	143	4

After centering fields and expanding grid to 601 by 601

Case	Δ(A, Obs)	rank
1	43	1
2	161	5
3	114	3
4	96	2
5	146	4

Values rounded to zero decimal places

Table from part of Table 1 in G. (WAF 2017)
Fig. 1 from Ahijevych *et al*. (2009, *WAF*, **24**, 1485 – 1497)

Mean Error Distance

- Magnitude of MED tells how good or bad the “misses/false alarms” are.
- Miss = Average distance of observed non-zero grid points from forecast.
  - Perfect score: MED(Forecast, Observation) = zero (no misses at all)
    - All observations are within forecasted non-zero grid point sets.
  - Good score = Small values of MED(Forecast, Observation)
    - all observations are near forecasted non-zero grid points, on average.
- False alarm = Average distance of forecast non-zero grid points from observations.
  - Perfect score: MED(Observation, Forecast) = zero (no false alarms at all)
    - All forecasted non-zero grid points fall overlap completely with observations.
  - Good score = Small values of MED(Observation, Forecast)
    - all forecasts are near observations, on average.
- Hit/Correct Negative
  - Perfect Score: MED(both directions) = 0
  - Good Value = Small values of MED(both directions)

Mean Error Distance

NCAR

MesoVICT core cases

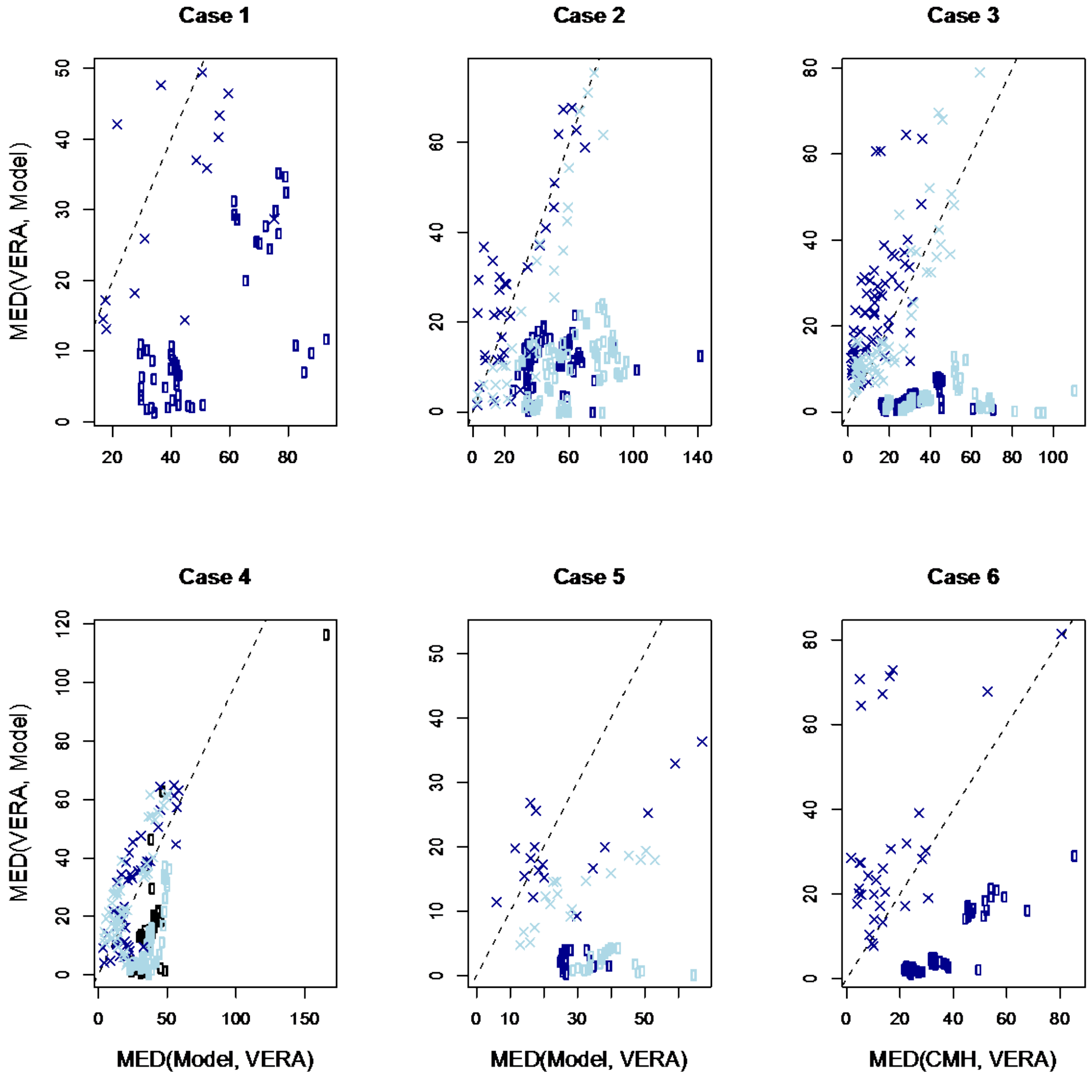
CMH

CO2

Threshold

o = 0.1 mm h^{-1}

× = 5.1 mm h^{-1}



Misses

False Alarms

MED Summary

- Mean Error Distance
  - Useful summary when applied in both directions
  - New idea of false alarms and misses (spatial context)
  - Computationally efficient and easy to interpret
- Properties
  - High sensitivity to small changes in one or both fields
  - Does not inform about bias per se
    - Could hedge results by over forecasting, but only if over forecasts are in the vicinity of observations!
  - No edge or position effects (unless part of object goes outside the domain)
  - Does not inform about patterns of errors
  - Does not directly account for intensity errors (only location)
  - Fast and easy to compute and interpret
- Complementary Methods include (but not limited to)
  - Frequency/Area bias (traditional)
  - Geometric indices (AghaKouchak et al 2011, doi:10.1175/2010JHM1298.1)

Baddeley's Δ Metric Summary

- Sensitive to differences in size, shape, and location
- A proper mathematical metric (therefore, amenable to ranking)
  - positivity ($\Delta(A, B) \geq 0$ for all A and B)
  - identity ($\Delta(A, A) = 0$ and $\Delta(A, B) > 0$ if $A \neq B$)
  - symmetry ($\Delta(A, B) = \Delta(B, A)$)
  - triangle inequality ($\Delta(A, C) \leq \Delta(A, B) + \Delta(B, C)$)
- Sensitive to position within the domain
  - Issue is overcome by centering (the pair of binary fields together) on a new square grid.
- Upper limit bounded only by domain size
  - Any comparisons across cases needs to be done on the same grid.
  - Grid should be square and comparisons should be done with object(s) centered on the grid.

Centroid Distance Summary

NCAR

- Is a true mathematical metric. So, conducive to rankings.
- Not sensitive to position within a field (or orientation of A to B; i.e., if A and B are rotated as a pair, the distance does not change)
- No edge effects
- Gives useful information for translation errors between objects that are similar in size, shape and orientation.
- Not sensitive to area bias
- Not as useful otherwise.
- Should be combined with other information.

- Thank you
- Questions?

- Gilleland, E., 2017. A new characterization in the spatial verification framework for false alarms, misses, and overall patterns. Weather Forecast., 32 (1), 187 - 198, DOI: 10.1175/WAF-D-16-0134.1.