Test Fields and Comparisons for Distance-based Spatial Forecast Verification Methods

Eric Gilleland

Research Applications Laboratory, National Center for Atmospheric Research

Co-authors: Gregor Skok, Barb Brown, Barbara Casati, Manfred Dorninger, Marion Mittermaier, Nigel Roberts, and Laurie Wilson

Support for the first author on this project was provided by the Developmental Testbed Center (DTC). The DTC Visitor Program is funded by the National Oceanic and Atmospheric Administration, the National Center for Atmospheric Research and the National Science Foundation. The second author acknowledges the financial support from the Slovenian Research Agency (research core funding No. P1-0188).

Spatial Forecast Verification

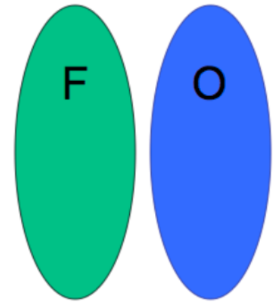
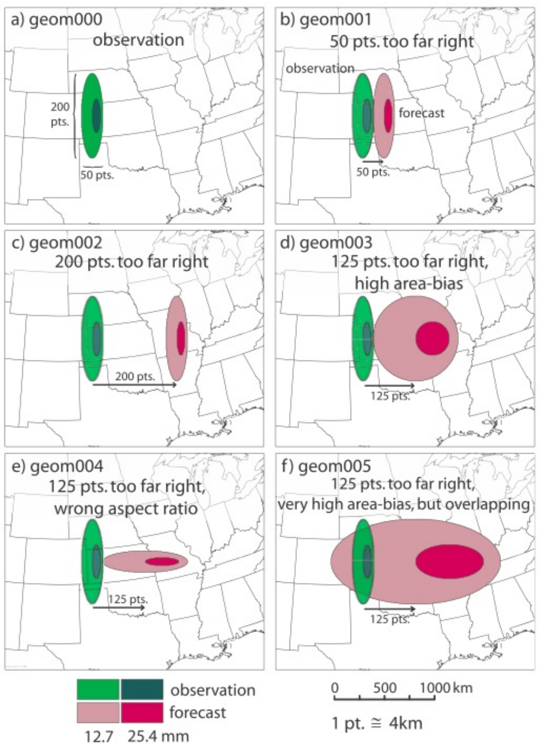


Figure from Barb Brown



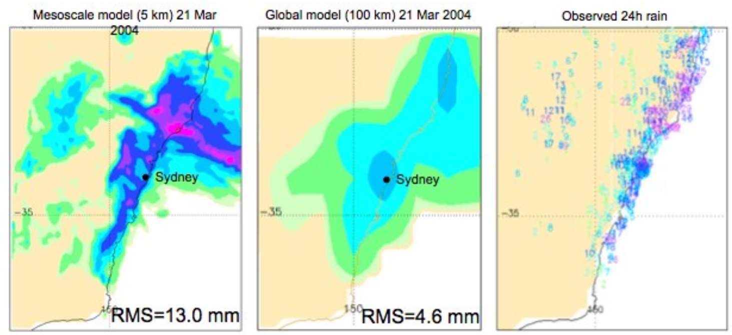
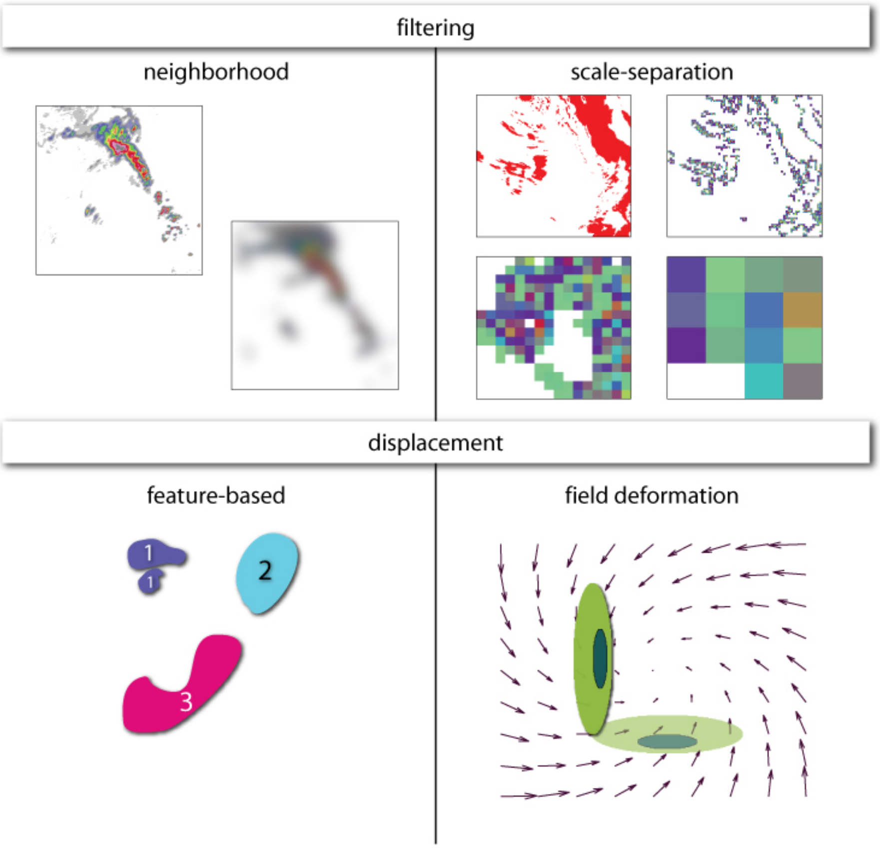


Figure from Beth Ebert

Traditional score	geom001/002/004	geom003	geom005
Accuracy	0.95	0.87	0.81
Frequency bias	1.00	4.02	8.03
Multiplicative intensity bias	1.00	4.02	8.04
RMSE (mm)	3.5	5.6	6.9
Bias-corrected RMSE (mm)	3.5	5.5	6.3
Correlation coefficient	−0.02	−0.05	0.20
Probability of detection	0.00	0.00	0.88
Probability of false detection	0.03	0.11	0.19
False alarm ratio	1.00	1.00	0.89
Hanssen–Kuipers discriminant (H-K)	−0.03	−0.11	0.69
Threat score or CSI	0.00	0.00	0.11
Equitable threat score or GSS	−0.01	−0.02	0.08
HSS	−0.03	−0.04	0.16

Far left figure and table from Ahijevych et al., 2009. *Weather Forecast.*, **24** (6), 1485 - 1497, doi: 10.1175/2009WAF2222298.1.

Spatial Forecast Verification



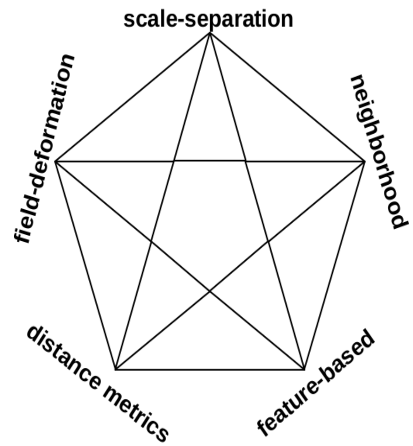
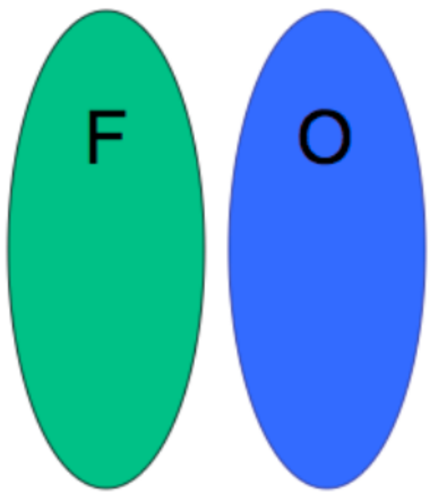


Fig. above from Barbara Casati

Fig. 2 from G. *et al*. (2010, *BAMS*, **91** (10), 1365 – 1373)

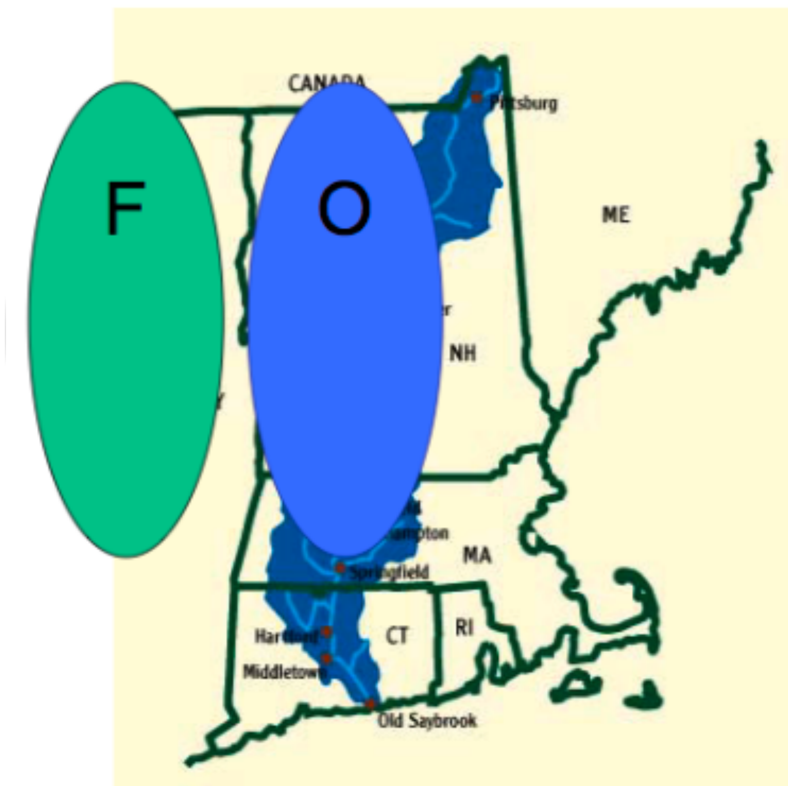
Spatial Forecast Verification



What properties about a forecast are most important? Is this forecast a good one?

Figure from Barb Brown

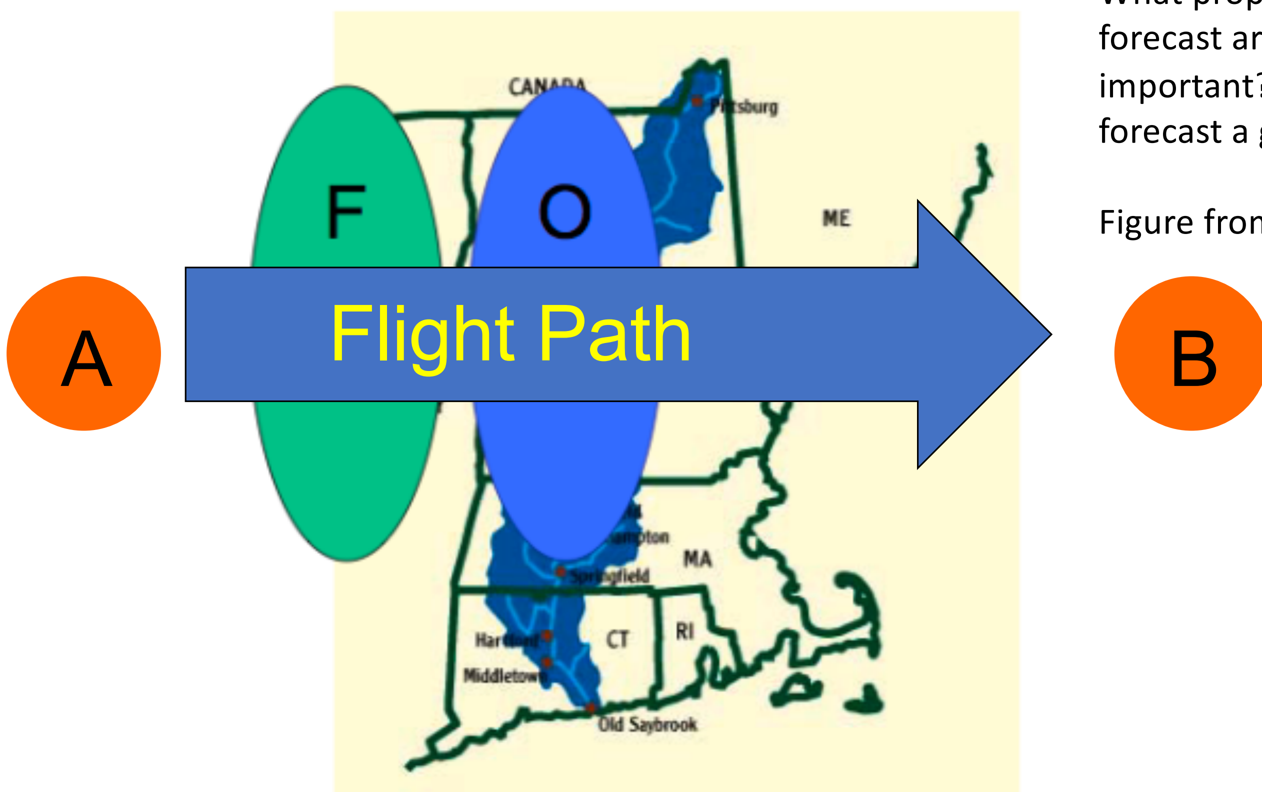
Spatial Forecast Verification



What properties about a forecast are most important? Is this forecast a good one?

Figure from Barb Brown

Spatial Forecast Verification



What properties about a forecast are most important? Is this forecast a good one?

Figure from Barb Brown

Mathematical metric

A measure $m(A, B) \geq 0$ is a *metric* if it satisfies

- Identity: $m(A, B) = 0$ if and only if $A = B$.
- Symmetry: $m(A, B) = m(B, A)$
- Triangle Inequality: $m(A, C) \leq m(A, B) + m(B, C)$

Ensures that if C is closer to A than B is to A, then $m(A, C) < m(A, B)$

Centroid Distance

Let $\boldsymbol{s}_i = (x, y) \in \mathcal{D}$ be grid-point locations within the domain, $\mathcal{D}$, with $i = 1, \dots, N$. Let $Z(\boldsymbol{s}_i)$ be the intensity (value) at location i for $i = 1, \dots, N$. Then, the ***centroid*** is the location of the center of mass of the domain and is calculated by

$$\mathbf{C}(\mathcal{D}) = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{s}_i \cdot Z(\boldsymbol{s}_i) = \frac{1}{|\mathcal{D}|}\sum_{\boldsymbol{s}\in\mathcal{D}} \boldsymbol{s} \cdot Z(\boldsymbol{s})$$

Can also be calculated for a subset, $A \subset \mathcal{D}$, of the domain (e.g., for a single feature within the field). In which case, replace $\mathcal{D}$ in the equation with A.

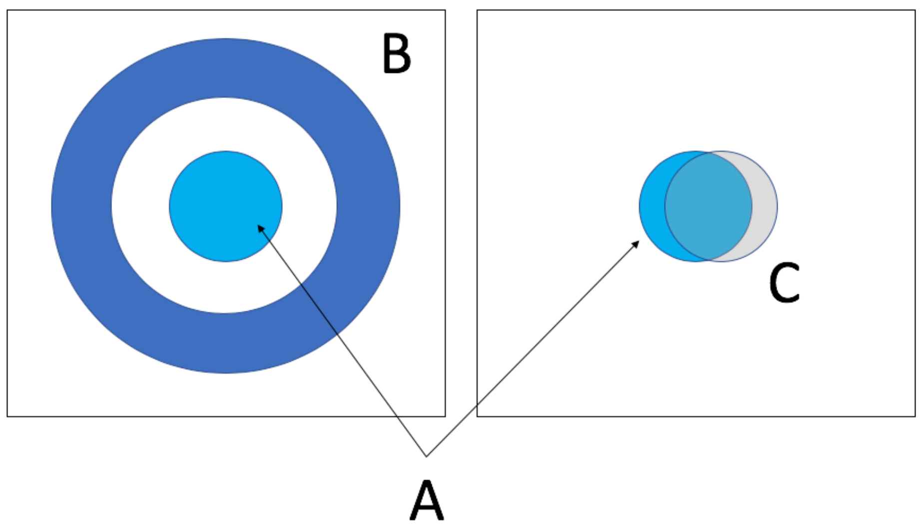
Centroid Distance

The ***centroid distance***, $C(A, B)$, between two sets (or entire fields) is the distance between their centers of mass.

The centroid distance is a true mathematical metric

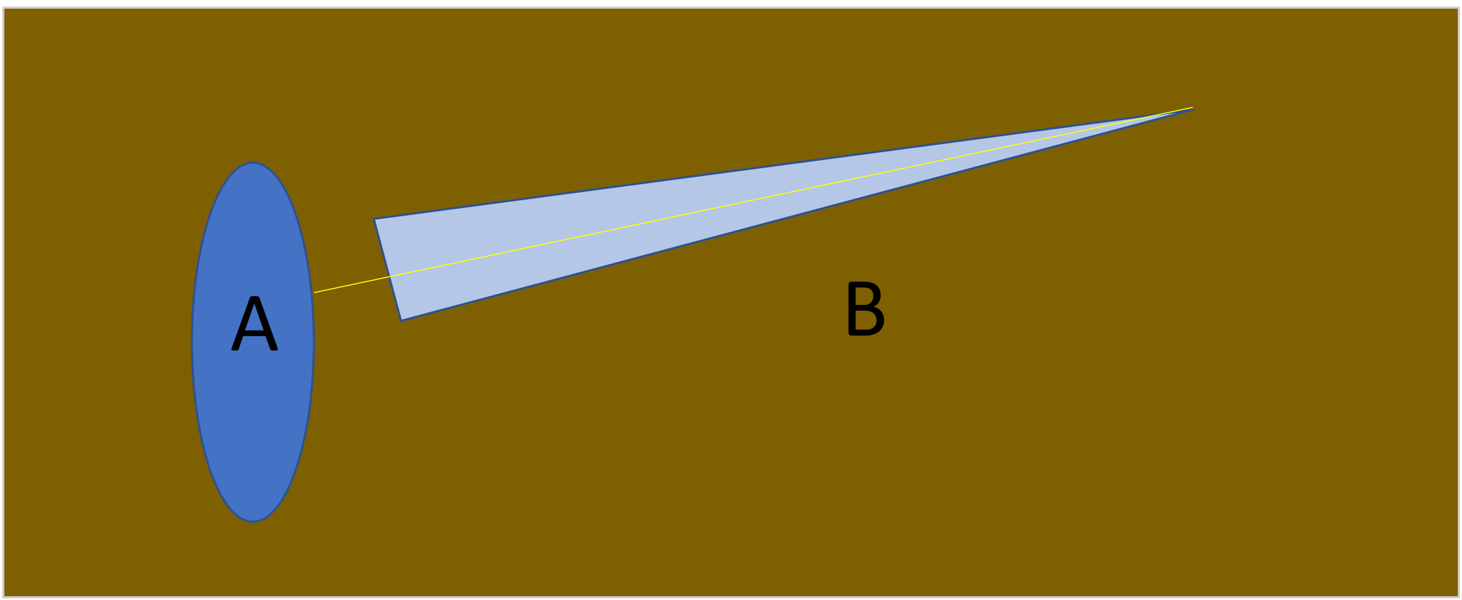
Centroid Distance

But, is it a good metric
for your purpose?



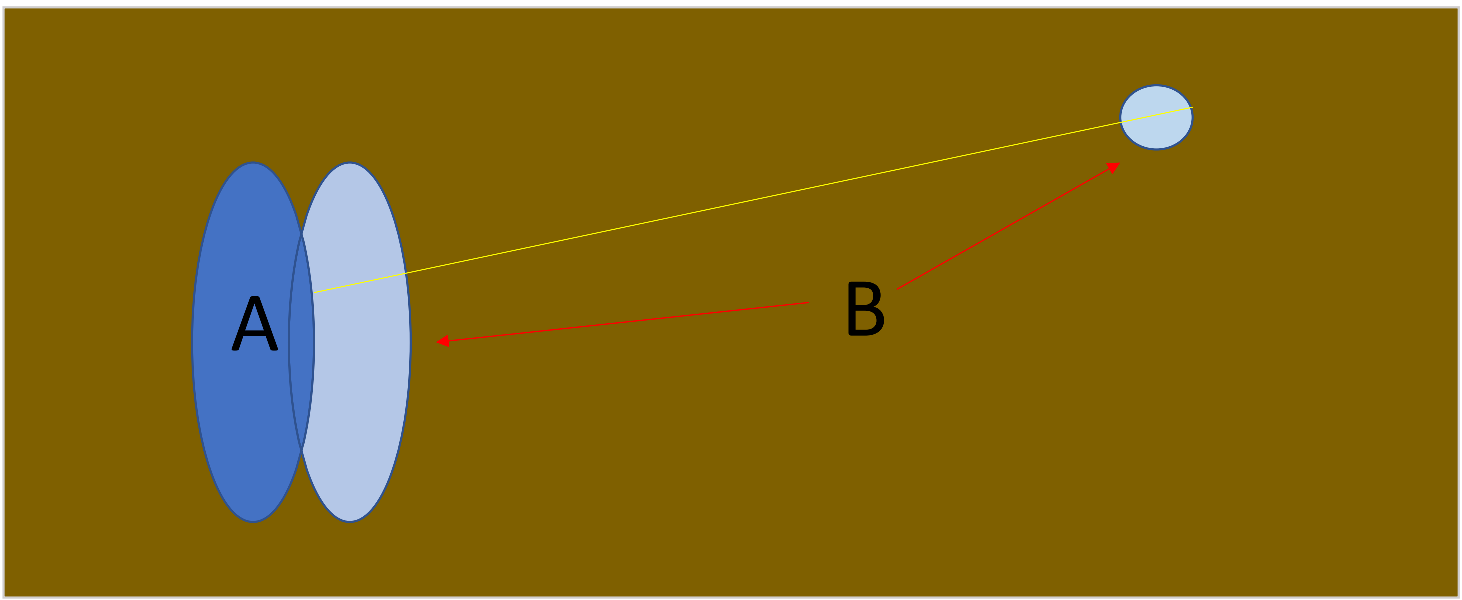
Hausdorff distance

The Hausdorff distance is defined to be the maximum of the shortest distance from each point in one set to the nearest point in another set.



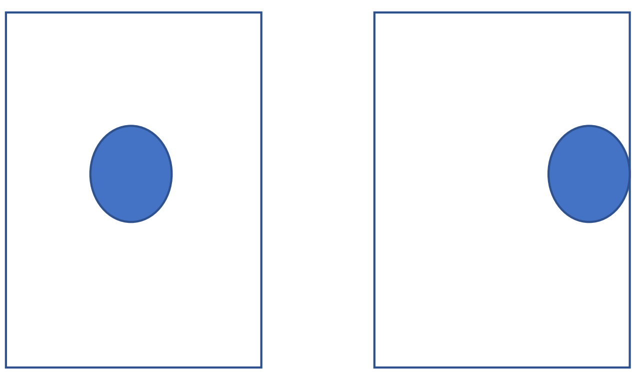
Hausdorff distance

The Hausdorff distance is defined to be the maximum of the shortest distance from each point in one set to the nearest point in another set.

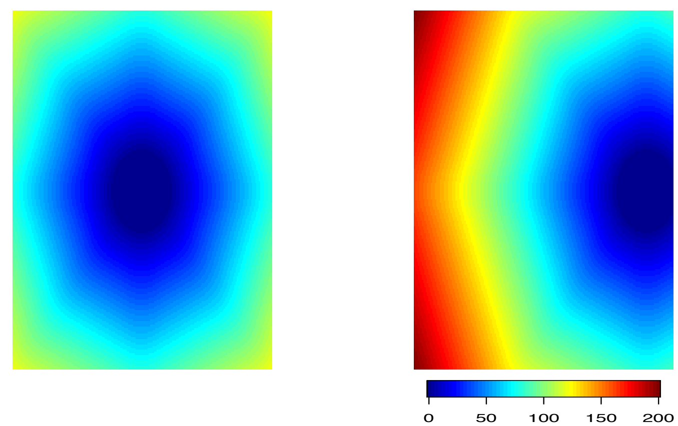


Distance maps

A B

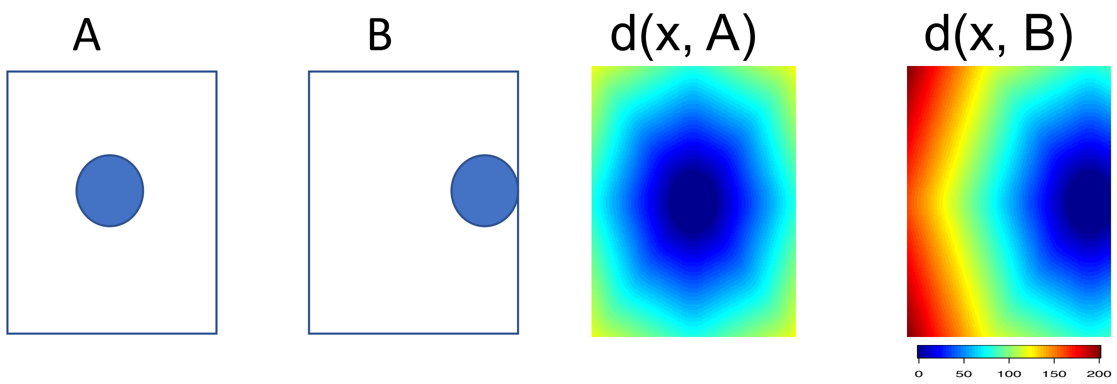


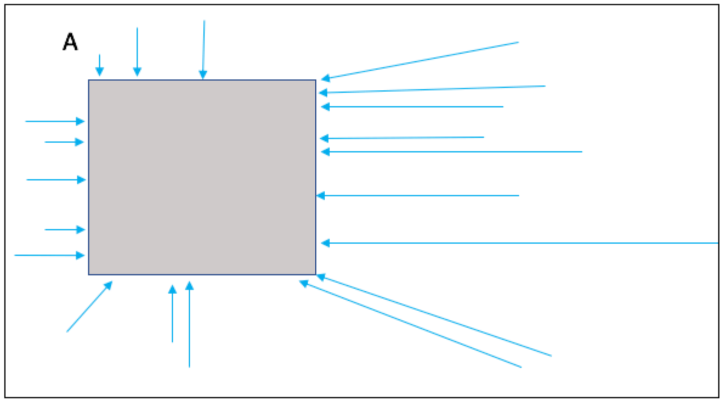
d(x, A) d(x, B)



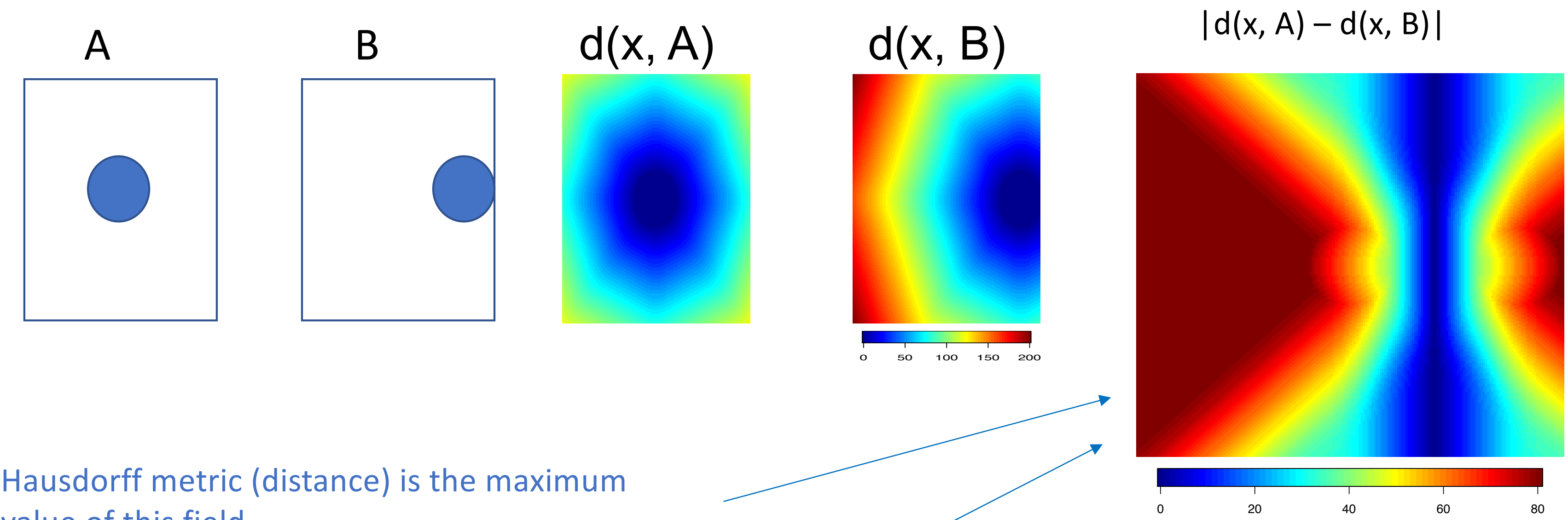
Distance maps for A and B. Note dependence on location within the domain.

Distance maps





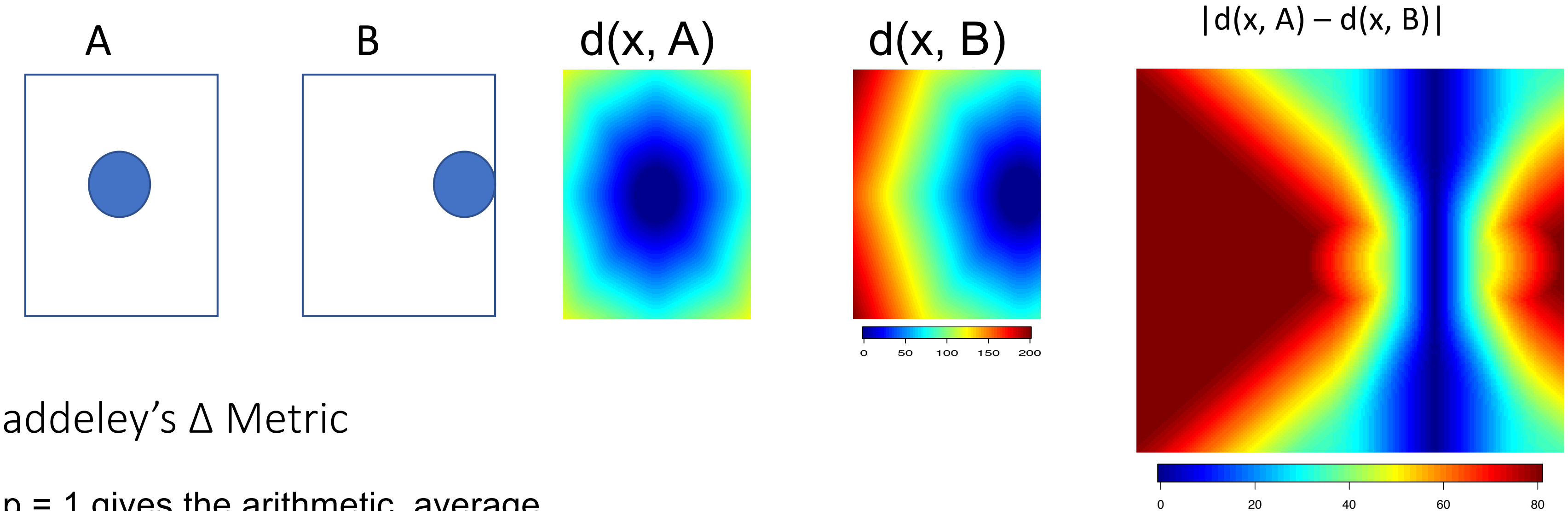
Distance maps



Hausdorff metric (distance) is the maximum value of this field

Baddeley's Δ metric is the L_p norm of this field, where the Hausdorff is the special case that $p = \infty$

Distance maps



Baddeley's Δ Metric

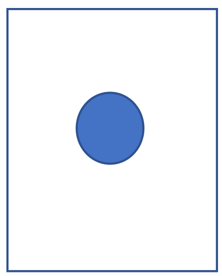
- p = 1 gives the arithmetic average
- p = 2 is the usual choice
- p = ∞ gives the Hausdorff distance

d(x, A) and d(x, B) may first be transformed by a convex function ω.

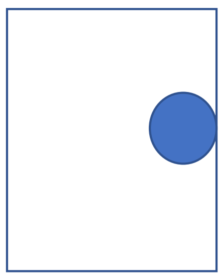
Usually,
ω(x) = max(x, constant),
but the picture here uses "∞" for the constant term.

Distance maps

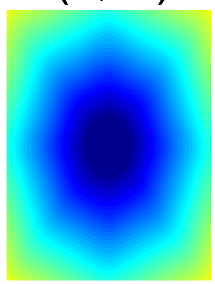
A



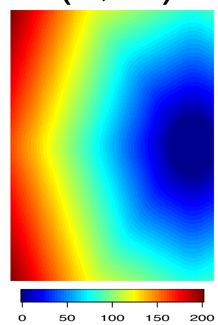
B



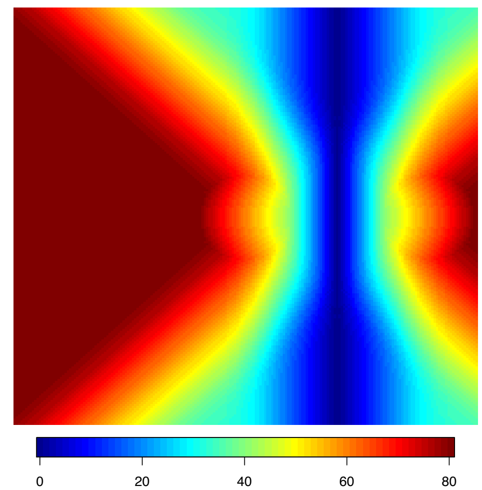
d(x, A)



d(x, B)



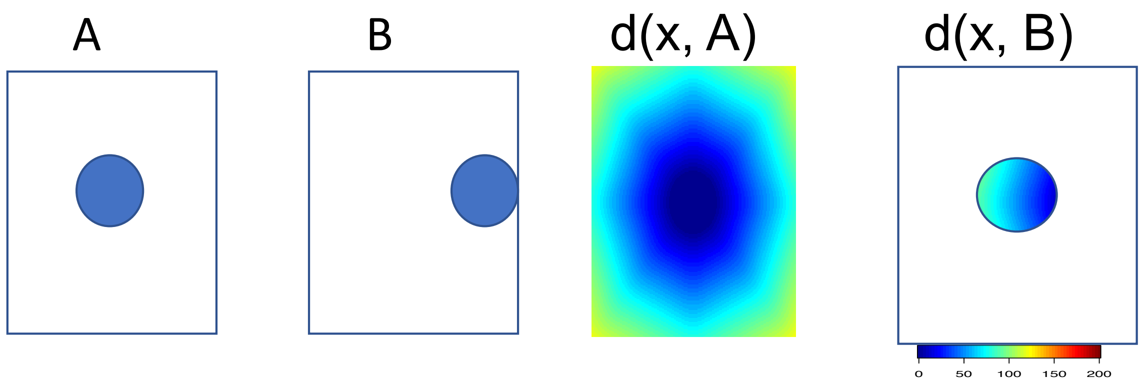
|d(x, A) – d(x, B)|



Baddeley's Δ Metric

$$\Delta = \left[\frac{1}{|\mathcal{D}|}\sum_{s\in\mathcal{D}}\left|\omega\big(d(\boldsymbol{s},A)\big)-\omega\big(d(\boldsymbol{s},B)\big|^{p}\right]^{1/p}$$

Distance maps



Pratt's figure of merit (FoM; not a metric) is given by

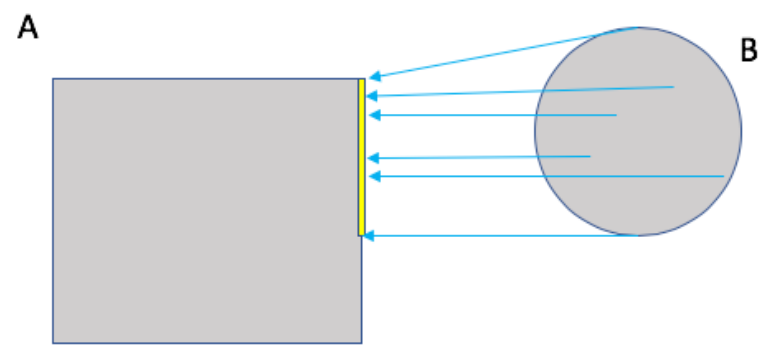
$$\text{FoM}(A,B) = \frac{1}{\max\{n_A, n_B\}} \sum_{\boldsymbol{s} \in B} \frac{1}{1 + \alpha d^2(\boldsymbol{s}, A)}$$

Mean Error Distance (not a metric!)

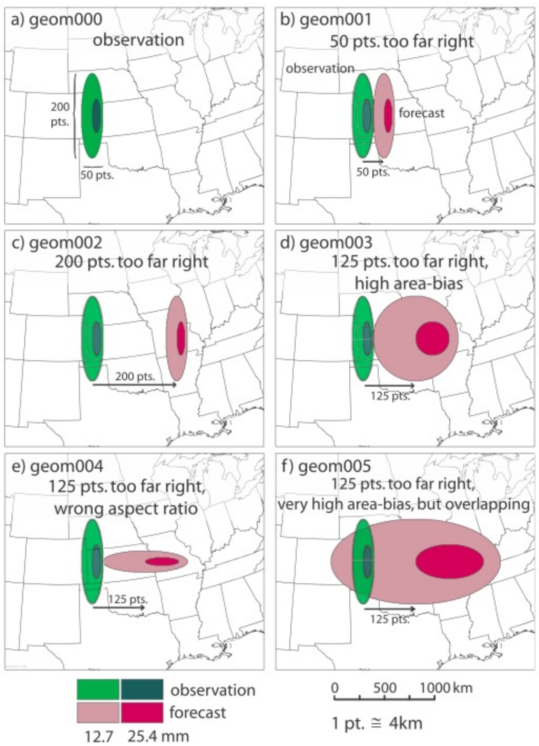
$$\text{MED}(A,B) = \frac{1}{N_B} \sum_{\boldsymbol{s} \in B} d(\boldsymbol{s}, A)$$

Zhu's measure is given by

$$Z(A,B) = \lambda \cdot \sqrt{\sum_{s} \left(\mathbb{I}_A(\boldsymbol{s}) - \mathbb{I}_B(\boldsymbol{s})\right)^2} + (1-\lambda) \cdot \text{MED}(A,B)$$



New Geometric Test Cases



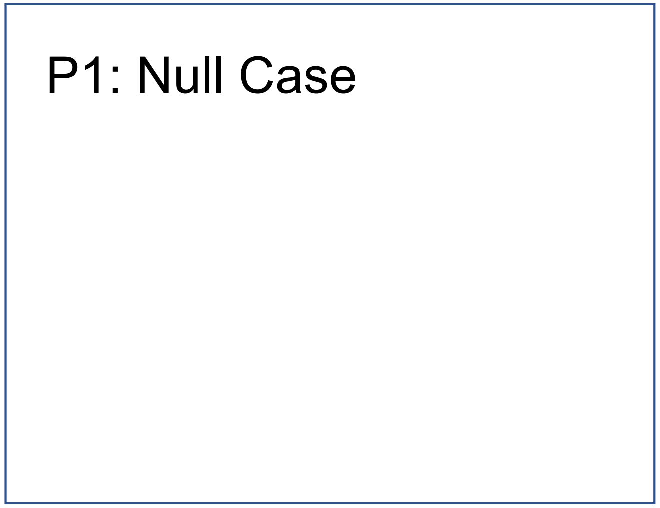
These cases from the ICP were very useful in gleaning information about how spatial methods summarized/ranked different types of forecast situations. They were gridded cases based on Barb Brown's illustration of some of the challenged faced when verifying high-resolution forecasts.

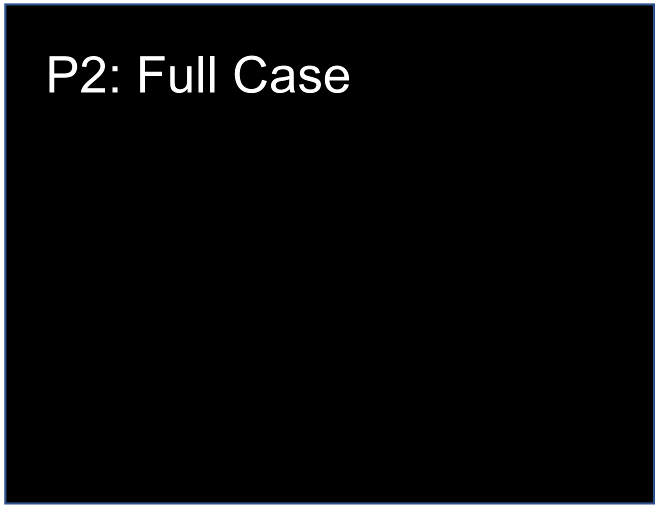
But, since then many new situations have come to light that needed attention.

All subsequent cases are placed on a 200 by 200 grid.

New Geometric Test Cases

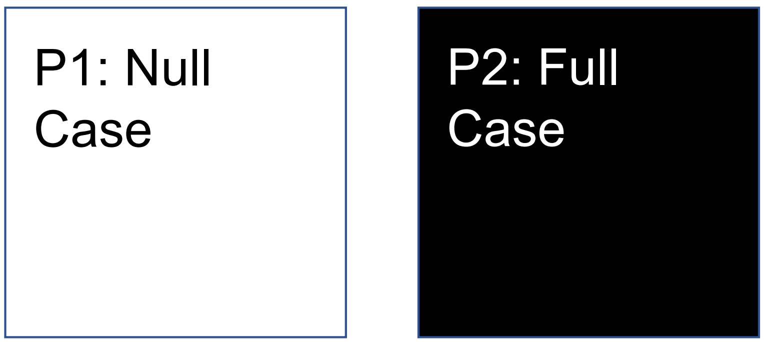
Pathological Cases



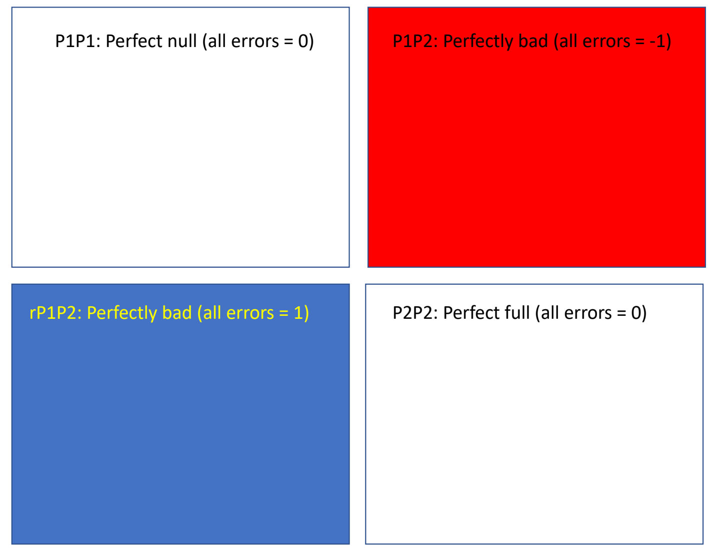


New Geometric Test Cases

Pathological



Δ (P1, P1) = H(P1, P1) = CD(P1, P1) = Z(P1,P1) = MED(P1, P1) = FoM(P1, P1) = undefined

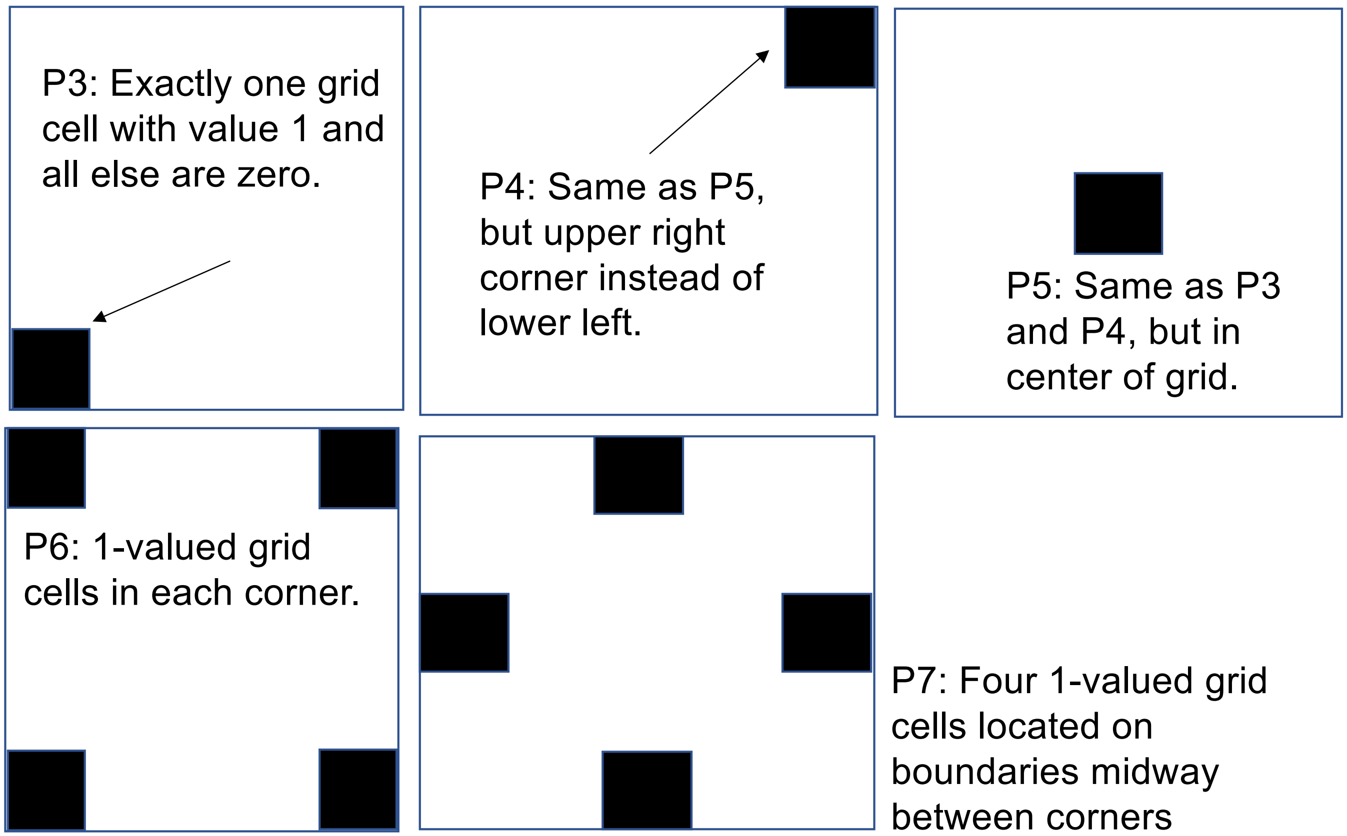


Remaining cases are essentially undefined for all measures. Can define distance map to be infinity for P1 (or the largest possible distance) to get a real value for most of them, but then they are all highly sensitive to a small change in the field, and such a value is somewhat arbitrary.

Δ(P2, P2) = H(P2, P2) = CD(P2, P2) = MED(P2, P2) = Z(P2, P2) = 0.00, FoM(P2, P2) = 1.00

New Geometric Test Cases

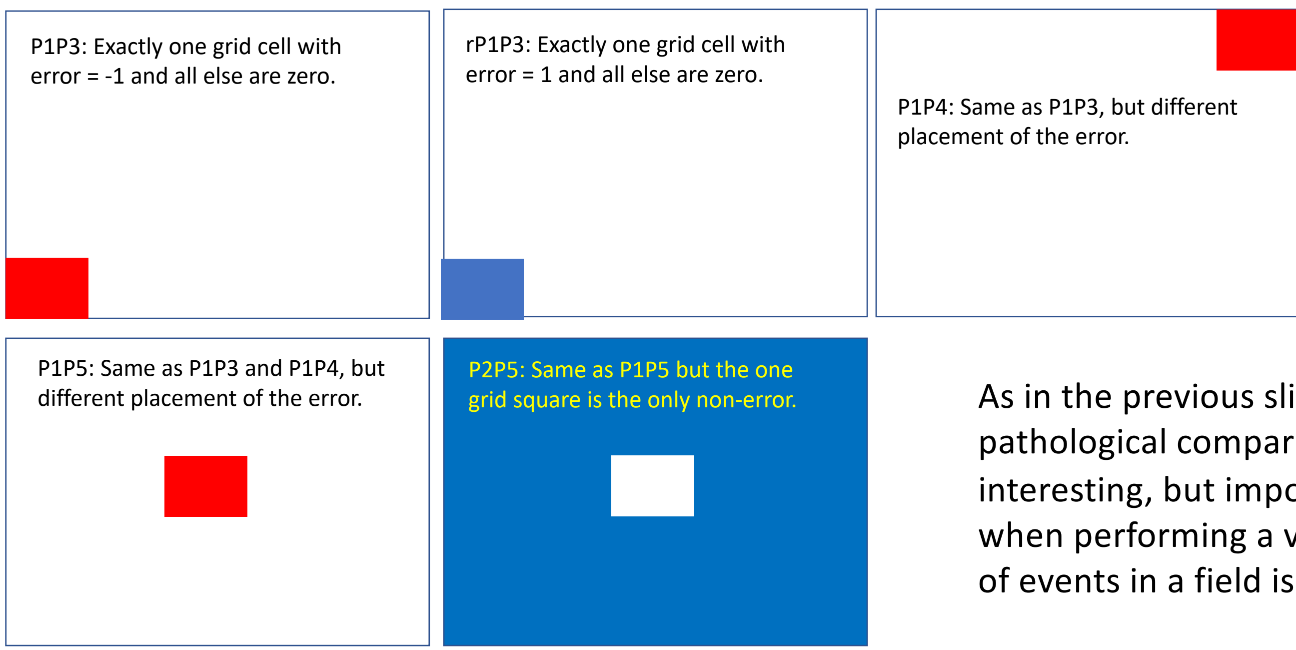
Pathological Cases



Centroid for P6 and P7 is the same, so CD(P6, P7) = 0 (perfect score!), but CD(P3, P6) = CD(P3, P7) is large.

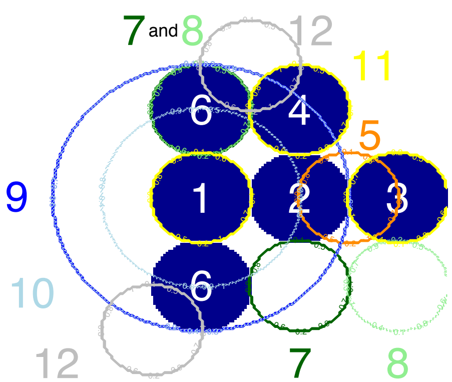
New Geometric Test Cases

Pathological Cases



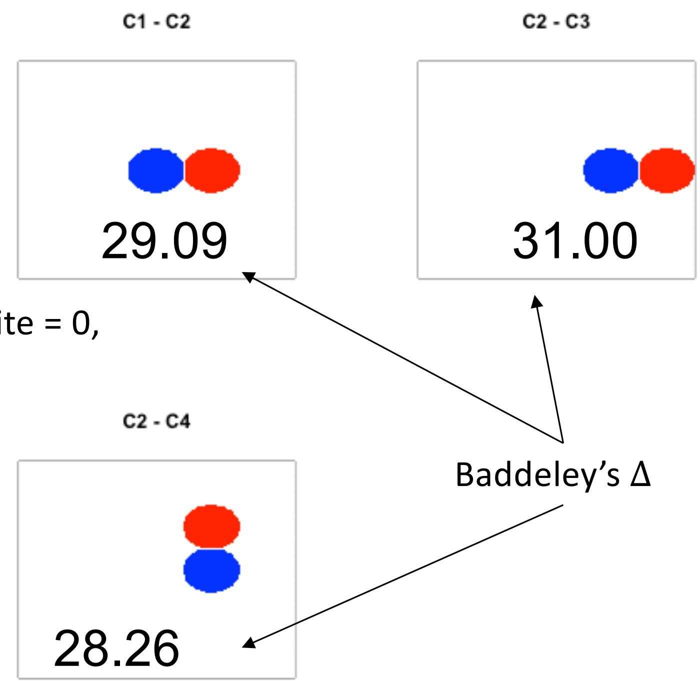
As in the previous slide, the remaining pathological comparisons are not very interesting, but important cases to consider when performing a verification. The number of events in a field is important!

New Geometric Test Cases



Circle Cases

New Geometric Test Cases



red = -1, white = 0,
blue = 1

Baddeley's Δ

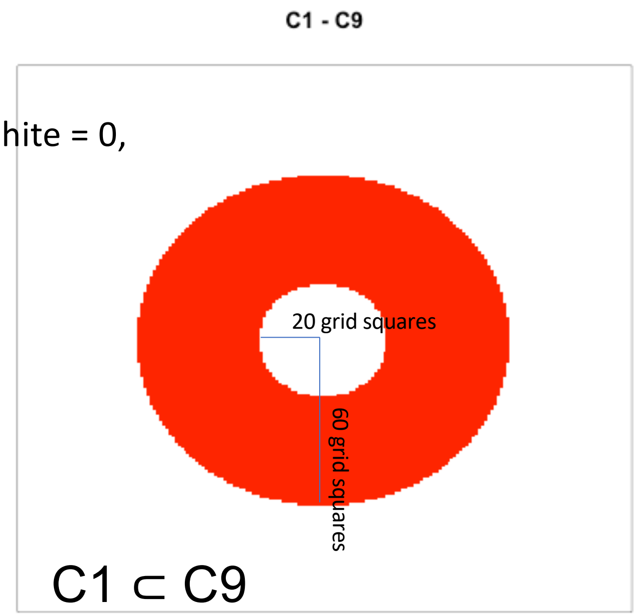
MED and FoM are symmetric here because the circles, A and B, are the same size and shape as each other

MED(A,B) = MED(B,A) = 21.92
FoM(A,B) = FoM(B,A) = 0.07

H(A,B) = 40.20
CD(A,B) = 40.00
Z(A,B) = 36.81

New Geometric Test Cases

red = -1, white = 0,
blue = 1

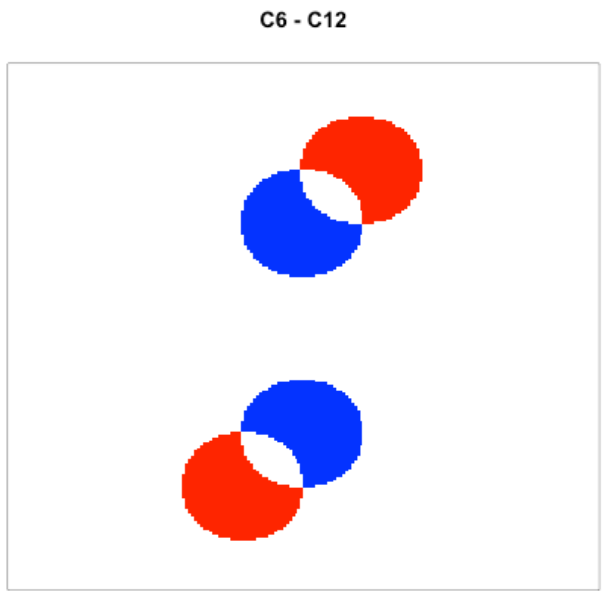


C1 – C9

Δ(C1, C9) = 38.13
H(C1,C9) = 43.43
CD(C1, C9) = 0.00
Z(C1, C9) = 50.5

MED(C1, C9) = 21.72
MED(C9, C1) = 0.00
FoM(C1, C9) = 0.12
FoM(C9, C1) = 0.18

New Geometric Test Cases



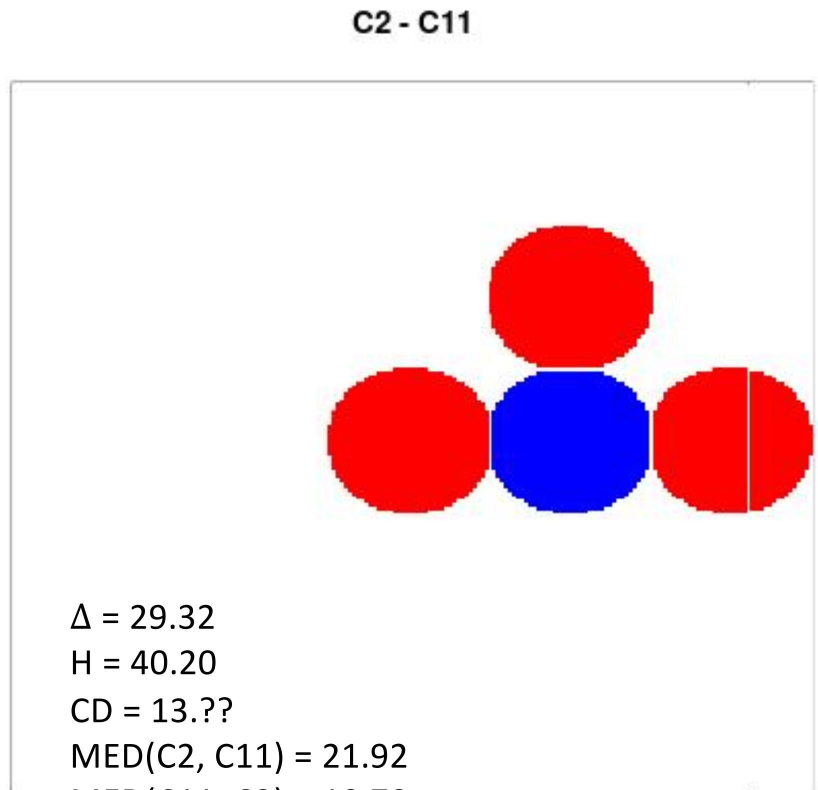
Δ = 18.84
H = 28.43
CD = 0.00
Z = 38.36

MED(C6, C12) = MED(C6, C12) = 11.24
FoM(C6, C12) = FoM(C6, C12) = 0.32

C6 – C12

red = -1, white = 0,
blue = 1

New Geometric Test Cases

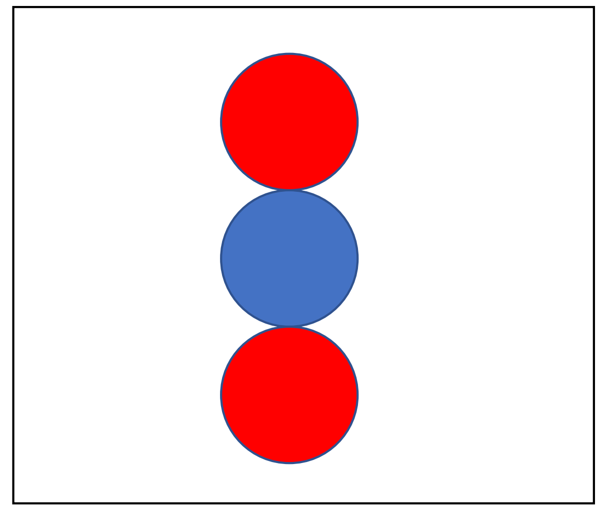


Δ = 29.32
H = 40.20
CD = 13.??
MED(C2, C11) = 21.92
MED(C11, C2) = 10.70
Z(C2, C11) = 41.84
Z(C11, C2) = 47.45

red = -1, white = 0, blue = 1

New Geometric Test Cases

C1 – C6 (illustration)



Δ = 24
H = 40.20
CD = 0
MED(C2, C6) = 22
MED(C6, C2) = 13
Z(C2, C6) = 43
Z(C6, C2) = 38

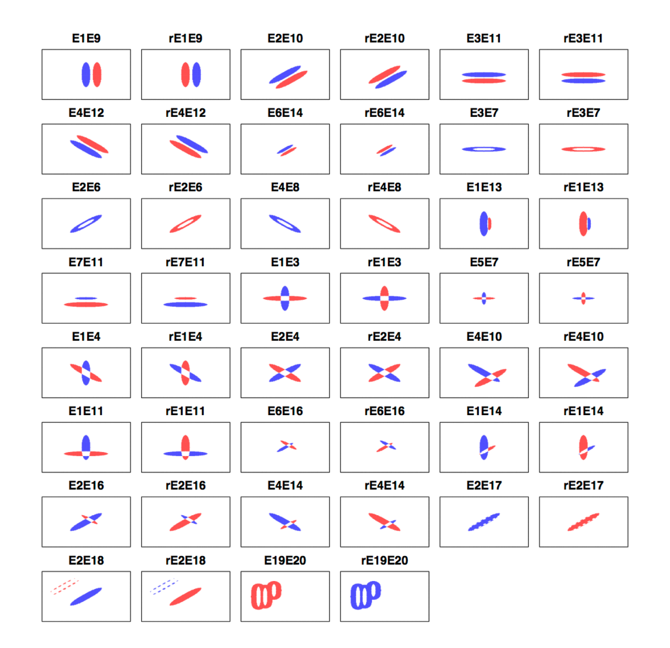
red = -1, white = 0, blue = 1

MED, FoM, Z

Modifications to these measures that are metrics

- avg MED(A, B) = $\frac{1}{2}\left(\mathrm{MED}(A,B)+\mathrm{MED}(B,A)\right)$
- min MED(A, B) = $\min\{\mathrm{MED}(A,B),\mathrm{MED}(B,A)\}$
- max MED(A, B) = $\max\{\mathrm{MED}(A,B),\mathrm{MED}(B,A)\}$

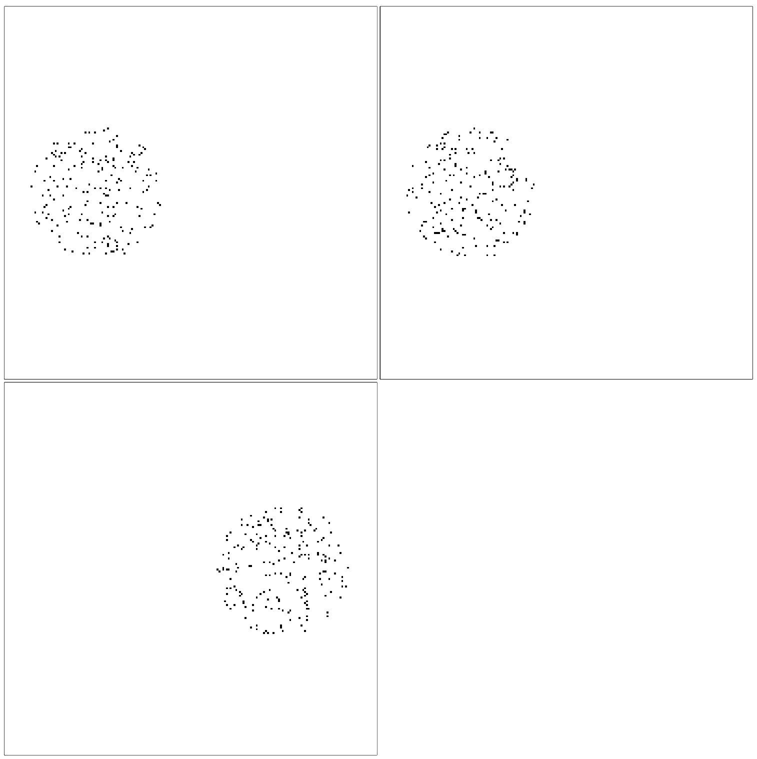
	$\Delta_{p=2,w=\infty}$	MED	rMED	Avg MED	dFSS
C1C2	29.0 (7.5)	22.0 (4.5)	22.0 (9.0)	22.0 (9.0)	34.0 (7.5)
C12C3	31.0 (9.0)	22.0 (4.5)	22.0 (9.0)	22.0 (9.0)	32.0 (5.0)
C2C4	28.0 (6.0)	22.0 (4.5)	22.0 (9.0)	22.0 (9.0)	34.0 (7.5)
C1C4	41.0 (15.0)	38.0 (14.5)	38.0 (14.5)	38.0 (14.5)	54.0 (11.0)
C3C4	38.0 (13.0)	38.0 (14.5)	38.0 (14.5)	38.0 (14.5)	48.0 (10.0)
C2C5	15.0 (1.0)	5.7 (1.5)	5.7 (2.5)	5.7 (1.5)	18.0 (3.0)
C3C5	16.0 (2.0)	5.7 (1.5)	5.7 (2.5)	5.7 (1.5)	16.0 (1.0)
C2C11	29.0 (7.5)	22.0 (4.5)	11.0 (5.0)	16.5 (6.0)	-
C1C6	24.0 (5.0)	22.0 (4.5)	13.0 (7.0)	17.5 (7.0)	-
C6C7	22.0 (4.0)	11.0 (3.5)	11.0 (5.0)	11.0 (4.0)	17.0 (2.0)
C6C8	35.0 (11.0)	31.0 (11.0)	28.0 (11.5)	30.0 (12.0)	33.0 (6.0)
C1C9	38.0 (13.0)	22.0 (4.5)	0.0 (1.0)	11.0 (4.0)	-
C1C10	38.0 (13.0)	32.0 (12.0)	28.0 (11.5)	30.0 (12.0)	-
C6C12	19.0 (3.0)	11.0 (3.5)	11.0 (5.0)	11.0 (4.0)	25.0 (4.0)
C13C14	32.0 (10.0)	35.0 (13.0)	35.0 (13.0)	35.0 (13.0)	43.0 (9.0)
1st tercile	2nd tercile	3rd tercile			



Other cases

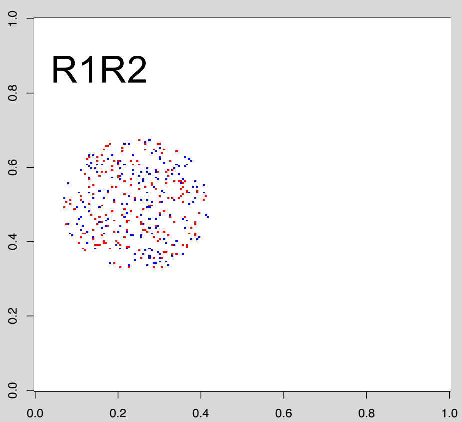
Complex Terrain Cases

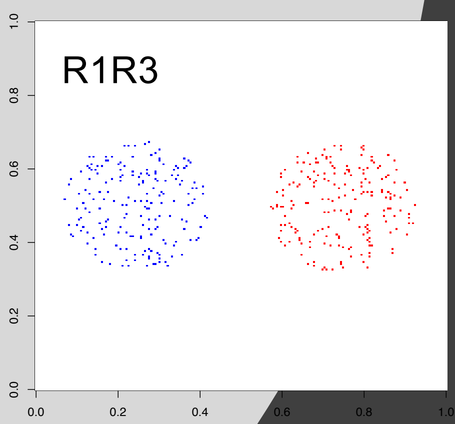
Other Cases



Random Rain Cases

Δ = 1.91
H = 9.12
CD = 1.30,
Z(R1, R2) = 11
Z(R2, R1) = 10
MED(R1, R2) = 2.37
MED(R2, R1) = 2.56





Δ = 63.39
H = 104.42
CD = 99.11,
Z = 45
FoM = 0.00
MED(R1, R3) = 70.04
MED(R3, R1) = 70.60

Other Cases

Random Rain Cases

Other cases

Additional Cases include:

- Holes (inverted C1 and C2)
- C1C4 with noise added
- C1C4 with P3 added
- C1C4 with P5 added

Summary

- Distance-based measures generally give similar information
  - Each has its caveats
- None handle pathological (but very common) situations very well
  - Keep track of the numbers of events in each field for later analysis of results
  - Consider what the best way to handle such cases is for specific purposes
- Centroid and Hausdorff distance give average translation errors
  - Centroid can give a perfect score in some situations where the comparisons are otherwise very different, less than perfect if two are close but slightly different
  - Hausdorff distance is highly sensitive to outliers
- MED/Baddeley give an average distance
  - Less sensitive to outliers
  - MED is insensitive to certain biases (Use frequency bias as a complementary measure!)
  - MED, Zhu and FoM can give a spatial form of false alarm v. miss information
  - FoM is unitless, so gives a better notion of good v. bad in that sense
  - Baddeley's delta metric is somewhat affected by domain, position within domain, etc., but not terribly so
- Zhu's measure accounts for bias directly
  - Together with MED, it gives a picture of MED and how much bias affects the result because it is an average of an overlap term with MED.

Thank you!

Eric Gilleland

Research Applications Laboratory

National Center for Atmospheric Research

https://ral.ucar.edu/staff/ericg/