Eric Gilleland

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

Research Applications Laboratory, National Center for Atmospheric Research,

Boulder, Colorado, U.S.A.

NCAR

NSF

DTC Developmental Testbed Center

A Comparison of Spatial Dissimilarity Measures

Dod Cloud Worskshop

14 September 2023

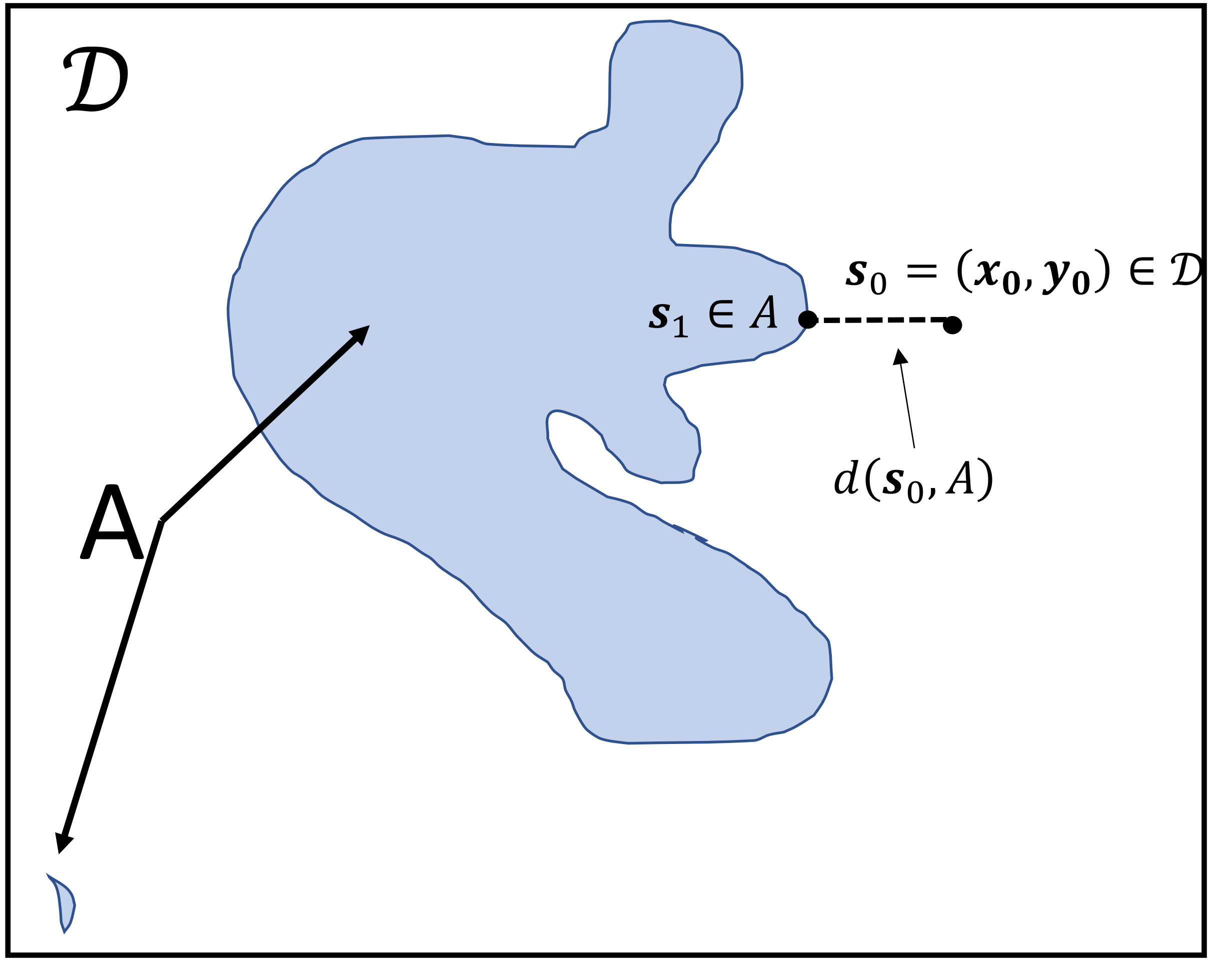
13 – 14 September 2023

Spatial Dissimilarity Measures

This talk mostly covers these papers (particularly the ones highlighted):

- Ahijevych, D., E. Gilleland, B.G. Brown, and E.E. Ebert, 2009. Application of spatial verification methods to idealized and NWP gridded precipitation forecasts. *Weather Forecast.*, **24** (6), 1485 - 1497, doi: 10.1175/2009WAF2222298.1.
- Baddeley, A. J., 1992. An error metric for binary images. Robust Computer Vision Algorithms, W. Forstner and S. Ruwiedel, Eds., Wichmann, 59–78.
- Gilleland, E., 2011. Spatial Forecast Verification: Baddeley's Delta Metric Applied to the ICP Test Cases. *Weather Forecast.*, **26** (3), 409 - 415, doi: 10.1175/WAF-D-10-05061.1.
- 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.
- Gilleland, E., 2021. Novel measures for summarizing high-resolution forecast performance. *Advances in Statistical Climatology, Meteorology and Oceanography*, **7** (1), 13 - 34, doi: 10.5194/ascmo-7-13-2021.
- Gilleland, E., 2022. Comparing spatial fields with SpatialVx: Spatial forecast verification in R. *Unpublished.* doi: 10.5065/4px3-5a05.
- Gilleland, E., D. Ahijevych, B.G. Brown, B. Casati, and E.E. Ebert, 2009. Intercomparison of Spatial Forecast Verification Methods. *Weather Forecast.*, **24**, 1416 - 1430, doi: 10.1175/2009WAF2222269.1.
- Gilleland, E., T.C.M. Lee, J. Halley Gotway, R.G. Bullock, and B.G. Brown, 2008. Computationally efficient spatial forecast verification using Baddeley's Δ image metric. *Mon. Wea. Rev.* **136** (5), 1747 - 1757, doi: 10.1175/2007MWR2274.1.
- Gilleland, E., G. Skok, B. G. Brown, B. Casati, M. Dorninger, M. P. Mittermaier, N. Roberts, and L. J. Wilson, 2020. A novel set of verification test fields with application to distance measures. *Mon. Wea. Rev.*, **148** (4), 1653 - 1673, doi: 10.1175/MWR-D-19-0256.1.

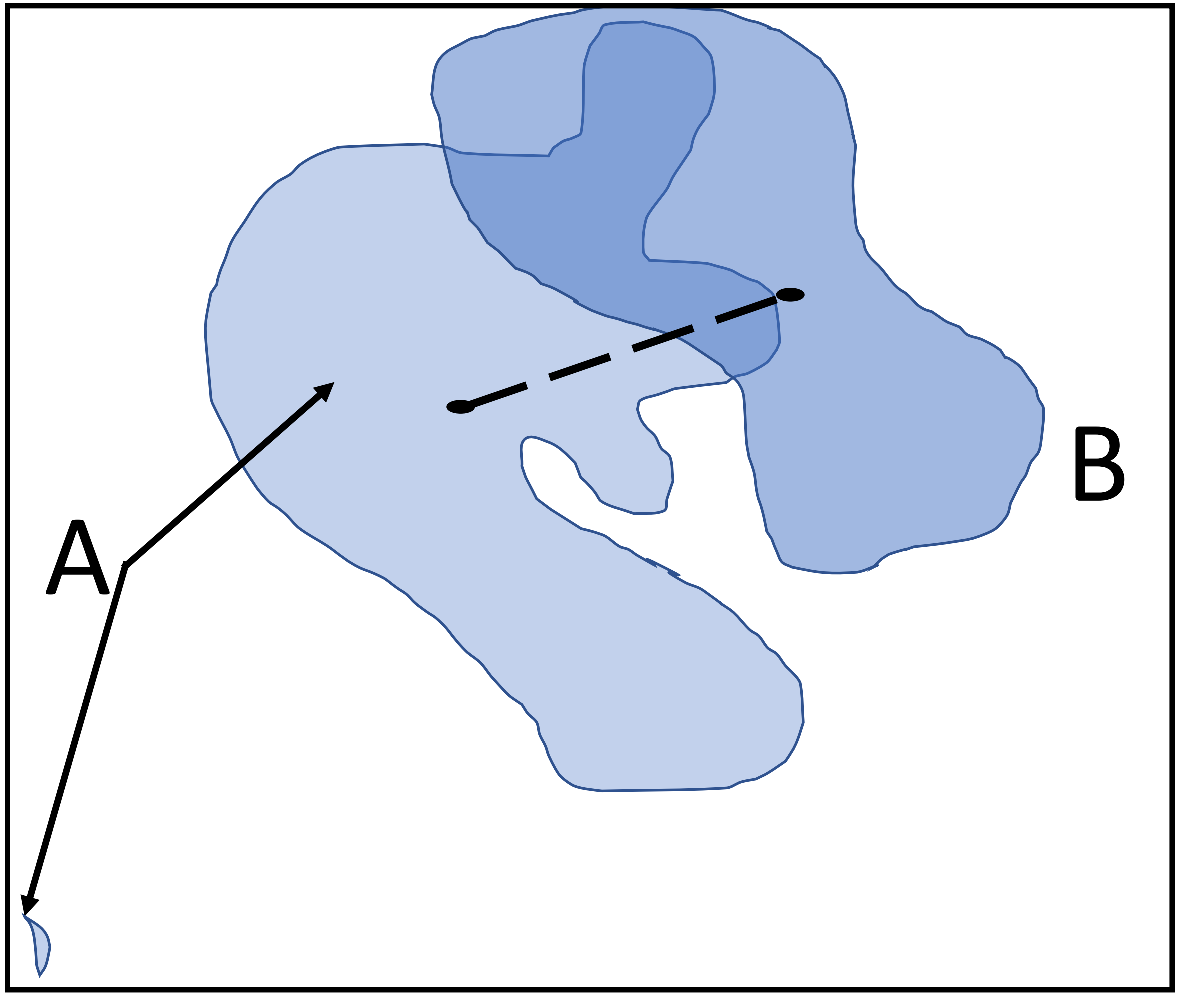
Spatial Dissimilarity Measures



Relatively easy to define a measure of (dis)similarity between a single point, $\boldsymbol{s}_0$, in the domain and a set of points, A, in the domain.

Here, the shortest distance from $\boldsymbol{s}_0$ to the *nearest* point in A is used, and called $d(\boldsymbol{s}_0, A)$.

Spatial Dissimilarity Measures

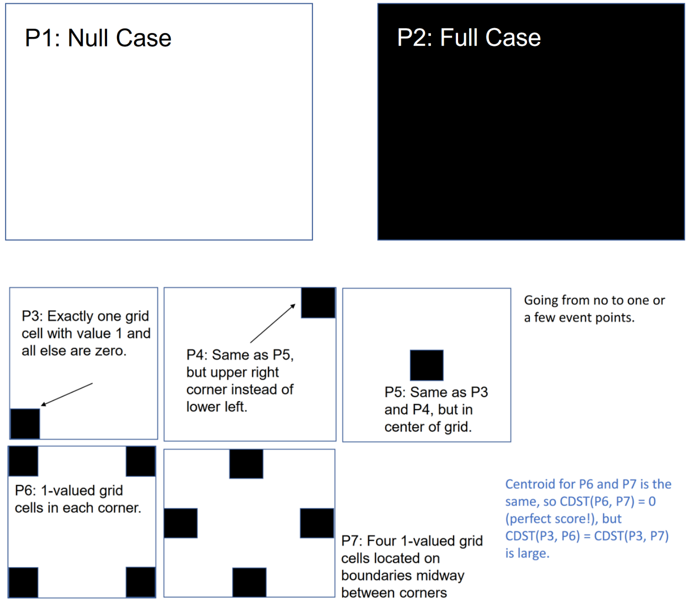


Considerably more challenging to identify a useful summary measure of (dis)similarity between two sets of points.

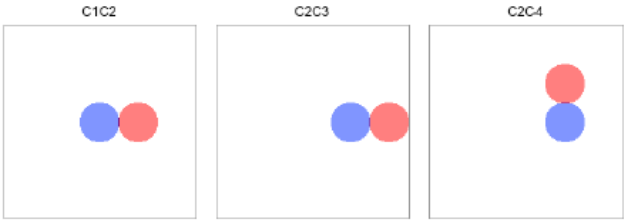
For example, when comparing a (gridded) forecast field against a (gridded) observation field.

Some Test Cases

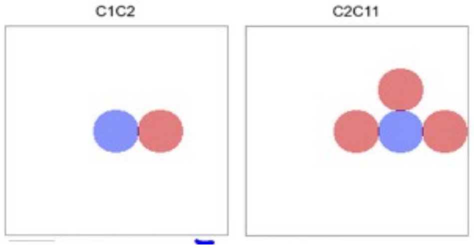
Pathologoical



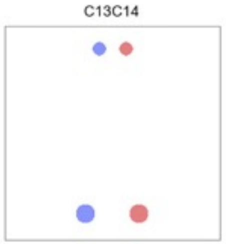
Positional and Boundary



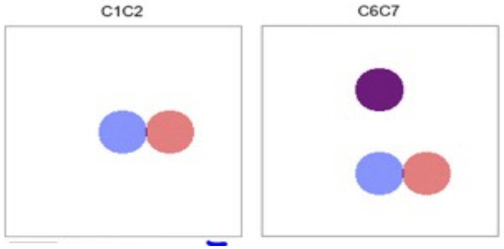
Frequency Bias



Rare Events

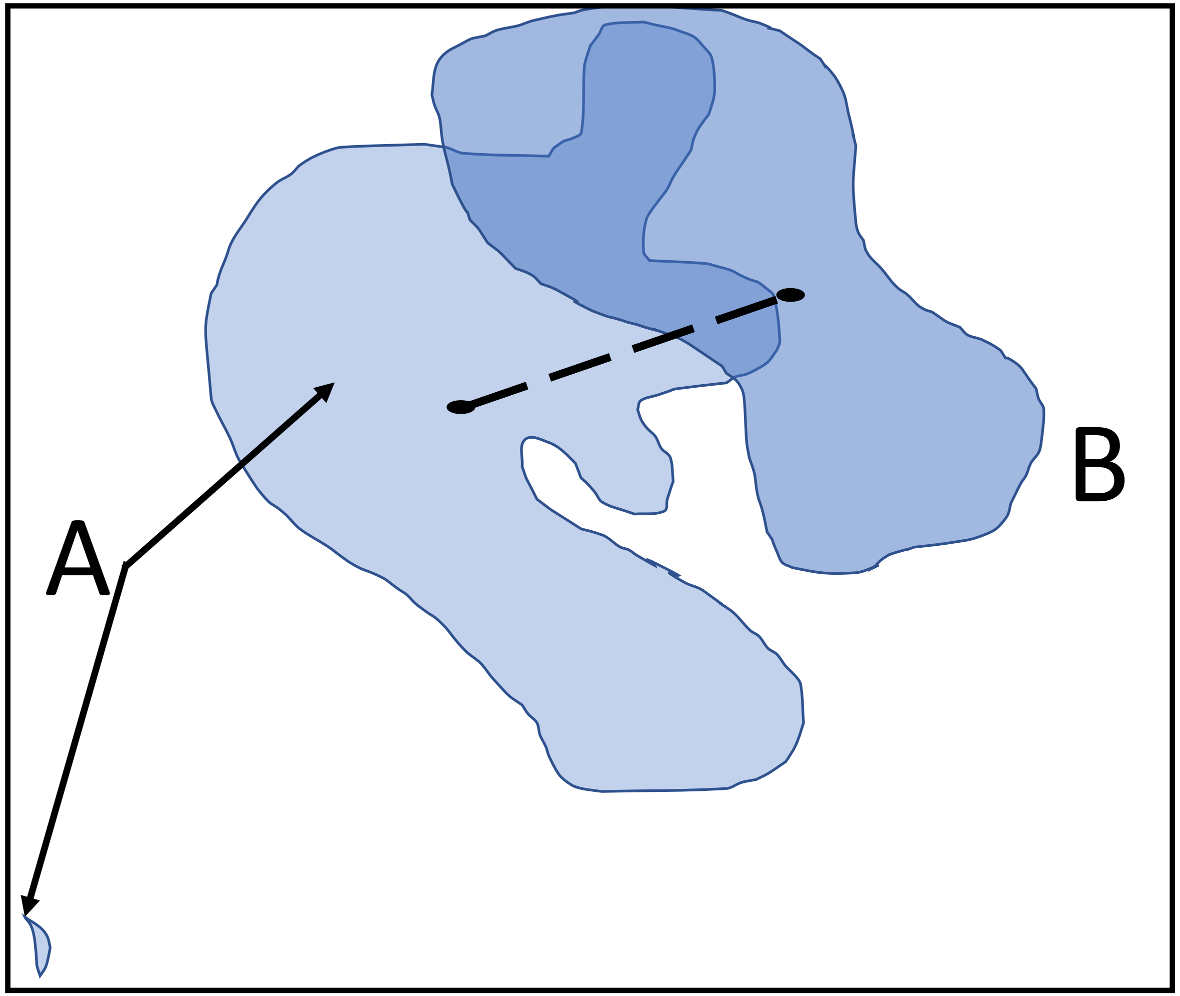


Partially Perfect Match



Subset of test cases from Gilleland (2017) and Gilleland et al (2020)

Spatial Dissimilarity Measures

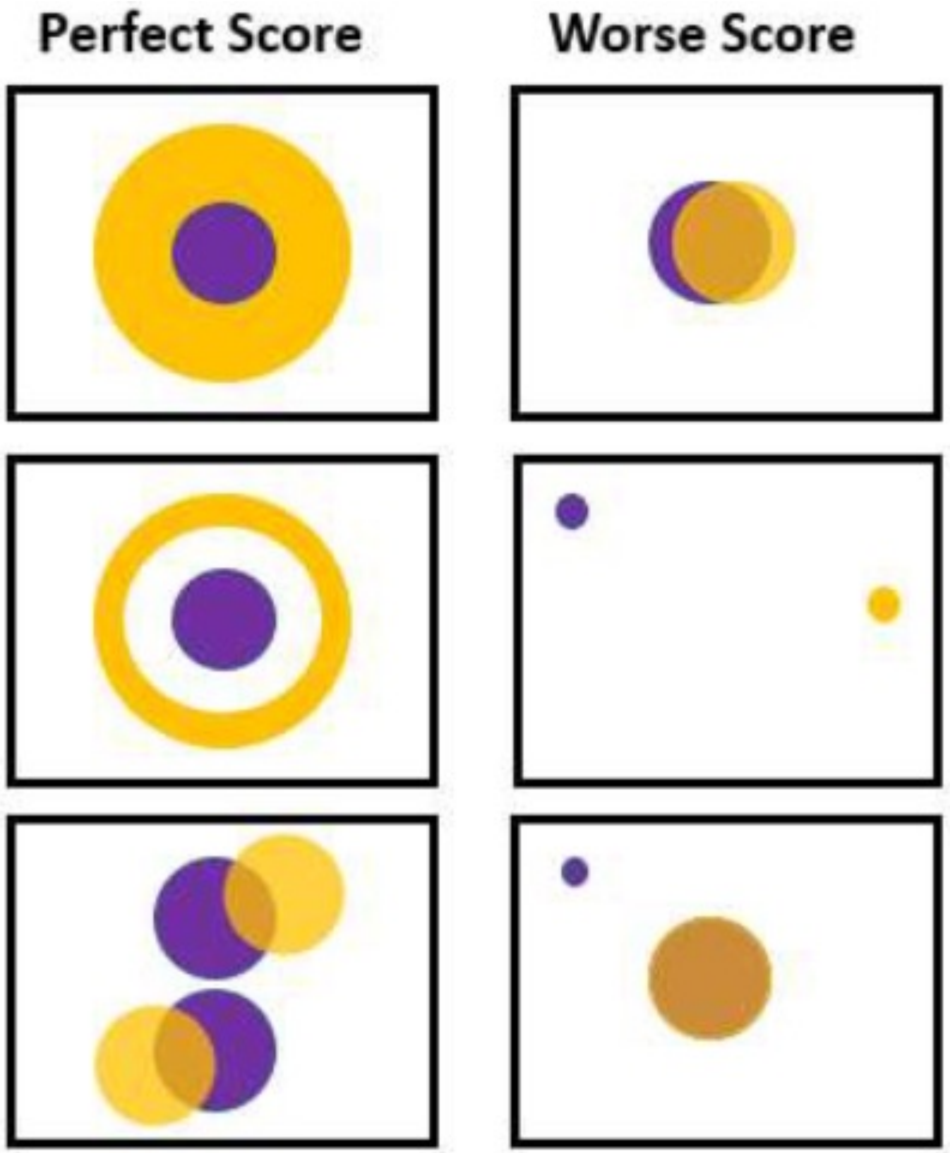


Centroid Distance is the distance between the centers of mass of the two sets.

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

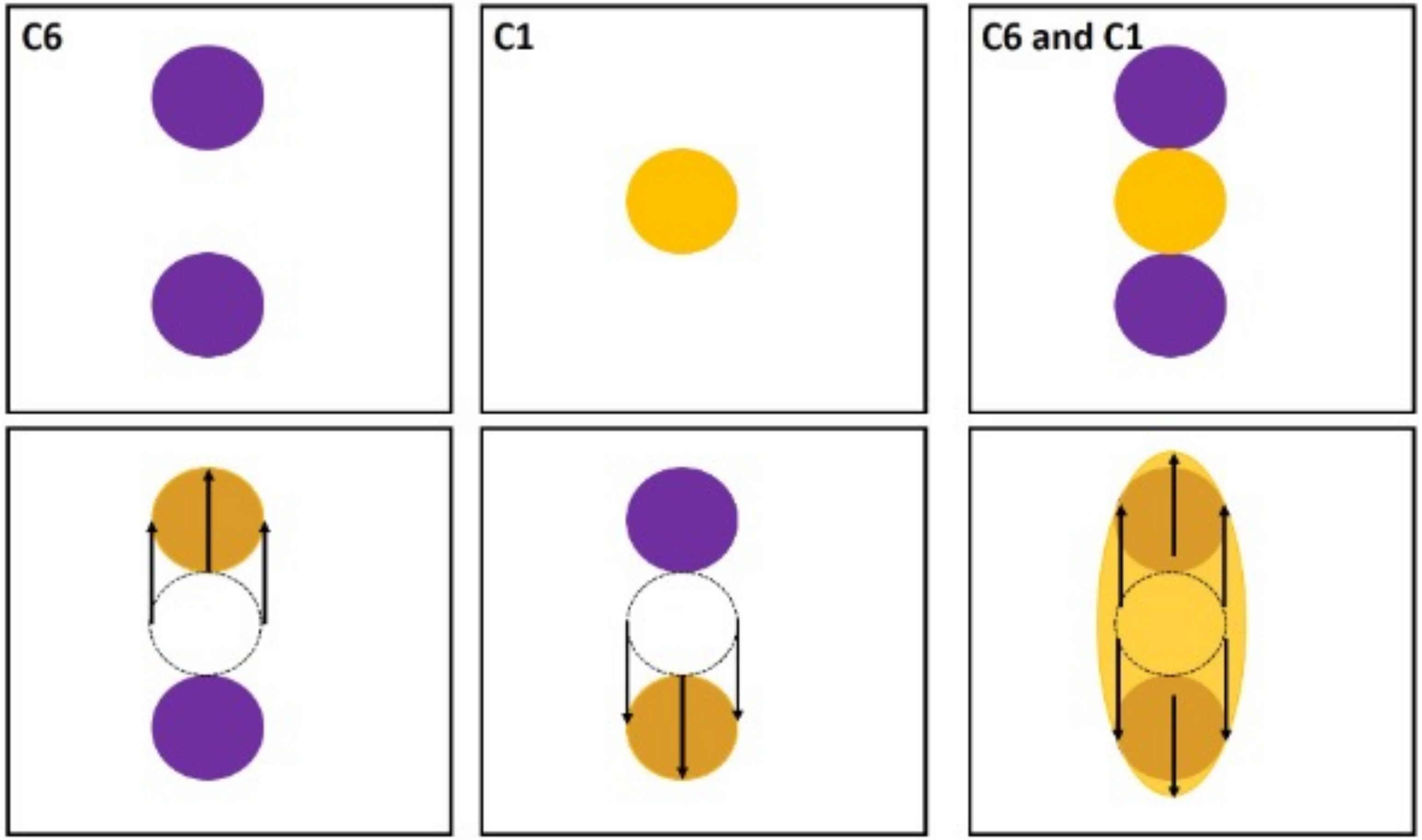
Replace $I(\cdot)$ with $Z(\cdot)$ if the field is not binary

Spatial Dissimilarity Measures: centroid distance



Spatial Dissimilarity Measures

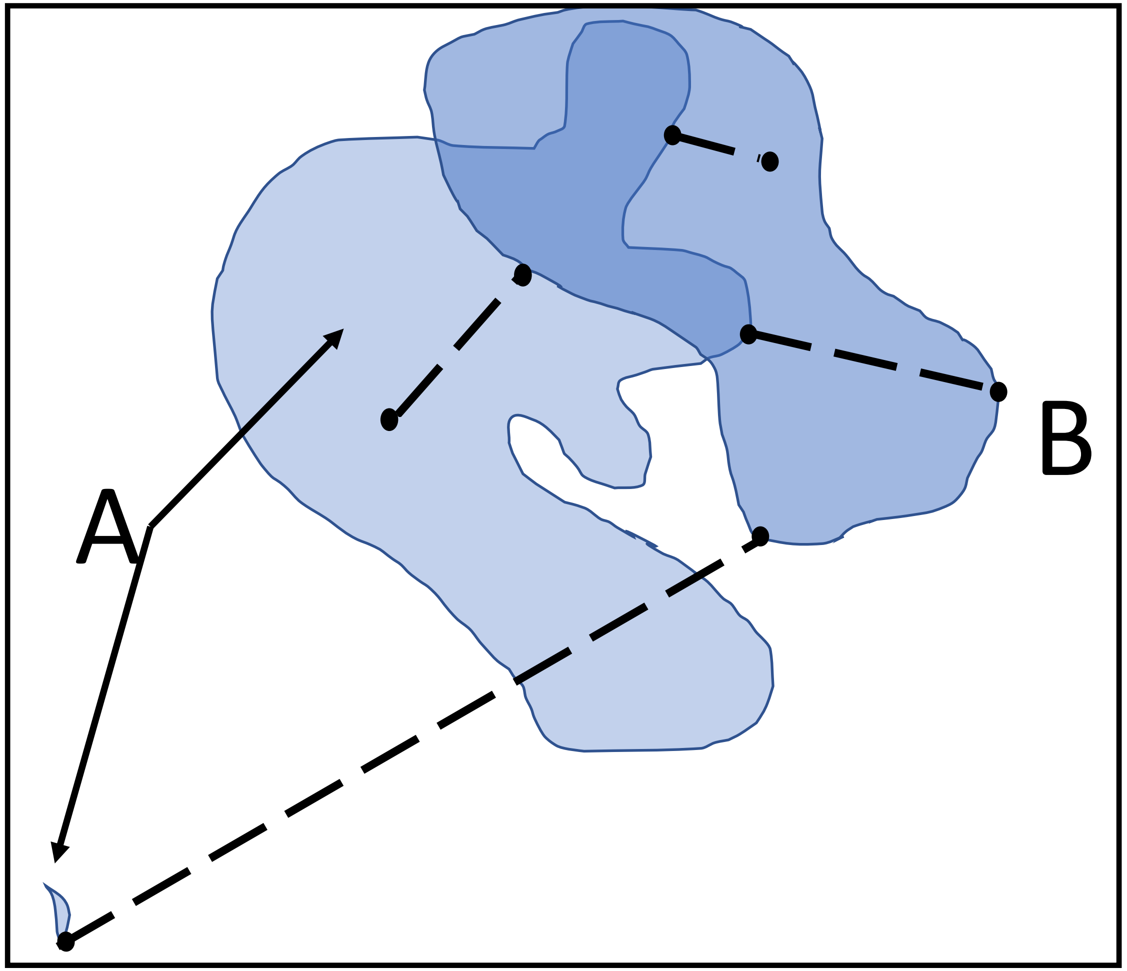
A more complicated solution is to either deform one field until it is better aligned with the other (e.g., image warping), or identify features and merge/match them across fields.



A (dis)similarity measure provides an overall summary of the quality of the match between the two fields, but no single summary measure provides all the information.

They can be applied to the entire field or to individual features within the fields.

Spatial Dissimilarity Measures



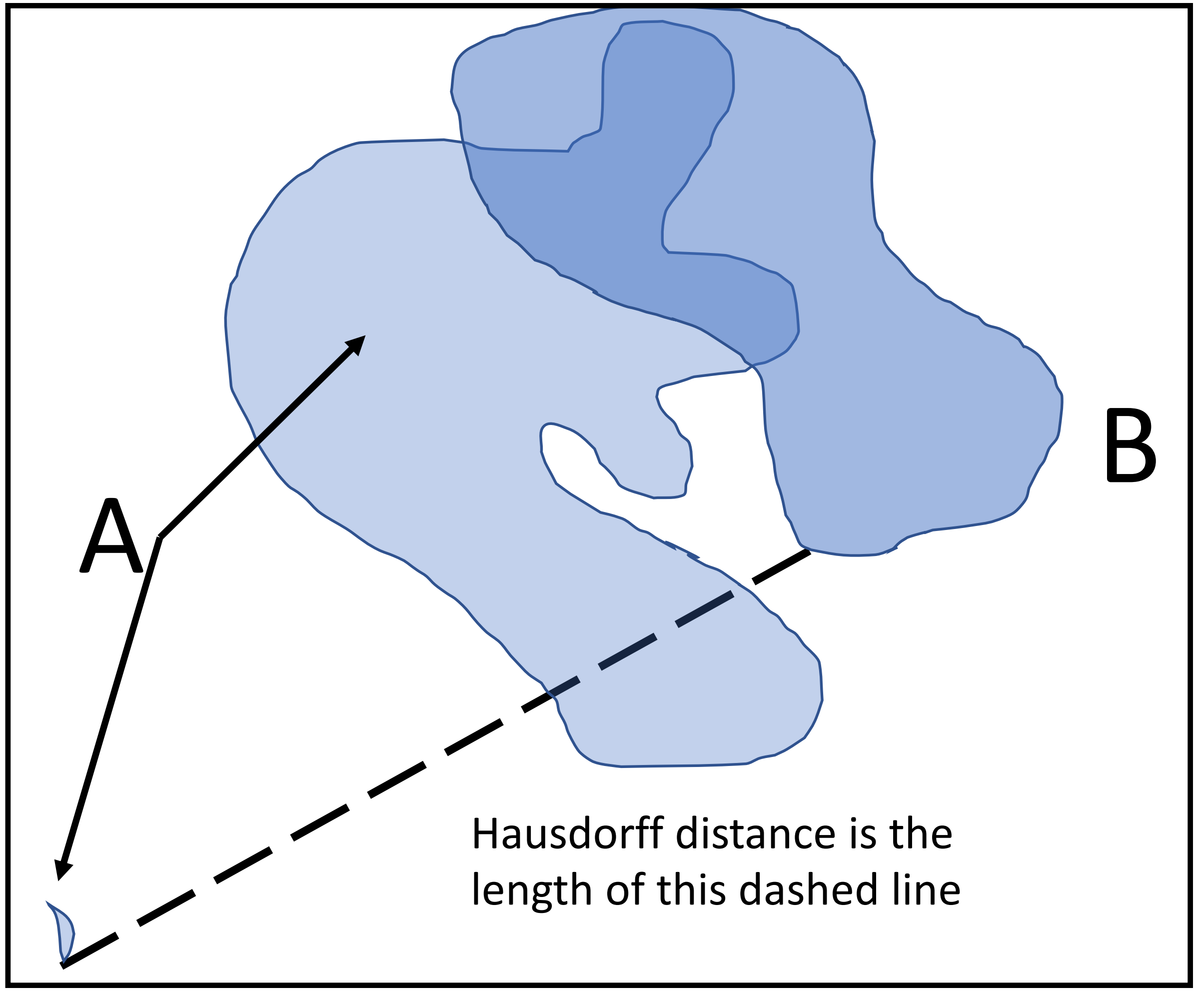
Calculate $d(\boldsymbol{s}, A)$ from each grid point $\boldsymbol{s} \in B$. Also find $d(\boldsymbol{s}, B)$ from each grid point $\boldsymbol{s} \in A$.

Baddeley's Delta Metric (Baddeley 1992) is a type of average of these distances (the L_p norm) given by

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

where N is the total number of grid points, p is a user chosen parameter (typical choice is $p = 2$ giving the Euclidean distance), and $\omega(\cdot)$ is a concave function, $\omega(t + u) \leq \omega(t) + \omega(u)$, usually taken to be $\omega(x) = \max(x, c)$ for some chosen constant c.

Spatial Dissimilarity Measures



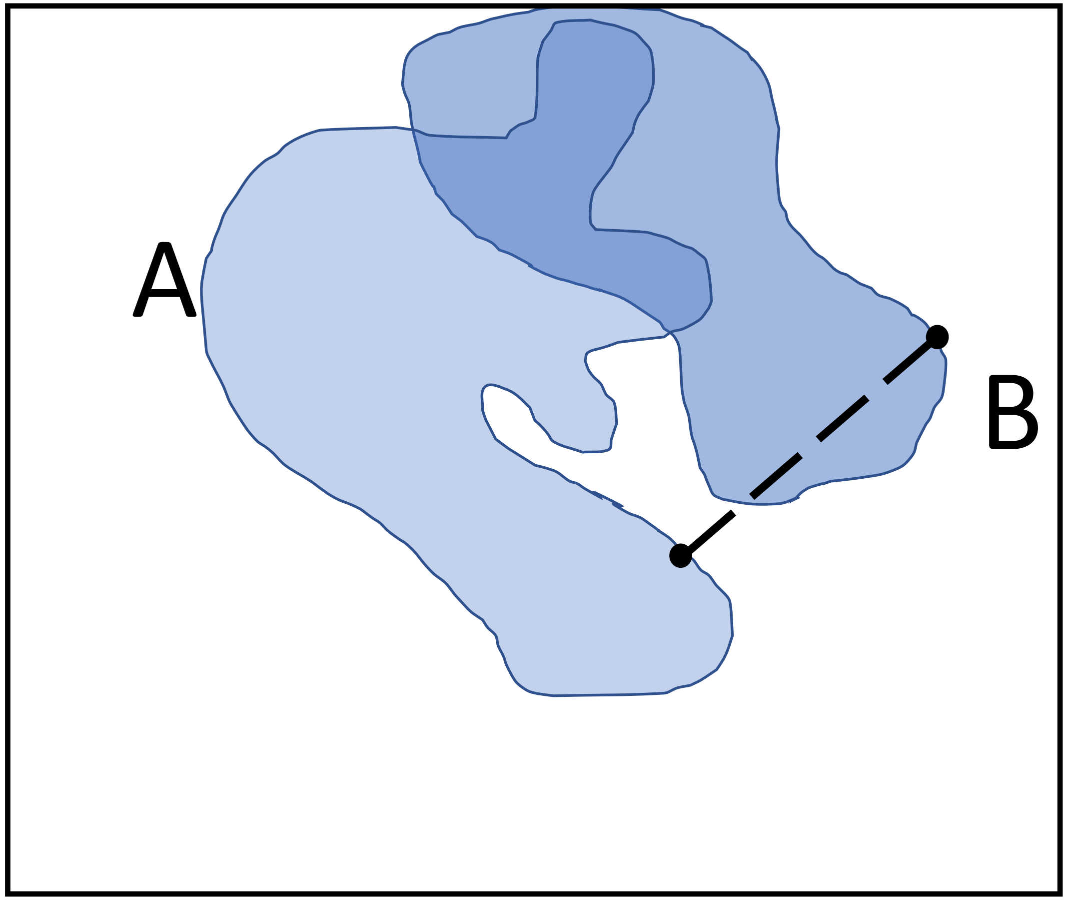
Calculate $d(\boldsymbol{s}, A)$ from each grid point $\boldsymbol{s} \in B$.
Also find $d(\boldsymbol{s}, B)$ from each grid point $\boldsymbol{s} \in A$.

The maximum of all of these distances gives the Hausdorff distance

$$H(A,B) = \max\left\{\max_{s \in B} d(\boldsymbol{s}, A), \max_{s \in A} d(\boldsymbol{s}, B)\right\}$$

(e.g., Baddeley 1992)

Spatial Dissimilarity Measures



The Hausdorff distance can be highly sensitive to small changes in the sets A and B.

So, perhaps not so useful in the context of cloud forecast verification. Very useful when interest is in small-spatial-scale severe weather.

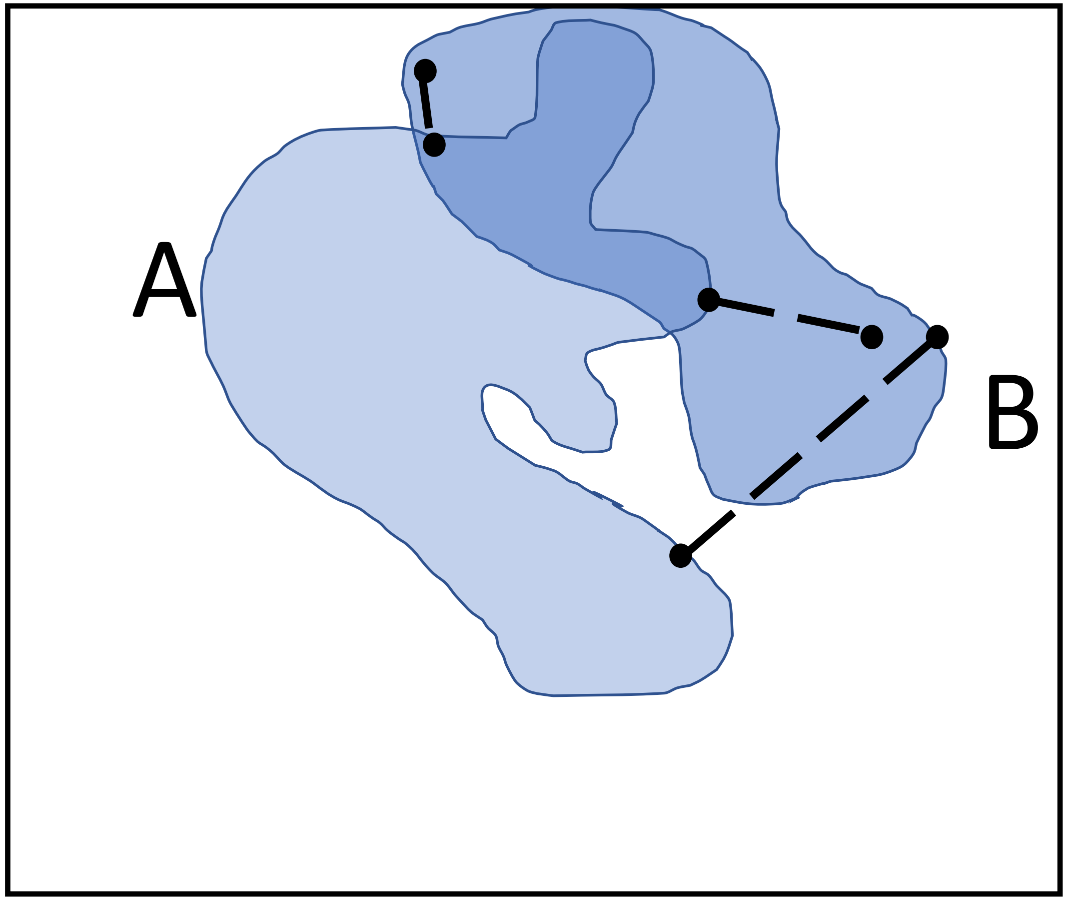
The maximum of all of these distances gives the

Hausdorff distance

$$H(A,B) = \max\left\{\max_{\boldsymbol{s}\in B} d(\boldsymbol{s},A), \max_{\boldsymbol{s}\in A} d(\boldsymbol{s},B)\right\}$$

(e.g., Baddeley 1992)

Spatial Dissimilarity Measures



The mean-error distance is given by

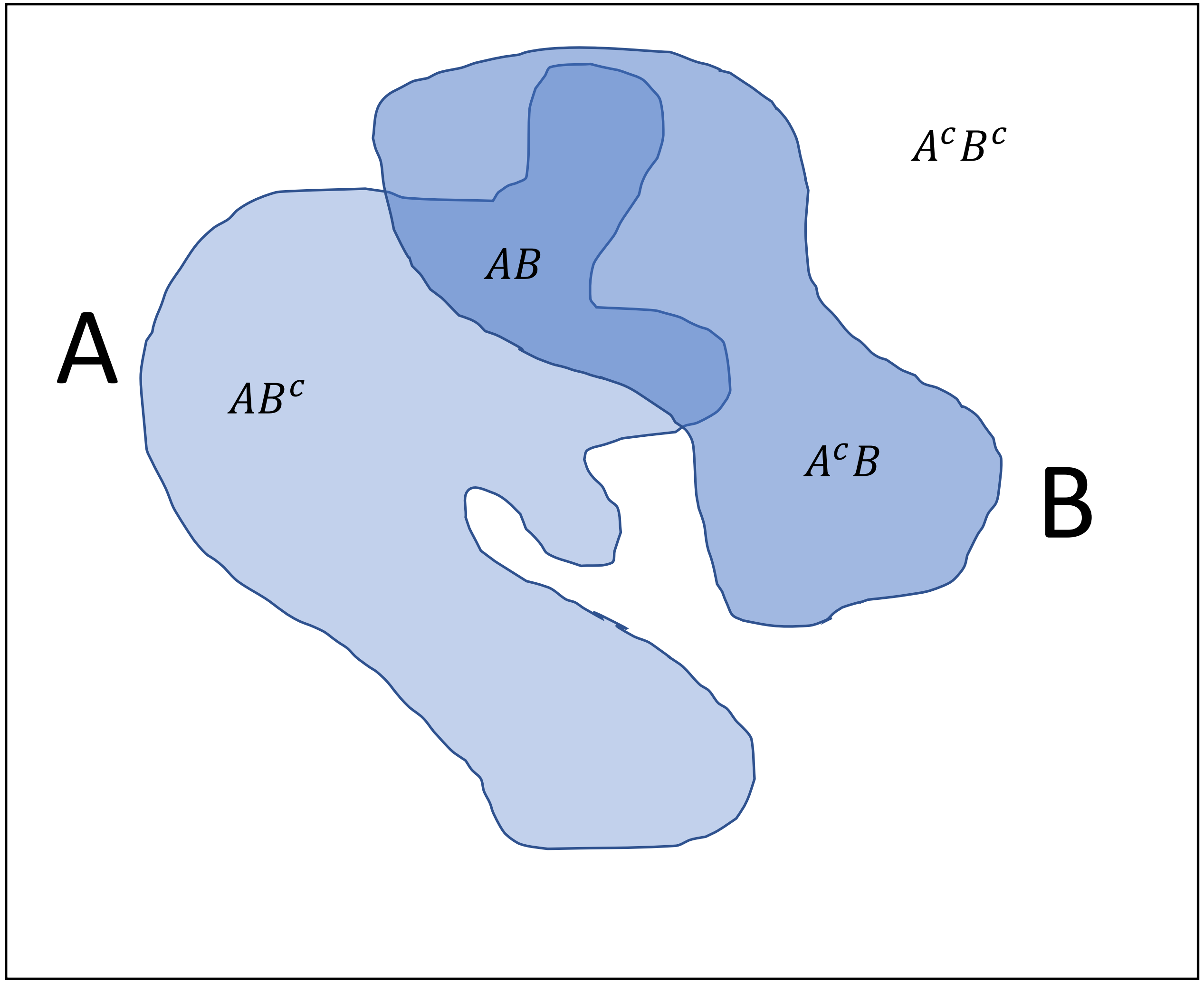
$$\mathrm{MED}(A,B) = \frac{1}{N_B}\sum_{\boldsymbol{s}\in B} d(\boldsymbol{s},A)$$

$$\mathrm{MED}(B,A) = \frac{1}{N_A}\sum_{\boldsymbol{s}\in A} d(\boldsymbol{s},B) \neq \mathrm{MED}(A,B)$$

(e.g., Baddeley 1992; Gilleland 2017)

New bias/distance performance measure, G and G_β

Gilleland (2021)



n_A = number of grid points in A,
n_B = number of grid points in B,
n_{AB} = number of grid points in AB,
$n_{A\Delta B} = n_{AB^c} + n_{A^cB} = n_A + n_B - 2n_{AB}$.

Let $y = y_1 y_2$ where

$$y_1 = n_{A\Delta B}$$

$$y_2 = \text{MED}(A,B)\cdot n_B + \text{MED}(B,A)\cdot n_A$$

The new measures are G and G_β:

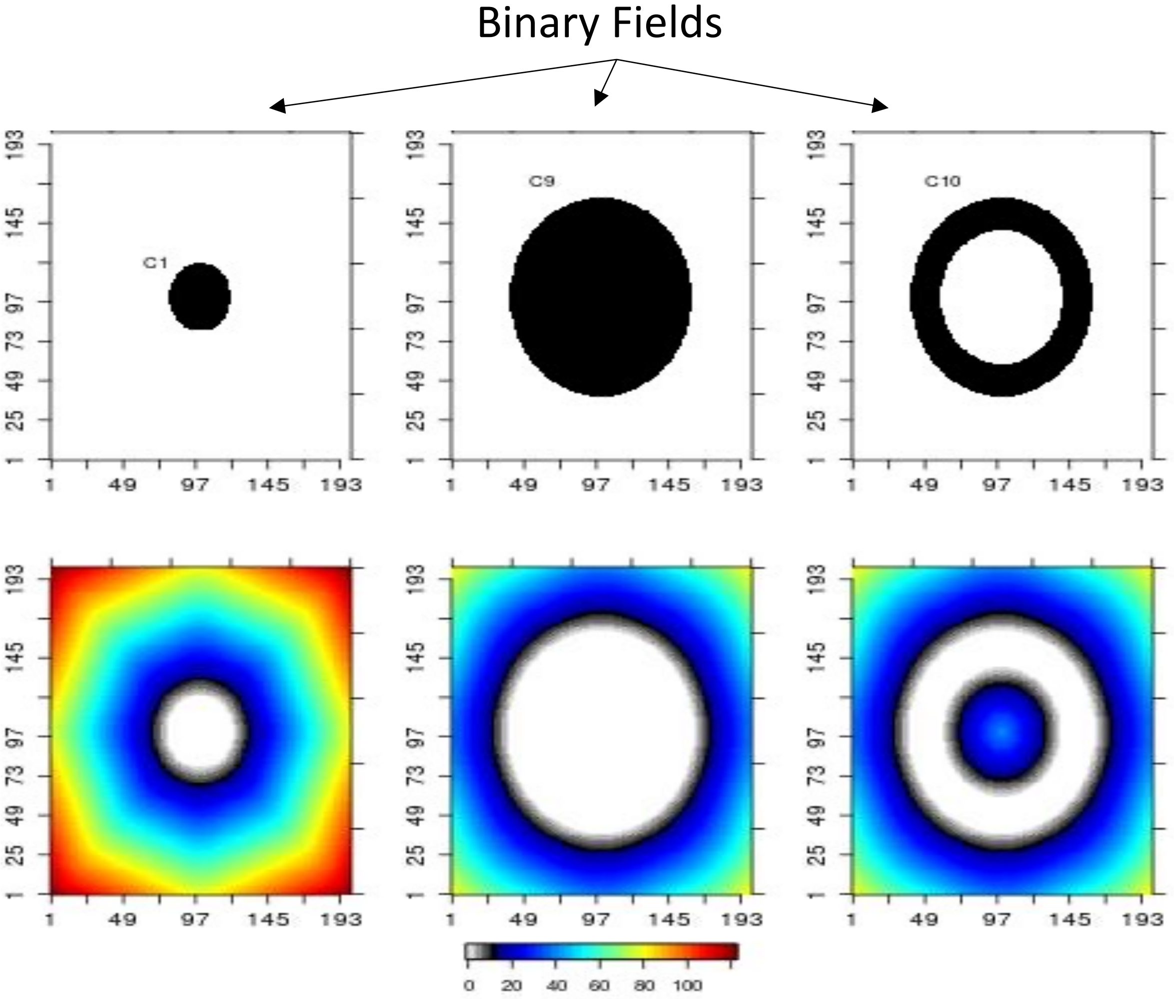
$$G = y^{1/3}$$

$$G_\beta(A,B) = \max\{1 - \frac{y}{\beta}, 0\}$$

$\sqrt{G}$ has units of g.p. and G_β is unitless.

Spatial Dissimilarity Measures

Distance Maps



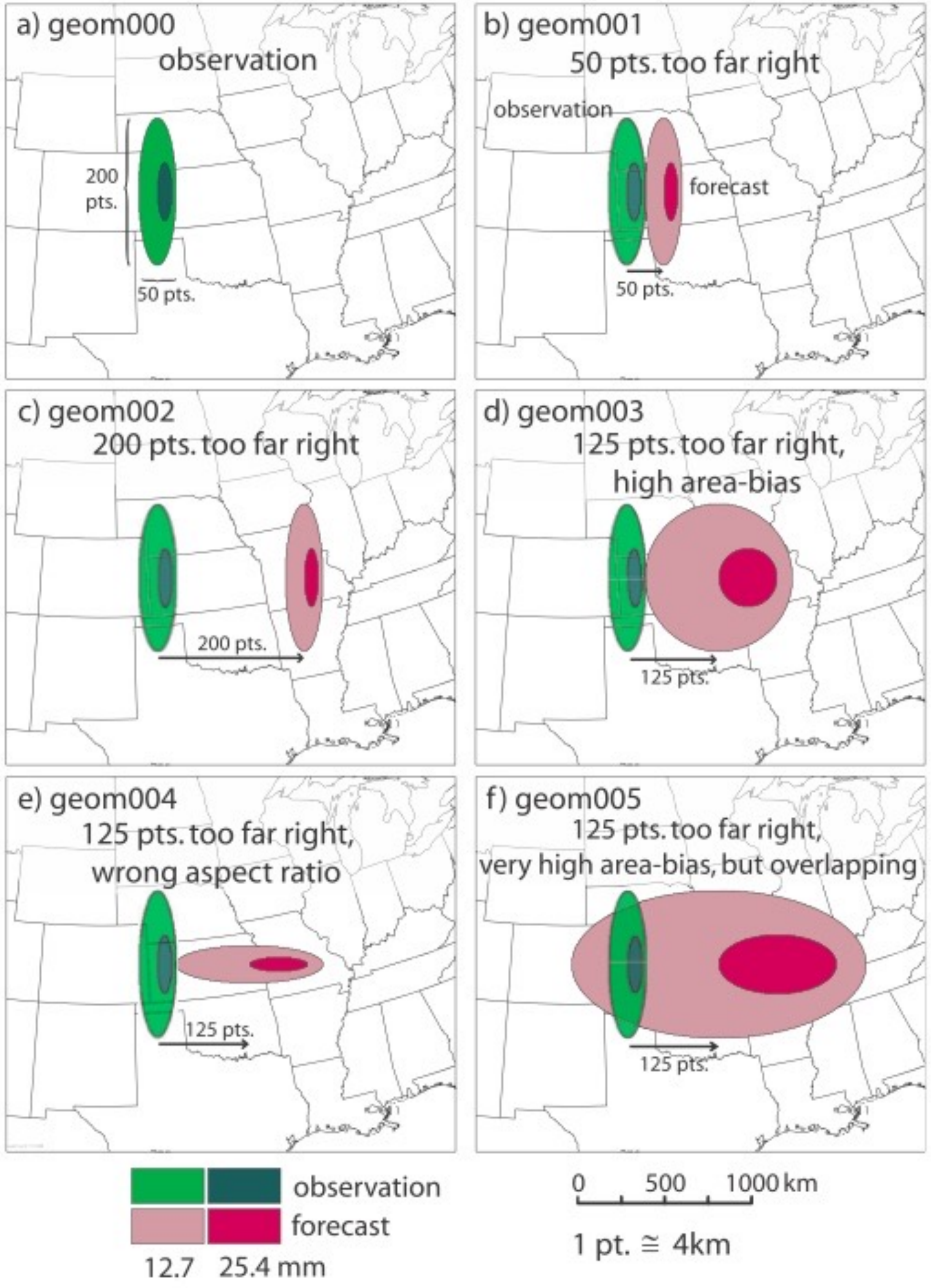
Distance maps of each binary field

Spatial Dissimilarity Measures Summary

	Handles Pathological Cases well?	No positional effects?	Sensitive to frequency bias?	Useful for rare events?	Reward partial perfect match?	Correctly penalize despite partial perfect match?
G	Yes	Yes	Yes	No	No	Yes
G_β	Yes*	Yes	Yes	Yes*	No	Yes
Centroid distance	No	Yes	No	No	No	No
Baddeley's Δ	No	No	Yes	No	Yes	No
Hausdorff	No	Yes	No	Yes	No	No
MED	No	Yes	No**	Yes**	Yes**	Yes**
FoM	No	Yes	Yes	Unclear	No	Yes

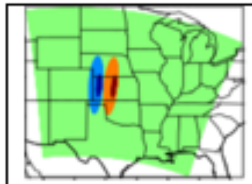
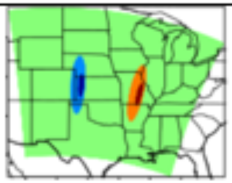
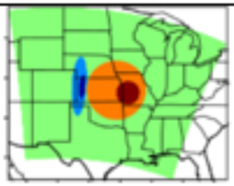
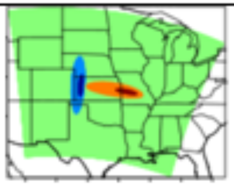
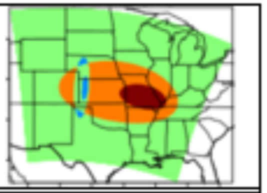
*Depending on choice of β

**Answer depends on the asymmetry of MED (i.e., may only be true in one direction but always true if looking at both directions).

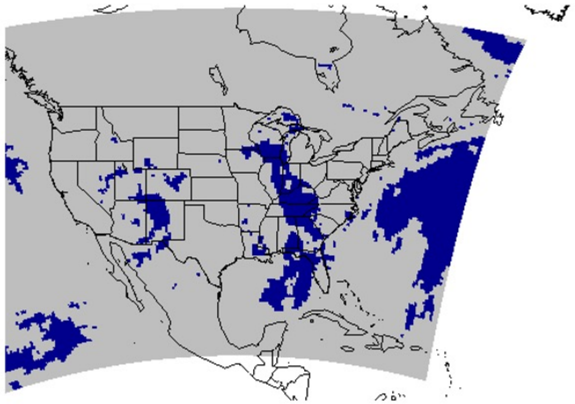


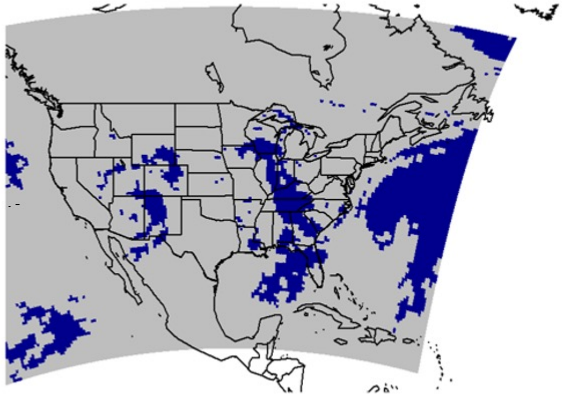
Traditional score	geom001/002/004	geom003	geom005
Accuracy	0.95	0.87	0.81
Frequency bias	1.00	4.02	8.03
Multiplicative intensity bias	1.00	4.02	8.04
RMSE (mm)	3.5	5.6	6.9
Bias-corrected RMSE (mm)	3.5	5.5	6.3
Correlation coefficient	−0.02	−0.05	0.20
Probability of detection	0.00	0.00	0.88
Probability 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

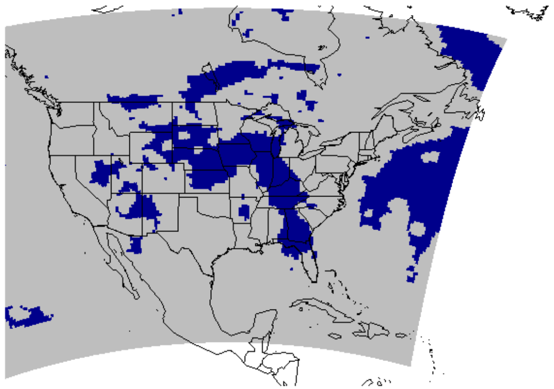
Far left figure and table from Ahijevych et al., 2009. *Weather Forecast.*, **24** (6), 1485 - 1497, doi: 10.1175/2009WAF2222298.1.

					
Method	geom001 translation-only error (50-pts)	geom002 translation-only (200-pts)	geom003 translation (125-pts) and large area bias	geom004 translation (125-pts) and aspect-ratio	geom005 translation (125-pts) and huge area bias (but overlapping)
$H(A,B)$	**Best**	**Tied for 2**	**Tied for 2**	**Tied for 2**	**Worst**
$G(A,B)$	**Best**	**3 (near tie for worst)**	**Tied for worst**	**2**	**Tied for worst**
$M(A,B)$ and $Z(A,B)$ Miss	**2 (near-tie with 3)**	**Worst**	**3 (near tie with 2)**	**4**	**Best**
$M(A,B)$ and $Z(A,B)$ False Alarm	**Best**	**Worst**	**3 (near tie with 2)**	**2 (near tie with 3)**	**4**
$F(A,B)$ Miss and False Alarm	**2**	**Worst**	**4**	**3**	**Best**
$\Delta(A,B)$	**Best**	**Worst**	**3**	**2**	**4**

GALWEM Cloud Amounts (%)







	Reanalysis (A) vs WWMCA (B)	Reanalysis (A) vs GALWEM (B)
Hausdorff distance metric	16.78	53.55
Baddeley Δ metric p = 2, c = infinity	3.45	11.03
MED(A, B)	0.147	1.87
MED(B, A)	0.207	2.95
$\sqrt{G}$	6.39	20.13
$G_\beta, \beta = \frac{N^2}{2}$	0.9998	0.76636

Summary

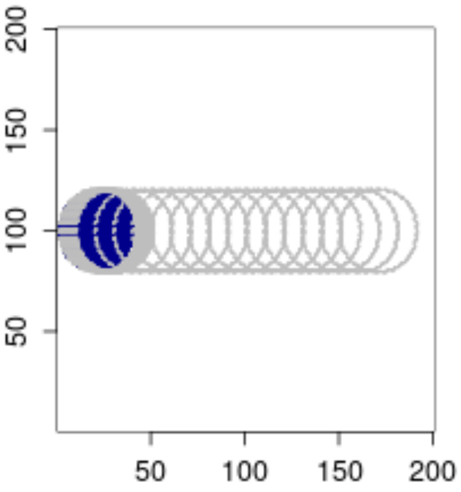
Dissimilarity Measures are:

- Relatively straightforward to understand.
- Fast to compute.
- Ideal for verifying cloud forecasts.
- Often used within more complicated spatial verification methods (e.g., MODE).
- May give complementary information so that more than one should be used.
- Should be applied in conjunction with other complimentary verification measures such as frequency bias.

Software:

- Model Evaluation Tools (Brown et al. 2021, doi: 10.1175/BAMS-D-19-0093.1)
- SpatialVx (doi: 10.5065/4px3-5a05)
- Others??
  - https://projects.ral.ucar.edu/icp/

New bias/distance performance measure, G_β



- $0 \leq G(A,B) \leq 1$
- $G(A,B) = 1$ is a perfect match between A and B.
- $G(A,B) = 0$ is a really bad match.