Eric Gilleland

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

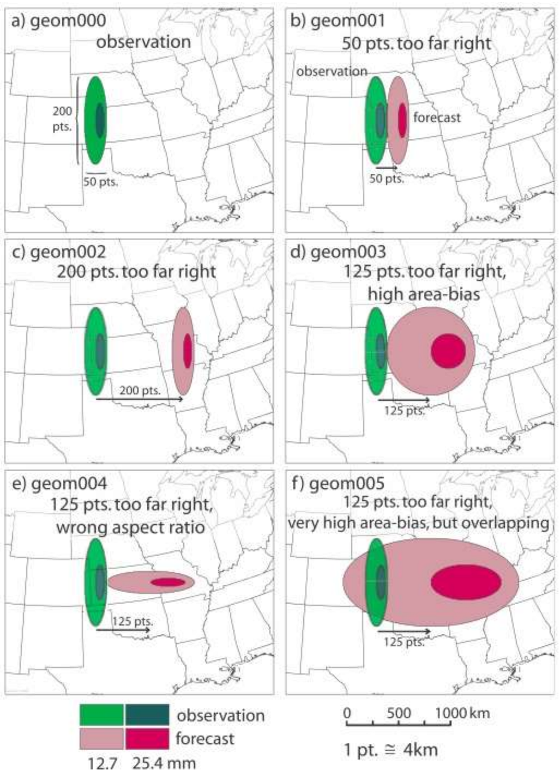
Research Applications Laboratory, National Center for Atmospheric Research,

Boulder, Colorado, U.S.A.

Spatial Forecast Verification
MPE Summer School

24 June 2021

Spatial Forecast Verification



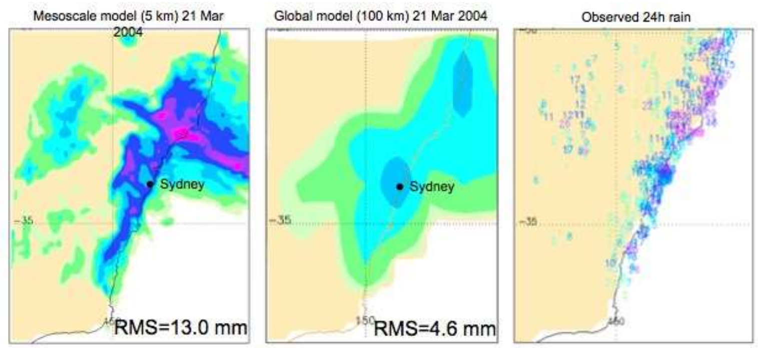
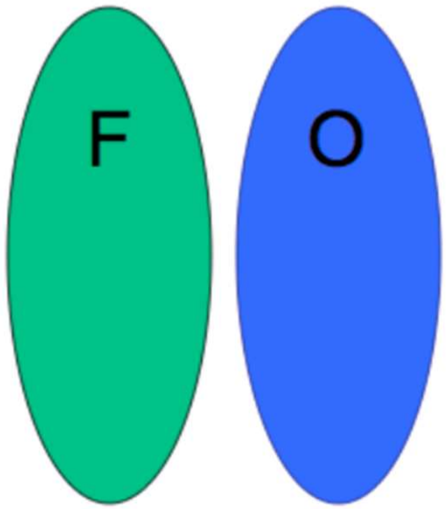


Figure from Beth Ebert

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.

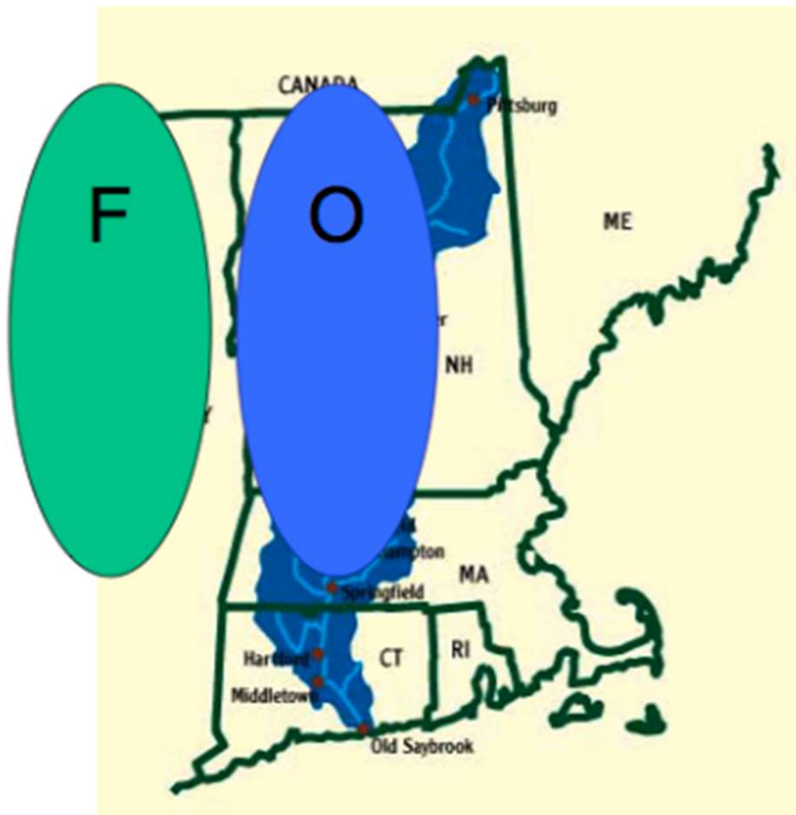
Spatial Forecast Verification



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

Figure from Barb Brown

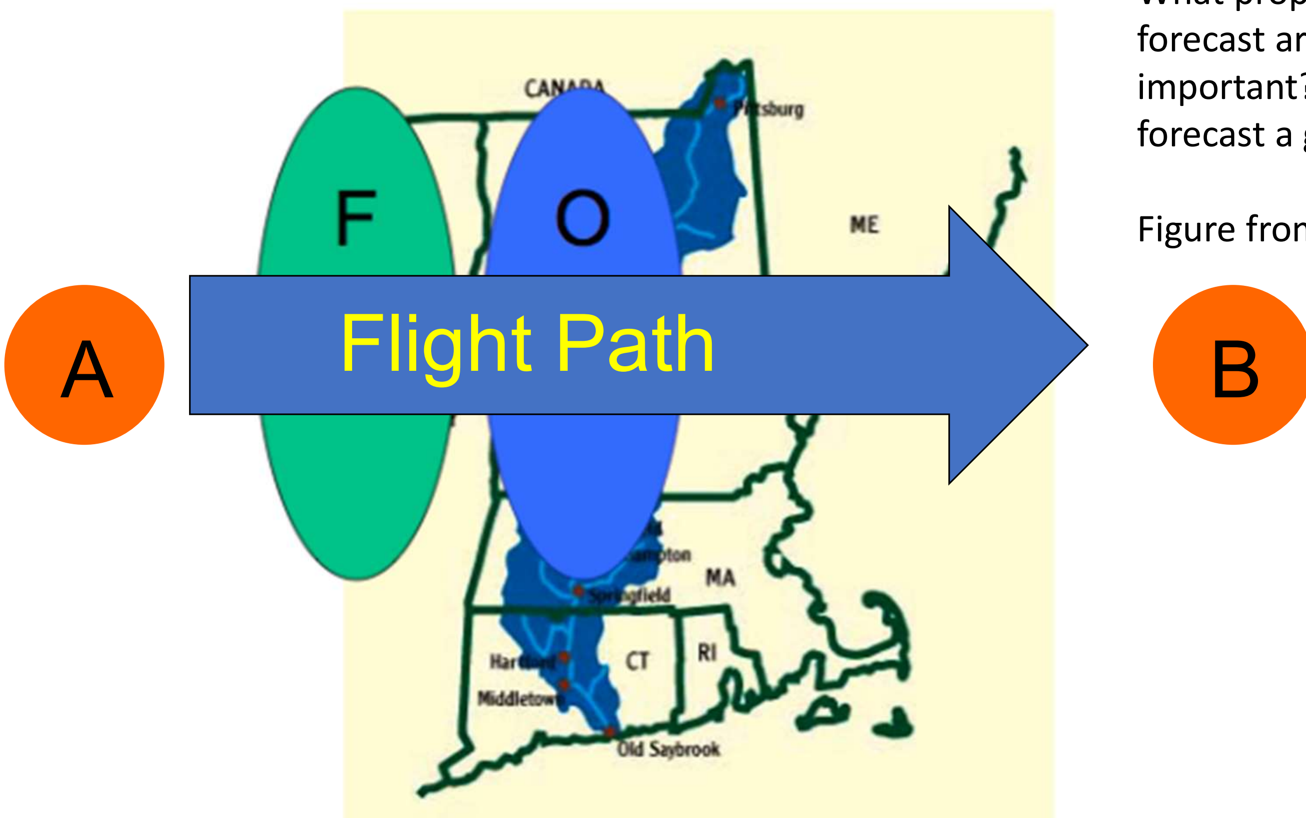
Spatial Forecast Verification



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

Displacement Methods: Field Deformation

Graphic by Johan Lindström

0-energy control points (landmarks, tie-points)

Image Warping

1-energy control points

Forecast ($\hat{Z}(\boldsymbol{s})$) Warped ($\hat{Z}(W(\boldsymbol{s}))$) Observed ($Z(\boldsymbol{s})$)

See, for example,

- *G. et al. (2010) Weather Forecast.*, **25** (4), 1249 - 1262, doi: 10.1175/2010WAF2222365.1
- *G. et al. (2010) NCAR Technical Note*, NCAR/TN-482+STR, 23pp.
- *G. (2013) Mon. Wea. Rev.*, **141**, (1), 340 - 355, doi: 10.1175/MWR-D-12-00155.1

See this paper for a list of references on the more general field deformation topic.

Displacement Methods: Field Deformation

Image Warping

$$X(\boldsymbol{s}) = \hat{X}(\boldsymbol{W}(\boldsymbol{s})) + \varepsilon = \hat{X}\left(W_x(x,y), W_y(x,y)\right) + \varepsilon$$

Warp function

$U(r) = r^2 \log r$

$$W_x(x,y) = a_{x,0} + a_{x,1} \cdot x + a_{x,2} \cdot y + \sum_{i=1}^{n_c} d_{x,i} \cdot U\left(\left\|\boldsymbol{p}_{X,i} - (x,y)\right\|\right)$$

Horizontal Translations

$\boldsymbol{p}_{X,i}$ is the $\boldsymbol{i}$-th 0-energy control point

Displacement Methods: Field Deformation

Image Warping

Pros

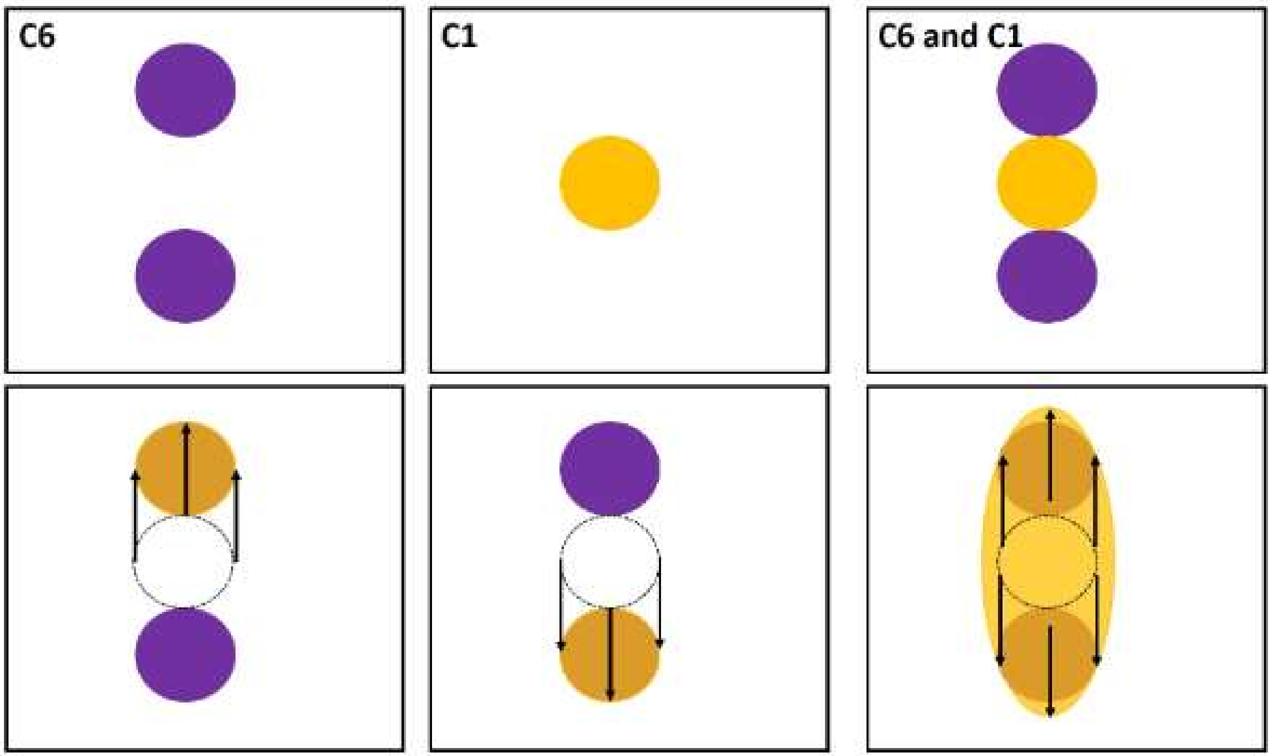
- The ability to write the deformation as a statistical model means that uncertainty information is elegantly obtainable, as well as spatial correlations!
- Finds an optimal spatial alignment whereby the amount of linear and nonlinear deformations can be taken into account, along with a percent reduction in error.
- Ultimately, familiar traditional measures can be applied to the deformed fields.

Cons

- The warp function, $\boldsymbol{W}$, is not unique.
- Can be difficult to implement.
- Can be difficult to determine what constitutes a good v. bad forecast.

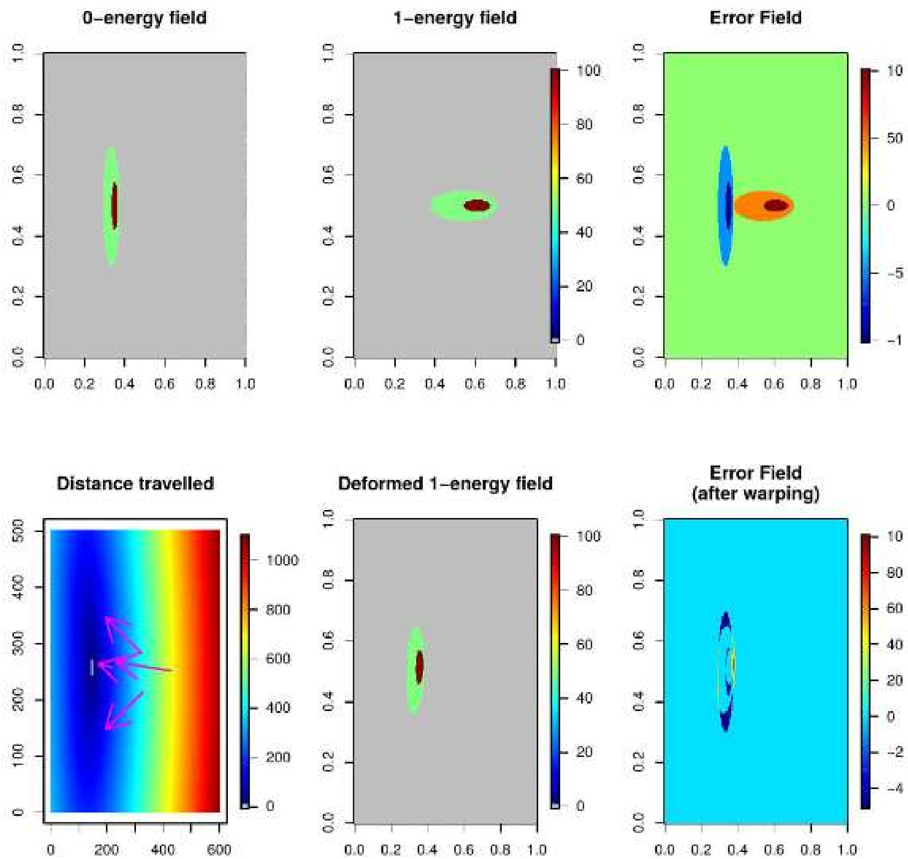
Displacement Methods: Field Deformation

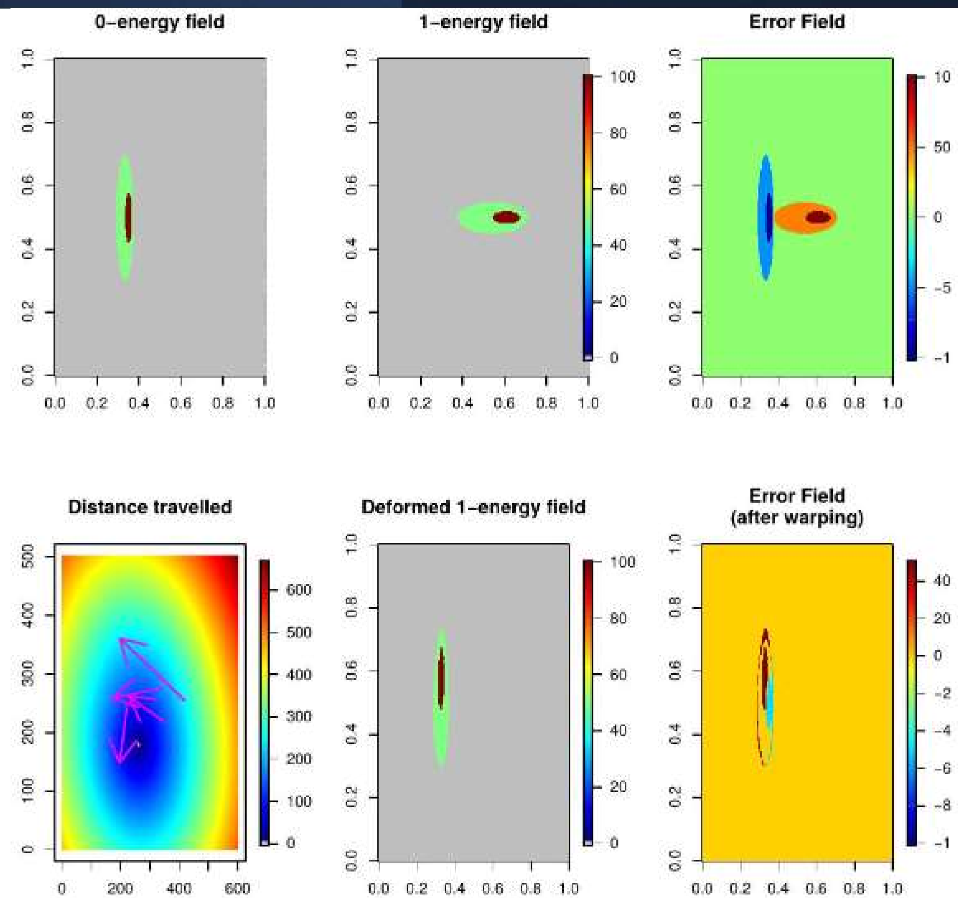
Image Warping



Displacement Methods: Field Deformation

Image Warping





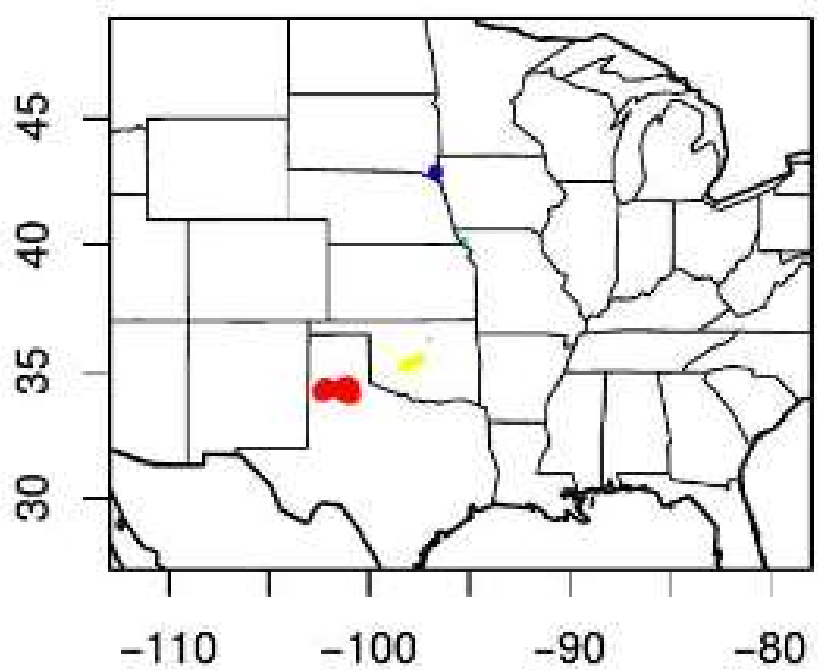
Displacement Methods: Field Deformation

Other deformation methods of note:

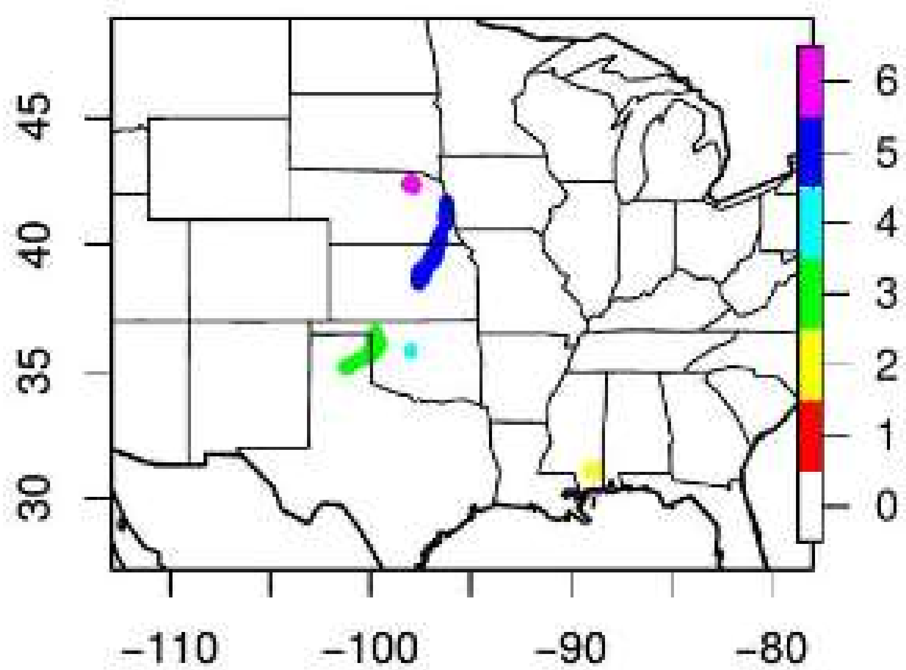
- Pyramid Scheme (Keil and Craig 2009, doi: 10.1175/2009WAF2222247.1.)
- Optical Flow (Marzban and Sandgathe 2010, doi: 10.1175/2010WAF2222351.1.)

Displacement Methods:Feature-based

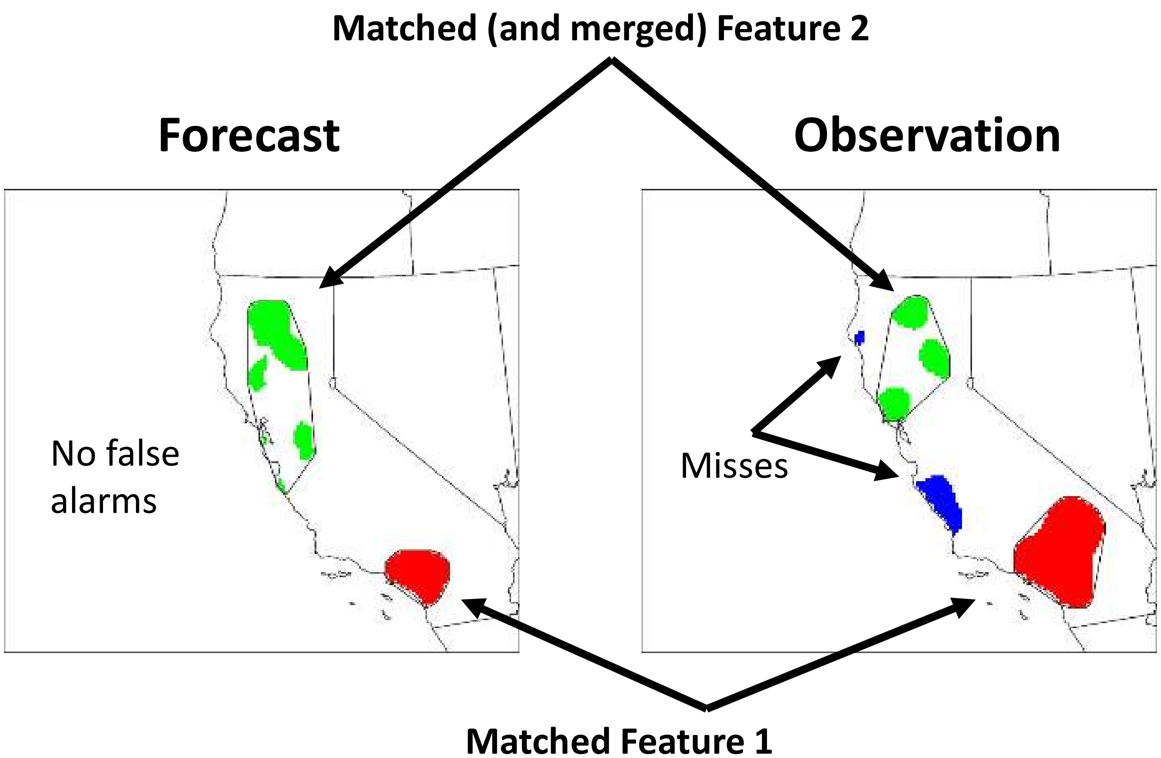
Observation



Forecast

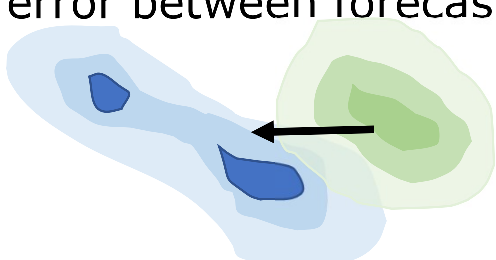


Displacement Methods: Feature-based



Displacement Methods:Feature-based

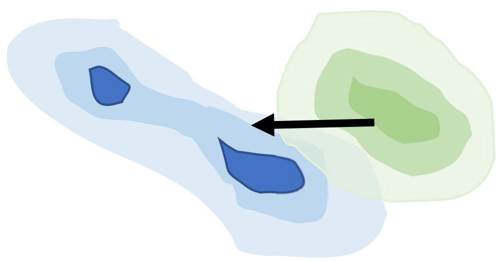
- Define entities using threshold (Contiguous Rain Areas; Ebert and McBride 2000, doi: 10.1016/S0022-1694(00)00343-7)
- Horizontally translate the forecast until a *pattern matching* criterion is met:
  - minimum total squared error between forecast and observations
  - maximum correlation
  - maximum overlap



- The displacement is the vector difference between the original and final locations of the forecast.

Displacement Methods:Feature-based

CRA decomposition of MSE gives the total MSE, call it m_t, as a sum of the MSE associated with *displacement errors*, m_d, MSE associated with volume errors, m_v, which is simply the squared difference between the mean of the forecast and observed values after translation, and MSE associated with pattern errors, m_p. Define m_s to be the MSE associated with the shift error (i.e., MSE after translation), then $m_d = m_t - m_s$.

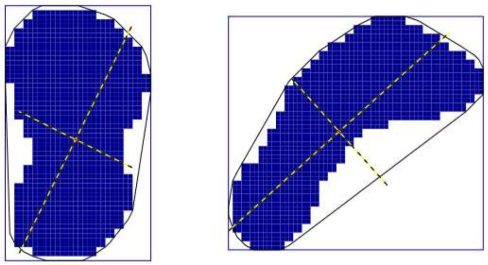


Displacement Methods:Feature-based

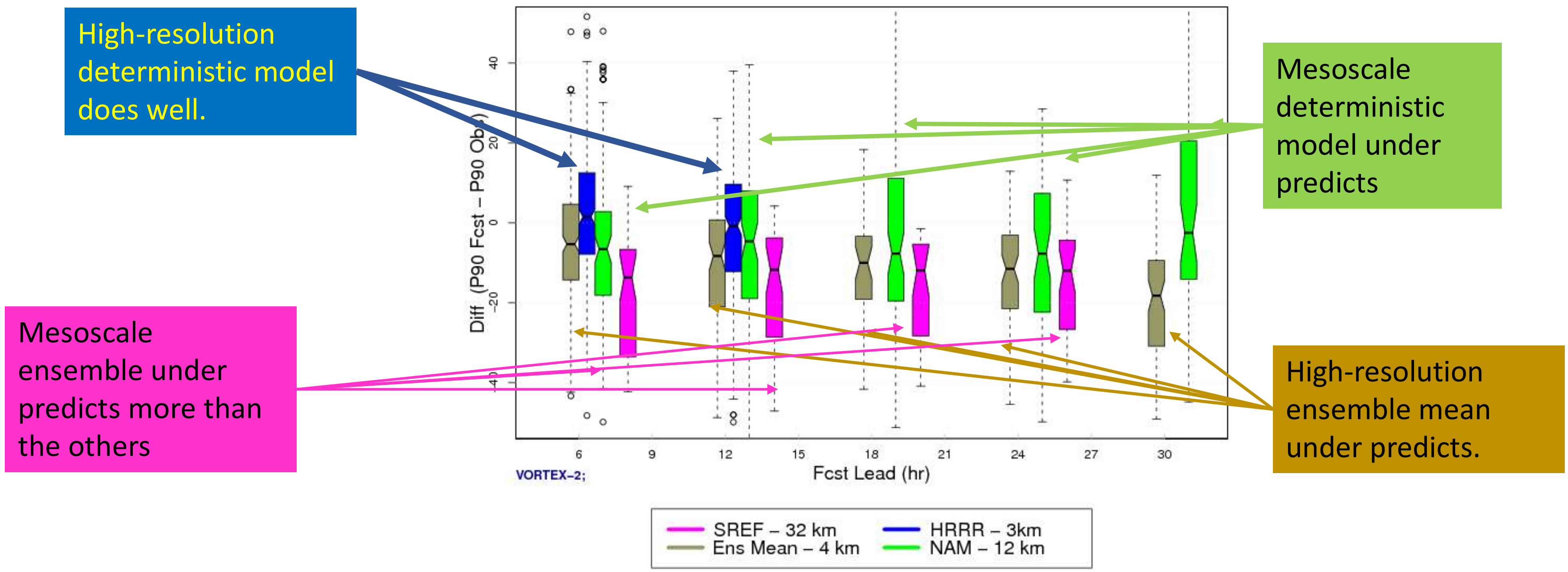
Method for Object-based Diagnostice Evaluation (MODE)

There are many ways to compare features, for example:

- Area ratio (F/O) = 1.3 means that the forecast is 30% too large.
- Centroid distance = 1 km means that the forecast's center of mass is translated by 1 km from that of the observation.
- Orientation angle difference of 15% means that the forecast is oriented 15^o from the orientation of the observed feature.
- Difference in 90-th percentiles of intensity equal to 0.5 means that peak rain is ½" too much.



Displacement Methods:Feature-based

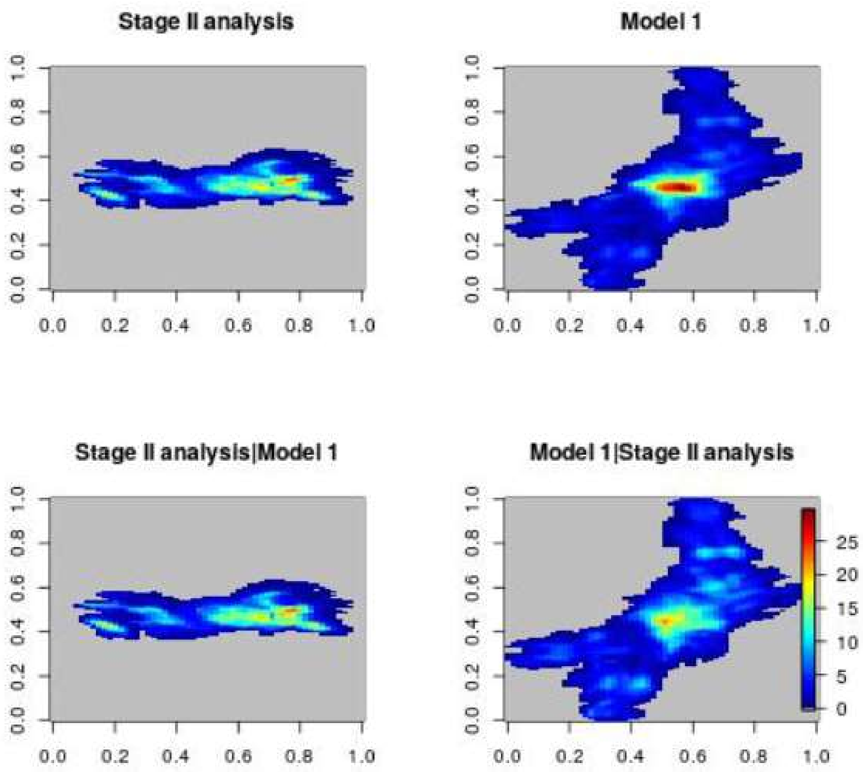


Method for Object-based Diagnostice Evaluation (MODE)

Displacement Methods:Feature-based

Distributional Summaries of Features

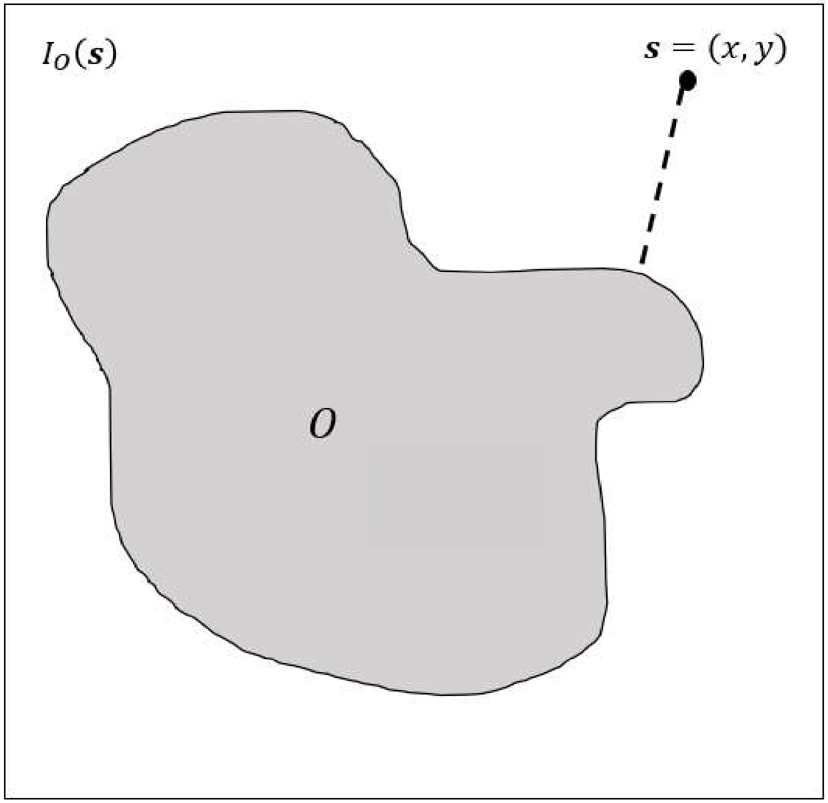
- Numbers/sizes of features
- Types of features
- Composites (ala Nachamkin 2009, https://doi.org/10.1175/2009WAF2222225.1). Figure on right is something similar.
- Boxplots of individual feature properties.



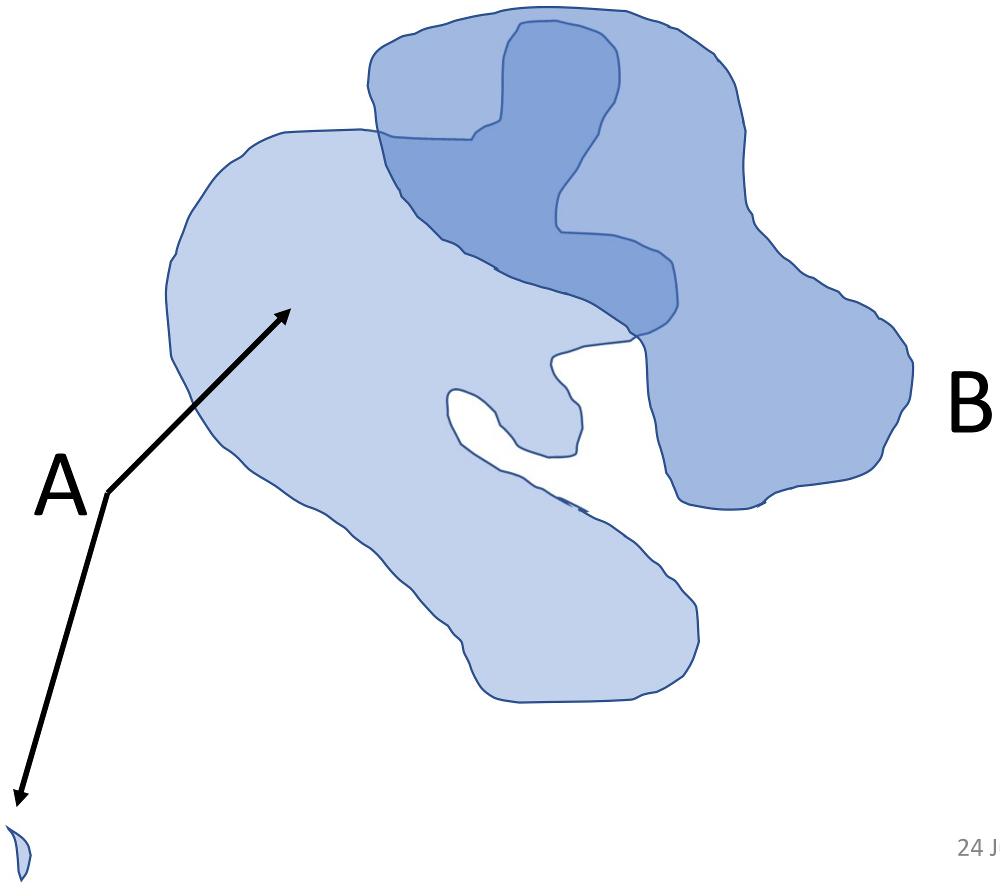
Spatial Dissimilarity Measures

$I_O(\boldsymbol{s}) = 1$ if $Z(\boldsymbol{s}) > u$ for example
$I_O(\boldsymbol{s}) = 0$ otherwise

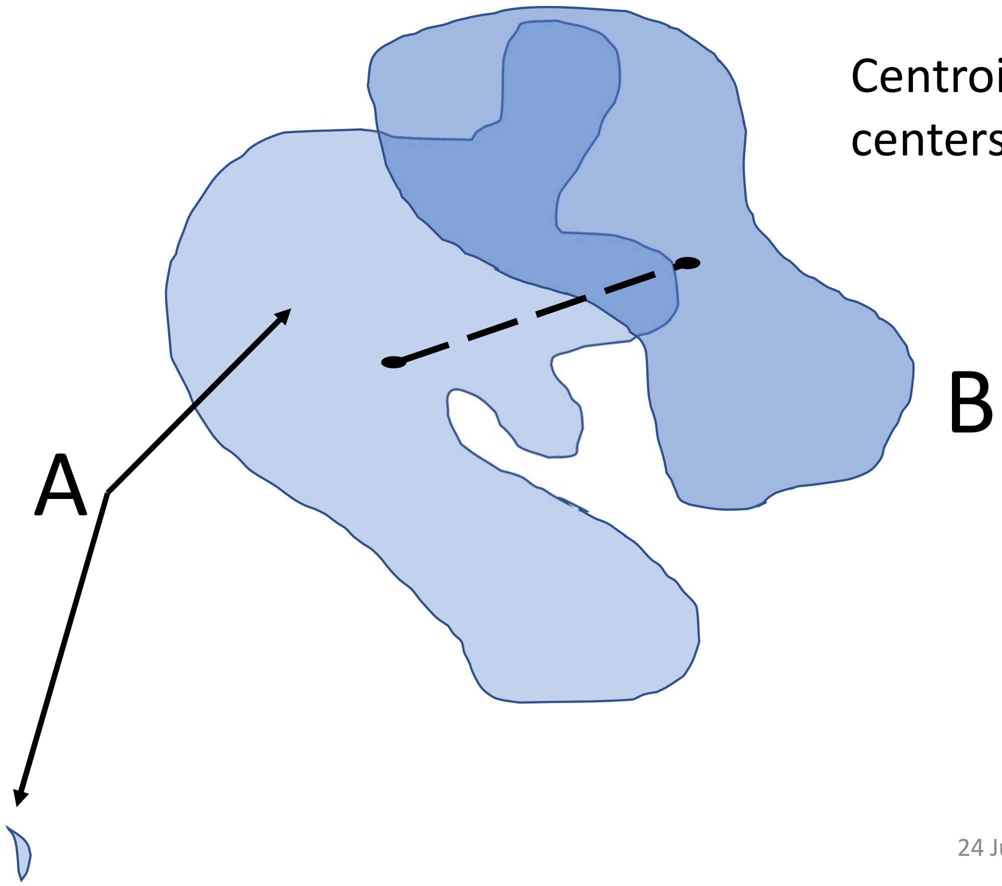
$\mathcal{D} \longrightarrow$



Spatial Dissimilarity Measures



Spatial Dissimilarity Measures

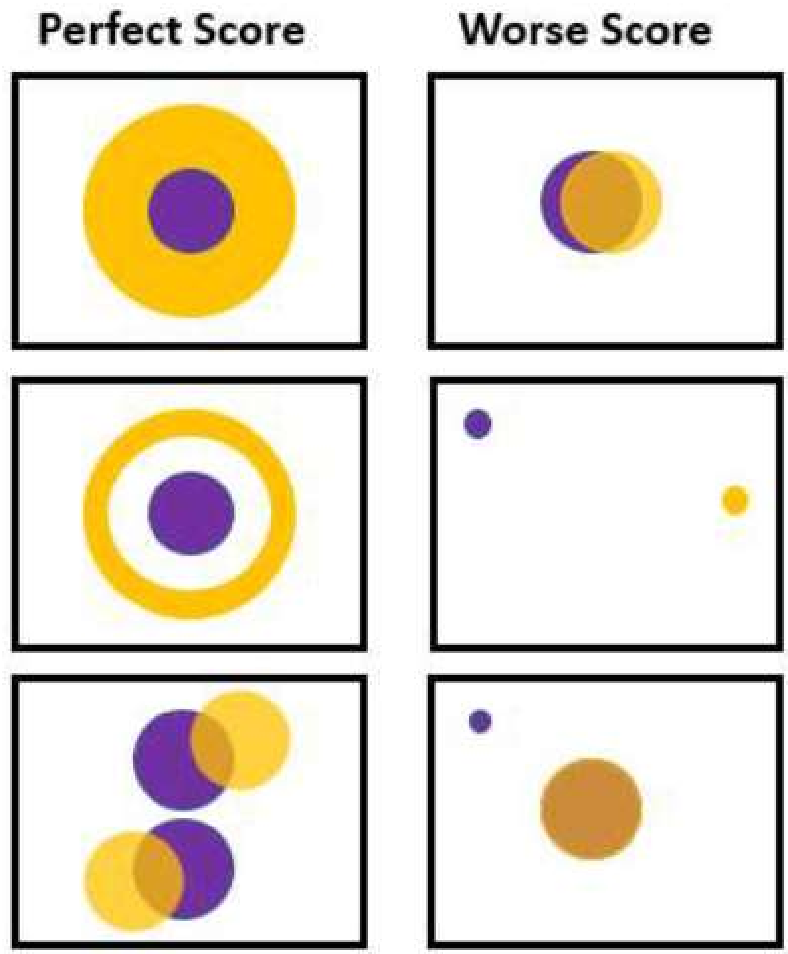


Centroid Distance is the distance between the centers of mass of the two sets.

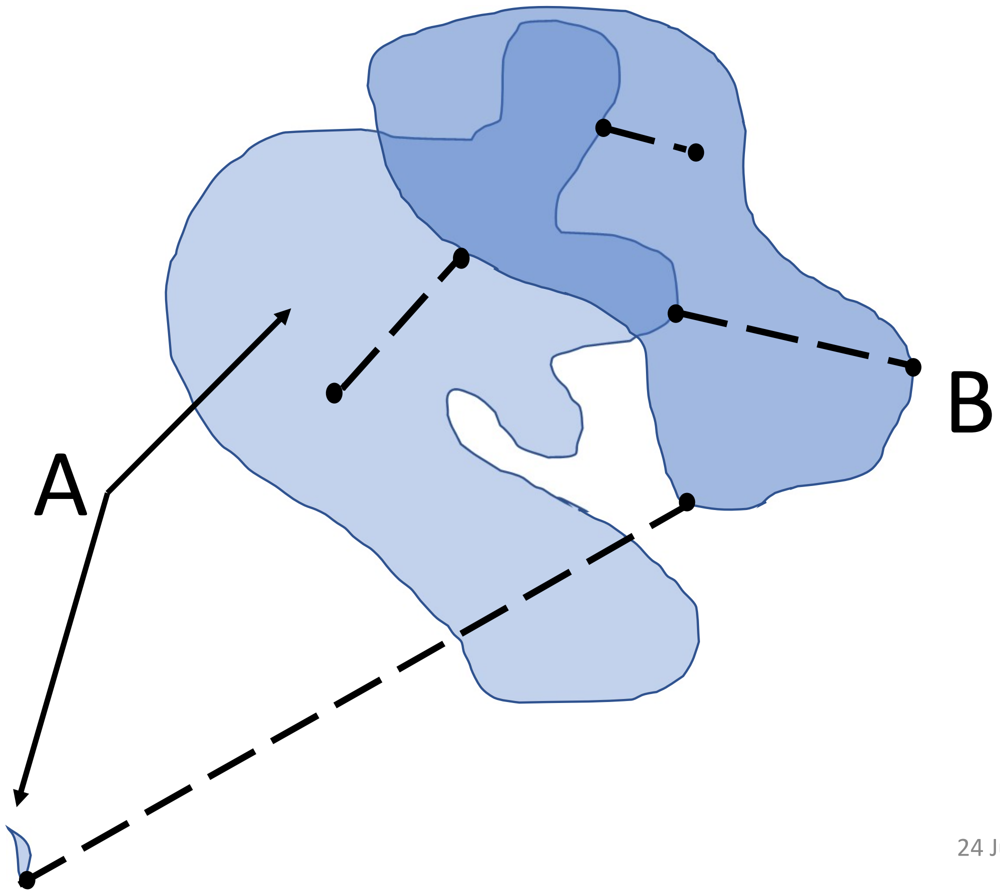
$$\mathrm{C}(\mathcal{D}) = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{s}_i \cdot I(\boldsymbol{s}_i) = \frac{1}{|\mathcal{D}|}\sum_{\boldsymbol{s}\in\mathcal{D}} \boldsymbol{s} \cdot I(\boldsymbol{s})$$

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

Spatial Dissimilarity Measures

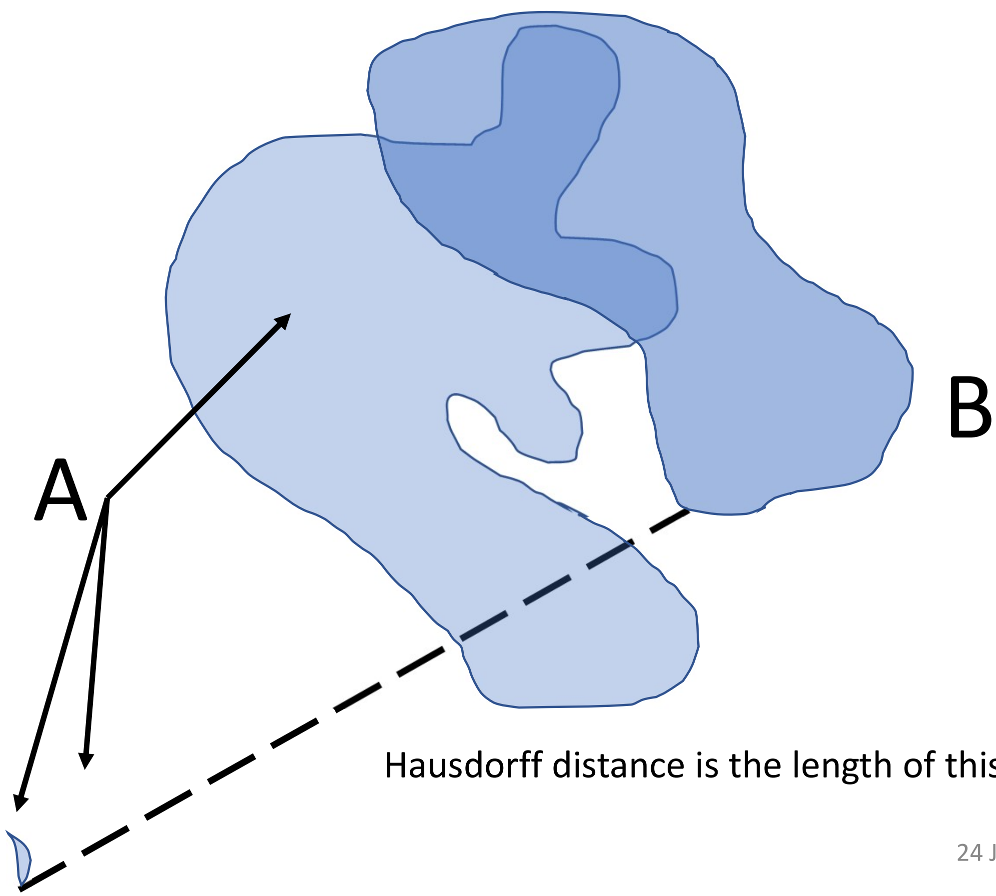


Spatial Dissimilarity Measures



$d(\boldsymbol{s}, A)$ is the shortest distance from a grid point $\boldsymbol{s} \in \mathcal{D}$ to the nearest grid point in the set A. Similarly for $d(\boldsymbol{s}, B)$.

Spatial Dissimilarity Measures

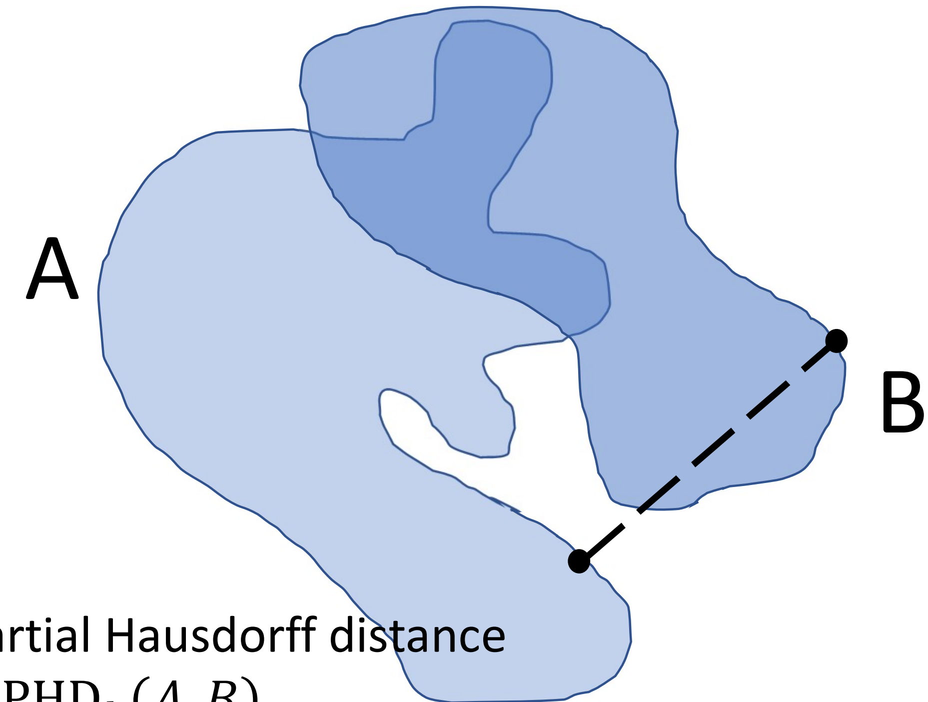


Hausdorff distance

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

Hausdorff distance is the length of this dashed line

Spatial Dissimilarity Measures



Hausdorff distance

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

= the length of the dashed line

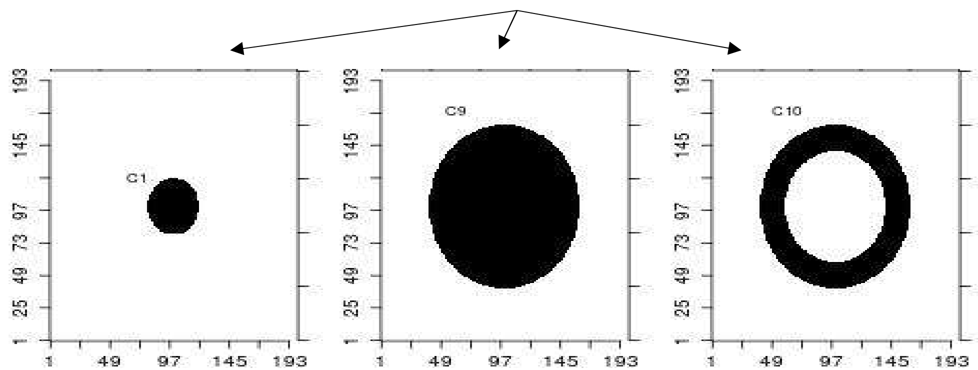
Partial Hausdorff distance

$$\mathrm{PHD}_k(A,B)$$
$$= \max\left\{k\text{−th}\underset{\boldsymbol{s}\in B}{\text{largest}}\, d(\boldsymbol{s},A), k\text{−th}\underset{\boldsymbol{s}\in A}{\text{largest}}\, d(\boldsymbol{s},B)\right\}$$

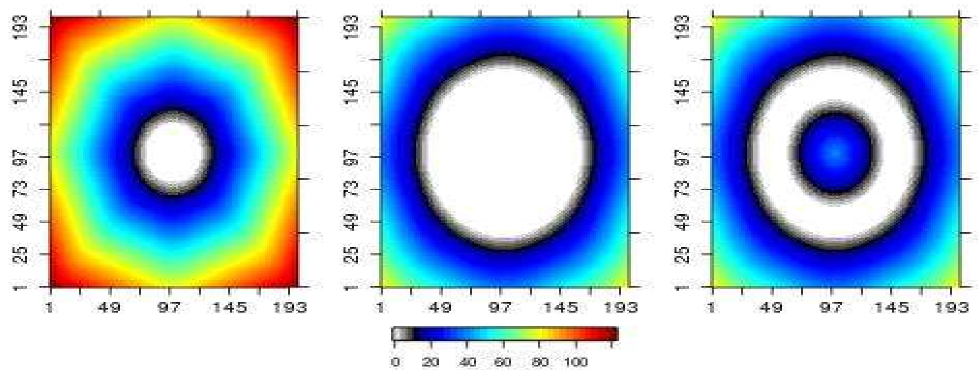
Spatial Dissimilarity Measures

Distance Maps

Binary Fields



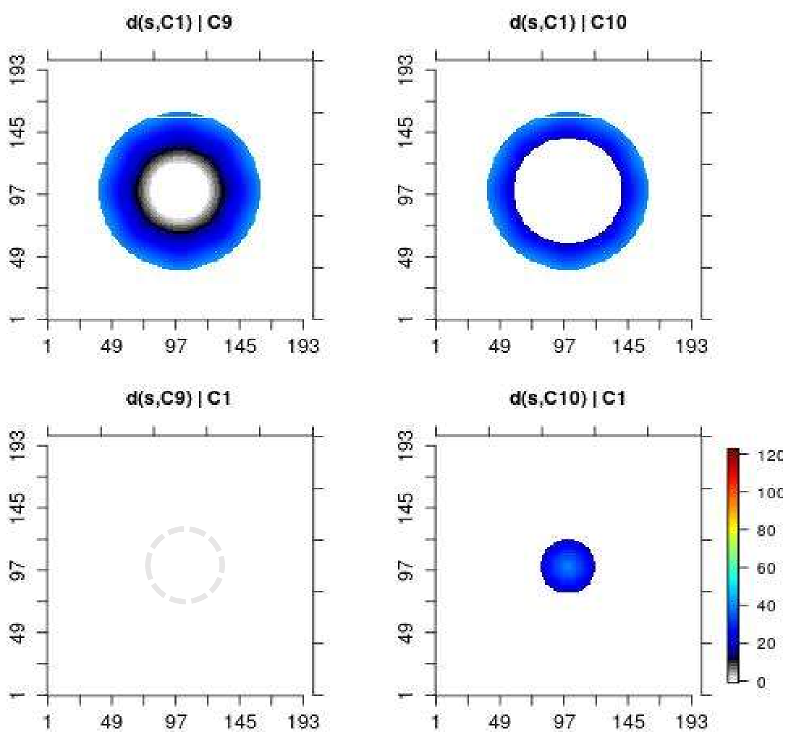
Distance maps of each binary field



Spatial Dissimilarity Measures

MED(C1,C9) is the average of the circle (including the inner white circle, but not the white part outside of the circle.

MED(C9,C1) is the average of the inner white circle (all zero-valued).



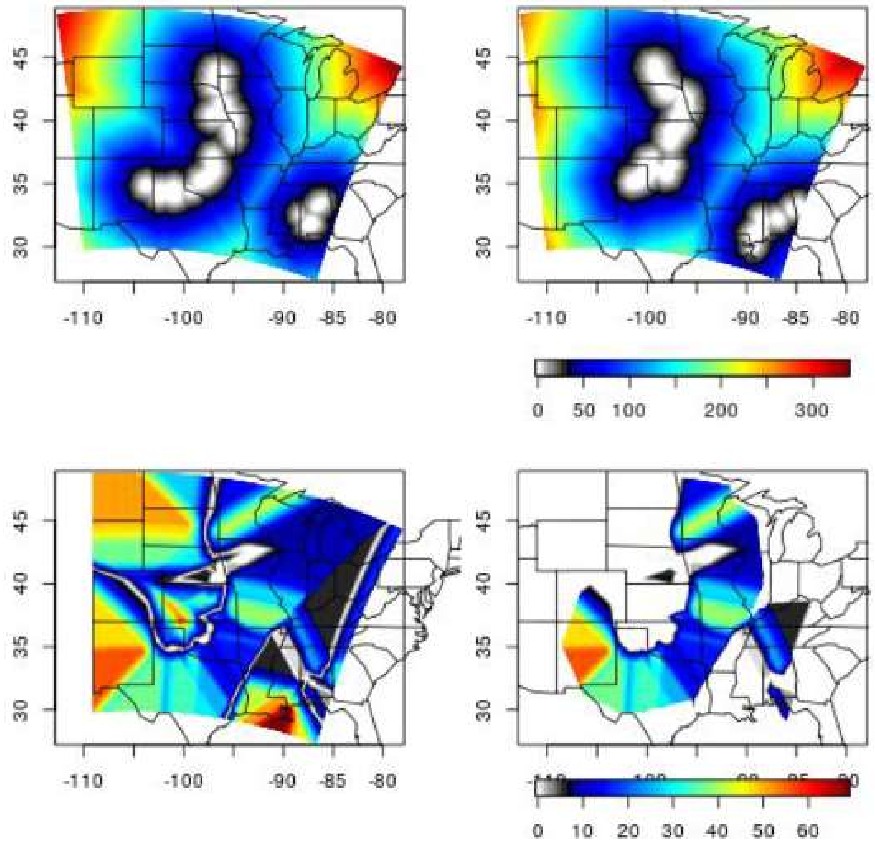
MED(C1,C10) is the average of the colored ring.

MED(C10,C1) is the average of the colored circle.

Spatial Dissimilarity Measures

Distance maps of the binary fields

Magnitude difference between the two distance maps in the top row.



Magnitude difference between the two distance maps in the top row after fist setting all distances larger than 150 grid squares to zero.

Spatial Dissimilarity Measures

Baddeley's Δ Metric

cut-off function $\omega(x) = \min\{x, \text{constant}\}$

user chosen parameter

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

Sum is over entire domain!

distance maps for A (giving $d(\boldsymbol{s},A)$ for every grid point $\boldsymbol{s} \in \mathcal{D}$) and B.

Spatial Dissimilarity Measures

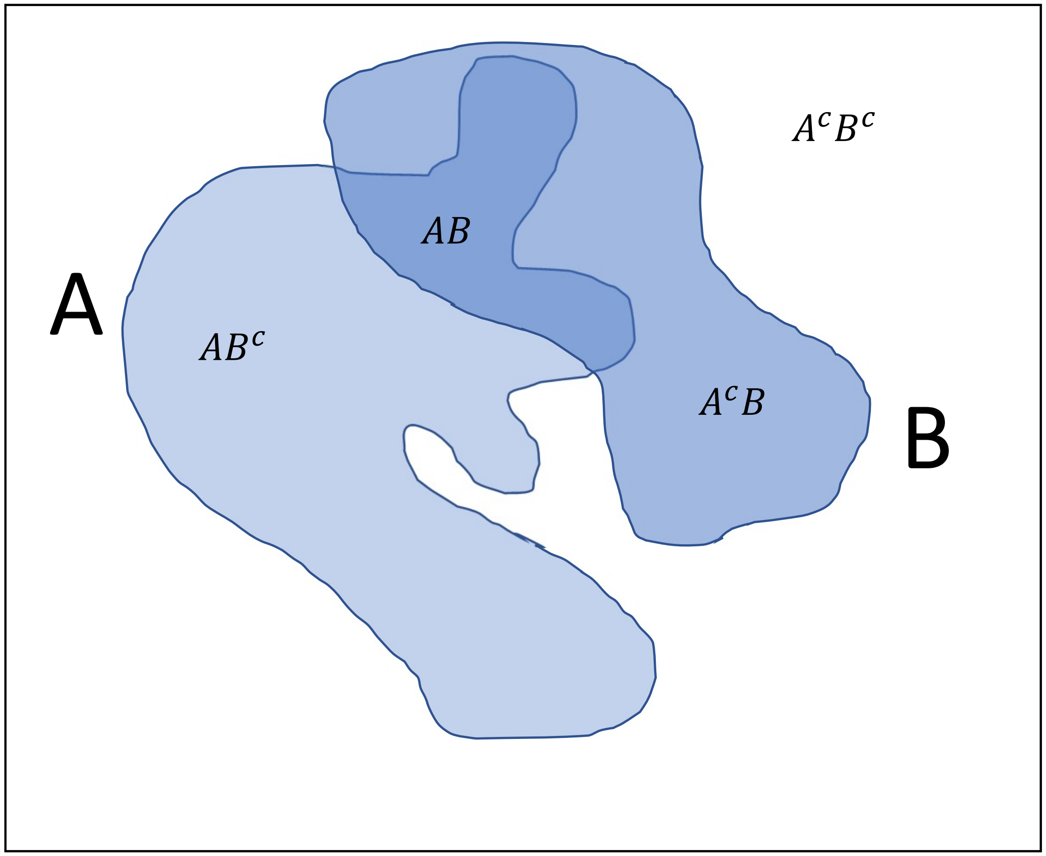
Baddeley's Δ Metric

user chosen parameter

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

- $p = 1$ gives a straight average of $\mathrm{Q} = \left|\omega\big(d(\boldsymbol{s}, A)\big) - \omega(d(\boldsymbol{s}, B)\right|$,
- $p = 2$ gives a Euclidean average of Q (emphasizes large differences more),
- $p = \infty$ gives the Hausdorff distance.

New bias/distance performance measure, G_β



n_A = number of grid points in A,
n_B = number of grid points in B,
n_{AB} = number of grid points in AB.

$$G = y^{1/3}$$
$$G_\beta(A,B) = \max\{1 - \frac{y}{\beta}, 0\}$$

where

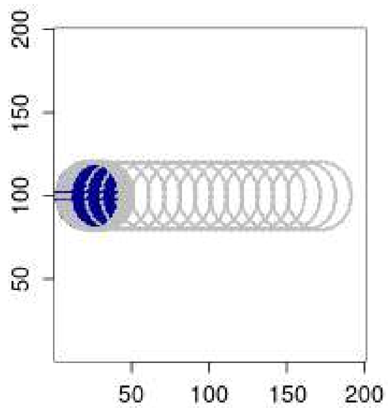
$$y = y_1 y_2$$

$$y_1 = n_A + n_B - 2n_{AB}$$

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

G. (2021, doi: 10.5194/ascmo-7-13-2021)

New bias/distance performance measure, G



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

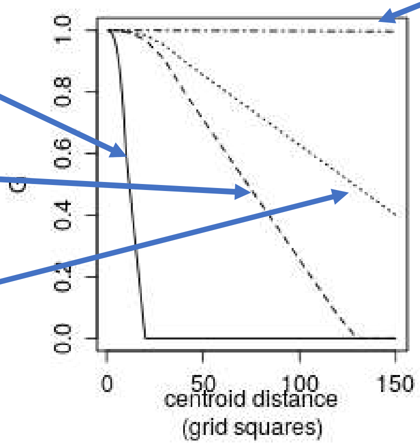
$\beta = N^3$

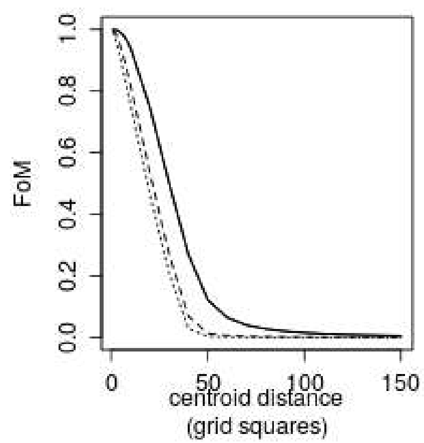
Maximum value of y_1 occurs when $n_A + n_B = N$ and $n_{AB} = 0$

$\beta = N\sqrt{N}$

$\beta = \frac{N^2}{2}$

$\beta = N^2$





Maximum value of y_2 depends on specific distance map, but should be approximately $N\sqrt{m^2 + n^2}$ for $|\mathcal{D}| = m \times n$, or $N^2\sqrt{2}$ if $n = m$. Occurs if $n_A = N$ and $n_B = 0$.

Equations for each of the distance-based measures compared here. Let $\boldsymbol{s} = (x, y) \in \mathcal{D}$ represent a grid point (coordinate) in the domain $\mathcal{D}$, N be the size of the domain with $A, B \subset D$ representing sets of grid points whose corresponding value is one (in the binary field). Let n_A and n_B represent the number of grid points in the sets A and B, respectively, and let n_{AB} represent the number of grid points in both sets. Further, let $I_A(\boldsymbol{s}) = 1$ if $\boldsymbol{s} \in A$ and zero otherwise, similarly for $I_B(\boldsymbol{s})$.

Measure Name	Measure Equation
Hausdorff distance	$H(A,B) = \max\left\{\max_{\boldsymbol{s}\in B}[d(\boldsymbol{s},A)], \max_{\boldsymbol{s}\in A}[d(\boldsymbol{s},B)]\right\}$
Baddeley's Δ	$\Delta(A,B) = \left[\frac{1}{N}\sum_{\boldsymbol{s}\in\mathcal{D}}\{d(\boldsymbol{s},A) + d(\boldsymbol{s},B)\}^2\right]^{1/2}$
Mean-error distance	$M(A,B) = \frac{1}{n_B}\sum_{\boldsymbol{s}\in B} d(\boldsymbol{s},A)$
$\boldsymbol{G_\beta}$	$G(A,B) = \max\left\{1 - \frac{1}{\beta}(n_A + n_B - 2n_{AB})(M(A,B)\mathrm{n}_B + M(B,A)\mathrm{n}_A), 0\right\}$

Spatial Dissimilarity Measures

Does the measure handle oft-encountered but pathological situations, such as, when one or both fields are empty, one field goes from being empty to having only one or a few points, etc.?

Baddeley's Δ, Hausdorff distance, centroid distance, MED, FoM, dFSS ✗ **No**

$\sqrt{G}, G_\beta$ ✓ **Yes**

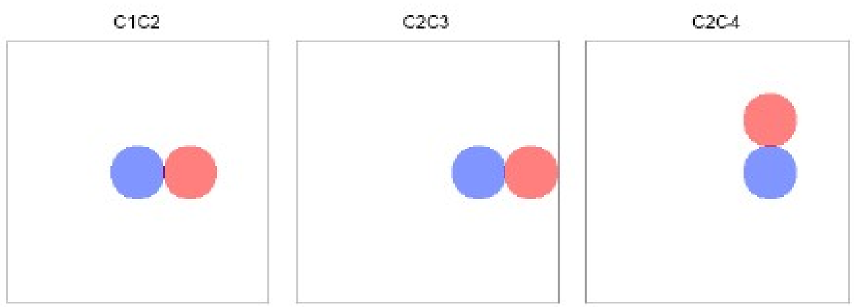
See G. (2017; https://doi.org/ 10.1175/WAF-D-16-0134.1), G. et al. (2020; https://doi.org/10.1175/MWR-D-19-0256.1) and G. (2021; https://doi.org/10.5194/ascmo-7-13-2021) for more details.

Spatial Dissimilarity Measures

Is the measure consistent regardless of the relative positions of features within the fields?

Baddeley's Δ, dFSS — **No**

Hausdorff distance, centroid distance, MED, FoM $\sqrt{G}$, G_β — **Yes**



See G. (2017; https://doi.org/ 10.1175/WAF-D-16-0134.1), G. et al. (2020; https://doi.org/10.1175/MWR-D-19-0256.1) and G. (2021; https://doi.org/10.5194/ascmo-7-13-2021) for more details.

Spatial Dissimilarity Measures

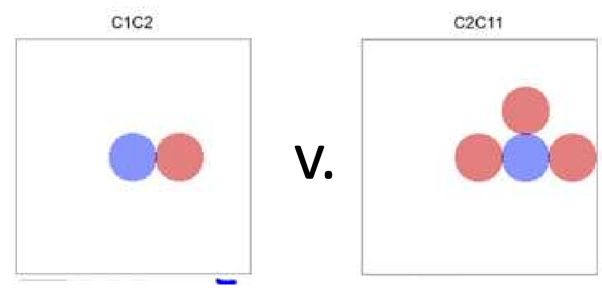
Does the measure degrade in the presence of frequency bias?

Hausdorff distance, MED, centroid distance, dFSS (undefined)

No

Baddeley's Δ, FoM $\sqrt{G}$, G_β

Yes



See G. (2017; https://doi.org/ 10.1175/WAF-D-16-0134.1), G. et al. (2020; https://doi.org/10.1175/MWR-D-19-0256.1) and G. (2021; https://doi.org/10.5194/ascmo-7-13-2021) for more details.

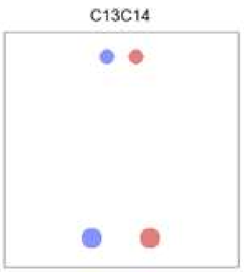
Spatial Dissimilarity Measures

Does the measure provide useful information about rare (i.e., spatially small) events?

Baddeley's Δ
centroid distance, dFSS (undefined), $\sqrt{G}$ — **No**

Hausdorff distance — **Yes**

MED, G_β — **Qualified Yes**



See G. (2017; https://doi.org/ 10.1175/WAF-D-16-0134.1), G. et al. (2020; https://doi.org/10.1175/MWR-D-19-0256.1) and G. (2021; https://doi.org/10.5194/ascmo-7-13-2021) for more details.

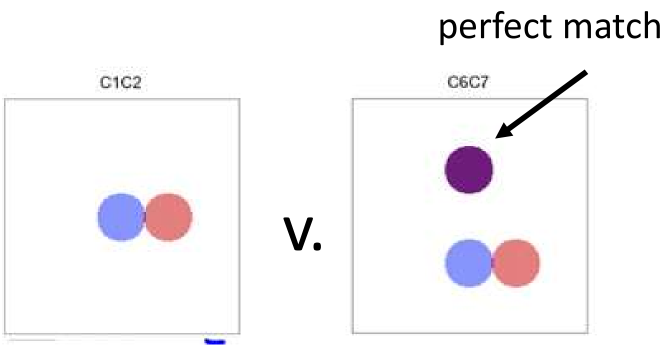
Spatial Dissimilarity Measures

Does the measure reward for partially perfect matches?

Hausdorff distance, centroid distance, $\sqrt{G}$, G_β ✗ **No**

Baddeley's Δ, dFSS ✓ **Yes**

MED ● **Qualified Yes**



See G. (2017; https://doi.org/ 10.1175/WAF-D-16-0134.1), G. et al. (2020; https://doi.org/10.1175/MWR-D-19-0256.1) and G. (2021; https://doi.org/10.5194/ascmo-7-13-2021) for more details.

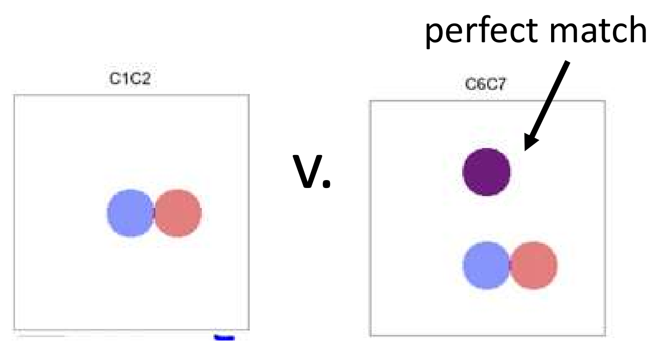
Spatial Dissimilarity Measures

Does the measure correctly penalize despite a partially perfect match?

Hausdorff distance, Baddeley's Δ, dFSS, centroid distance — **No**

$\sqrt{G}, G_{\beta}$ — **Yes**

MED — **Qualified Yes**



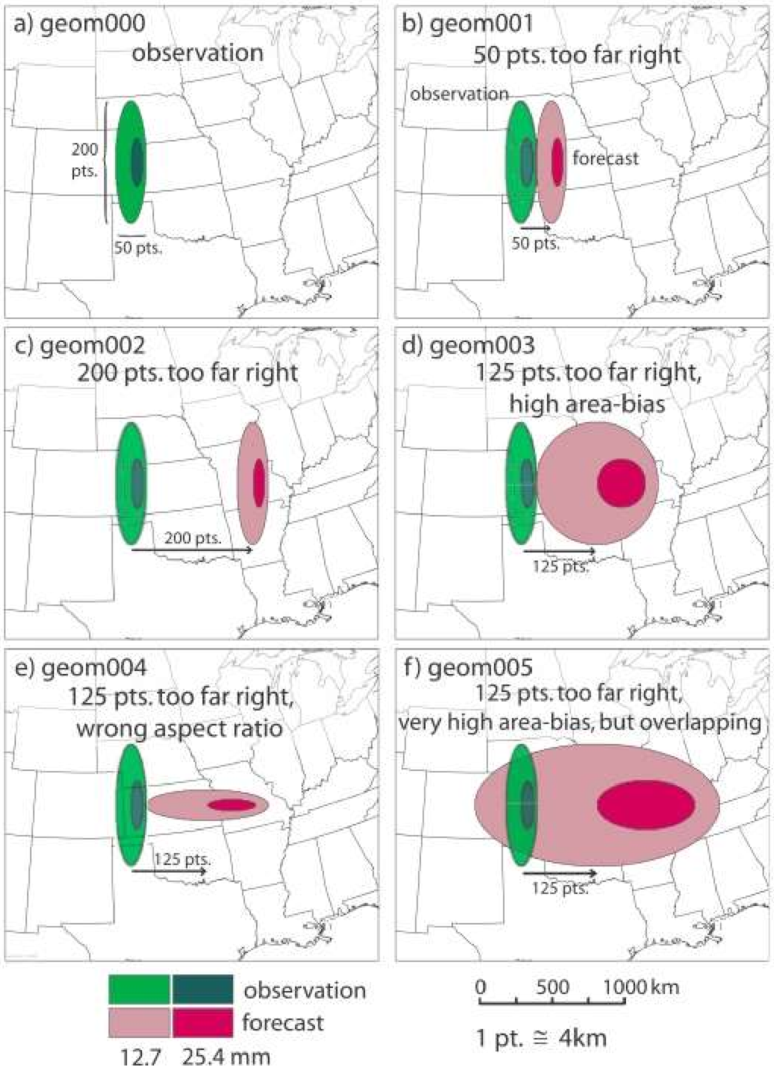
See G. (2017; https://doi.org/ 10.1175/WAF-D-16-0134.1), G. et al. (2020; https://doi.org/10.1175/MWR-D-19-0256.1) and G. (2021; https://doi.org/10.5194/ascmo-7-13-2021) for more details.

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

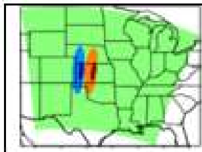
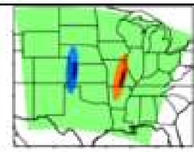
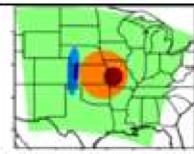
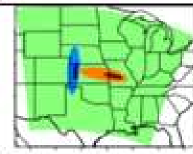
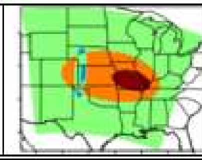
*Depending on choice of β

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

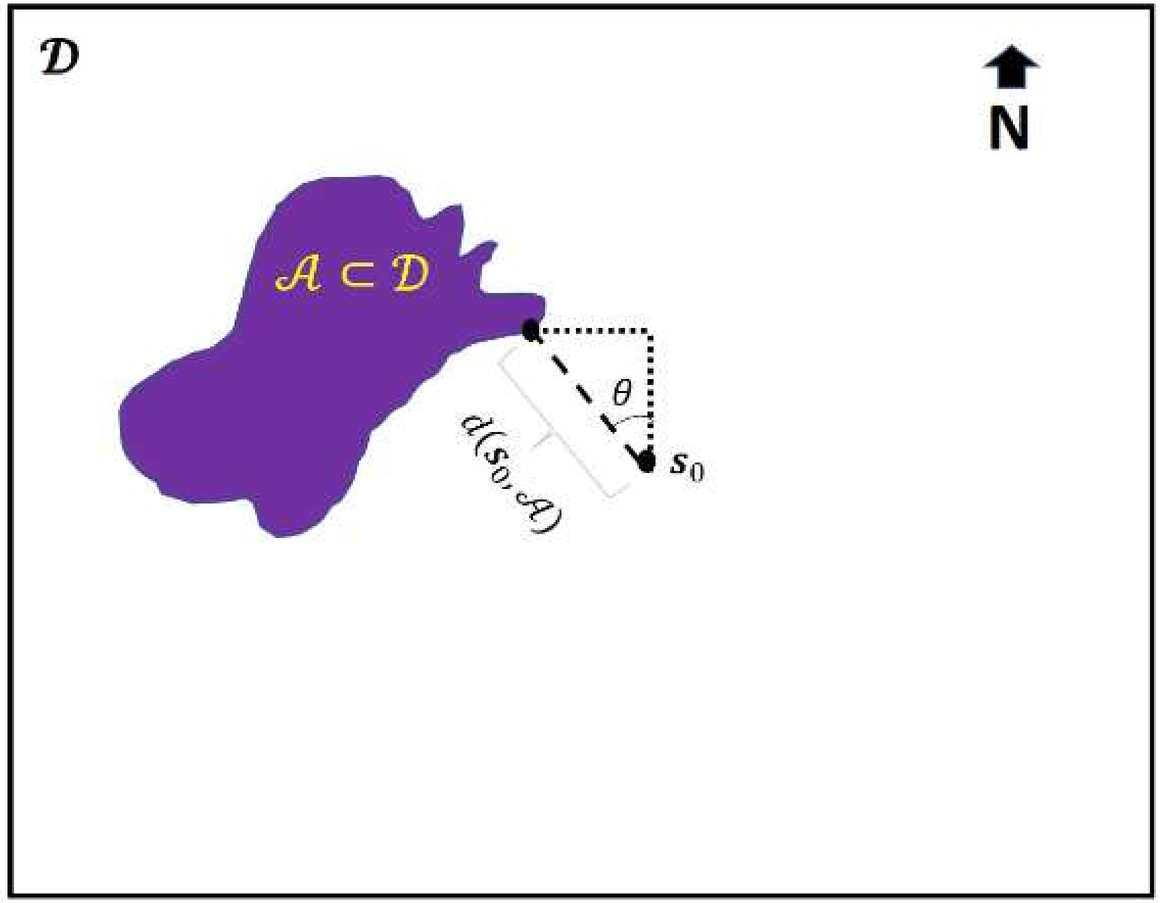


Traditional score	geom001/002/004	geom003	geom005
Accuracy	0.95	0.87	0.81
Frequency bias	1.00	4.02	8.03
Multiplicative intensity bias	1.00	4.02	8.04
RMSE (mm)	3.5	5.6	6.9
Bias-corrected RMSE (mm)	3.5	5.5	6.3
Correlation coefficient	−0.02	−0.05	0.20
Probability of detection	0.00	0.00	0.88
Probability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

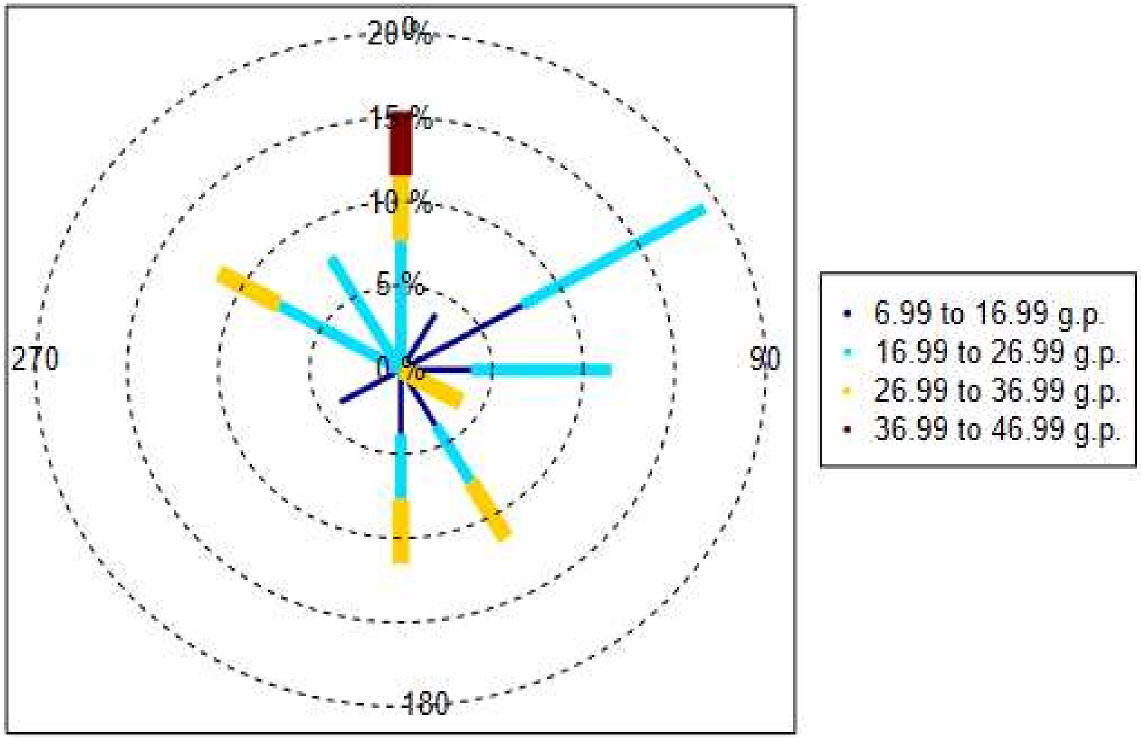
					
Method	geom001 translation-only error (50-pts)	geom002 translation-only (200-pts)	geom003 translation (125-pts) and large area bias	geom004 translation (125-pts) and aspect-ratio	geom005 translation (125-pts) and huge area bias (but overlapping)
$H(A,B)$	**Best**	**Tied for 2**	**Tied for 2**	**Tied 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**

Point-to-grid?

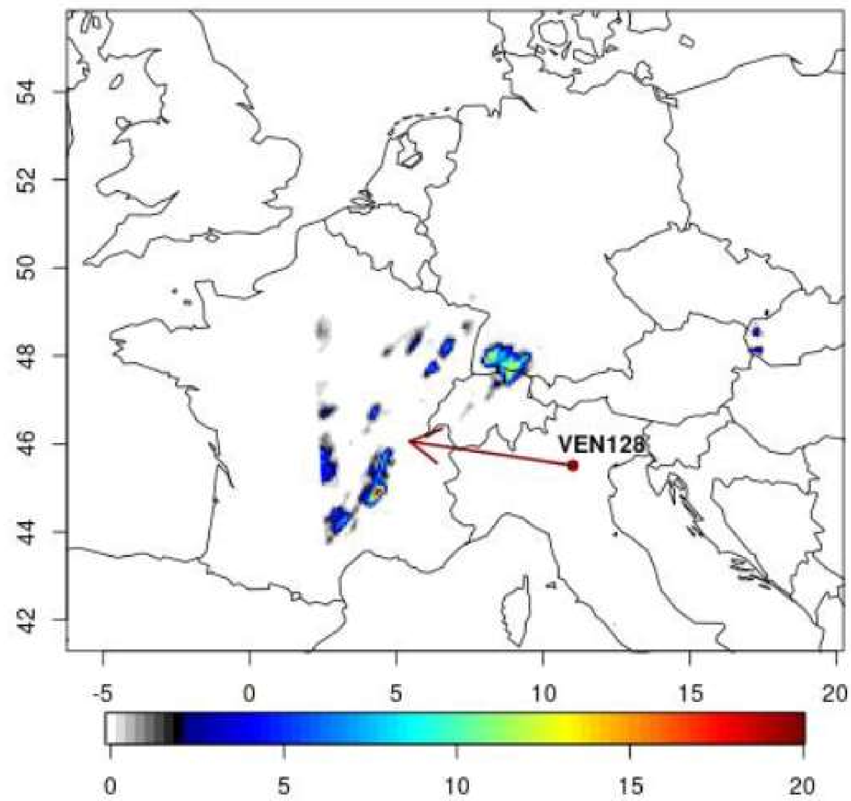


cf. Brunet et al. 2018, doi:10.1127/metz/2018/0883

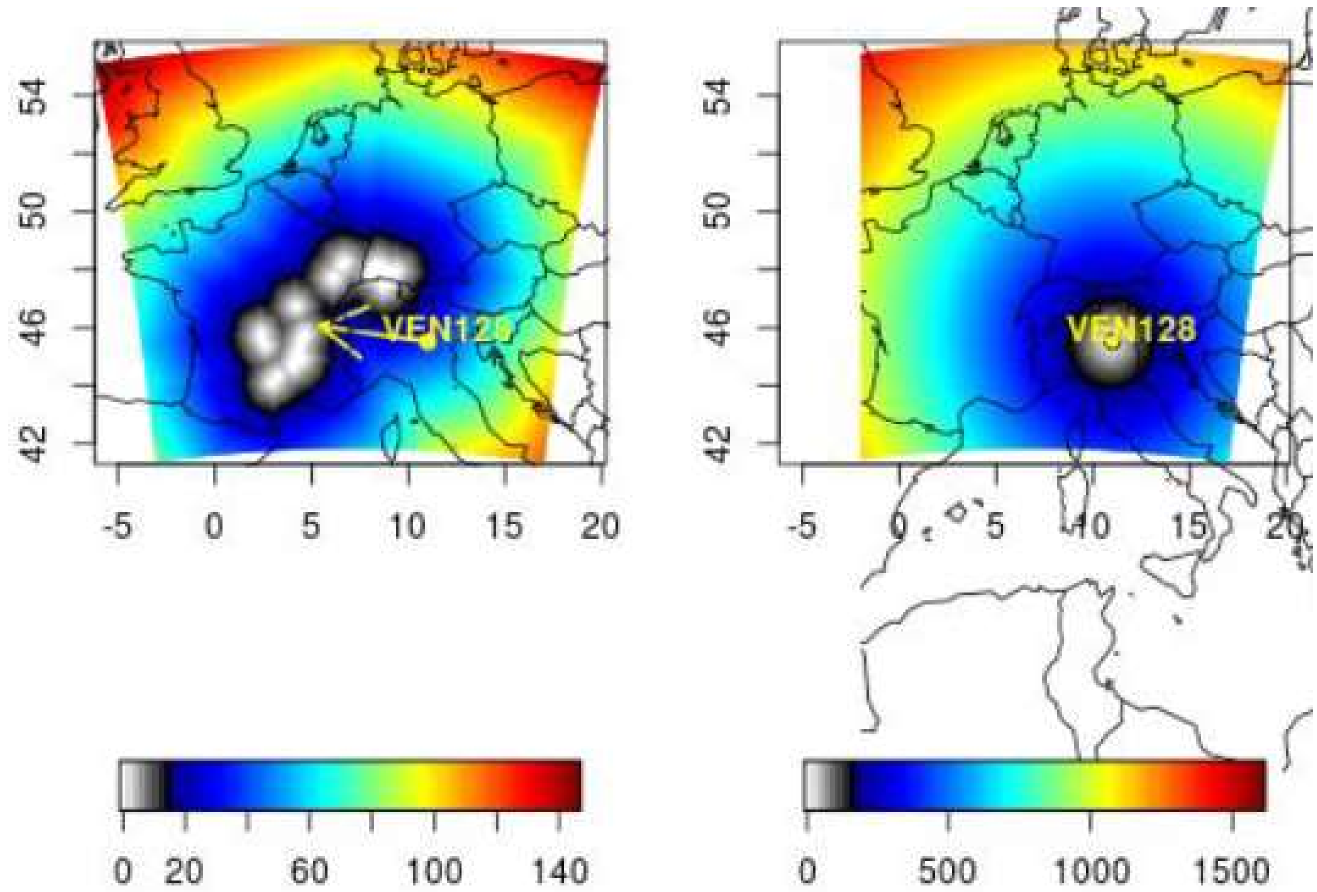
Point-to-grid?



Point-to-grid?

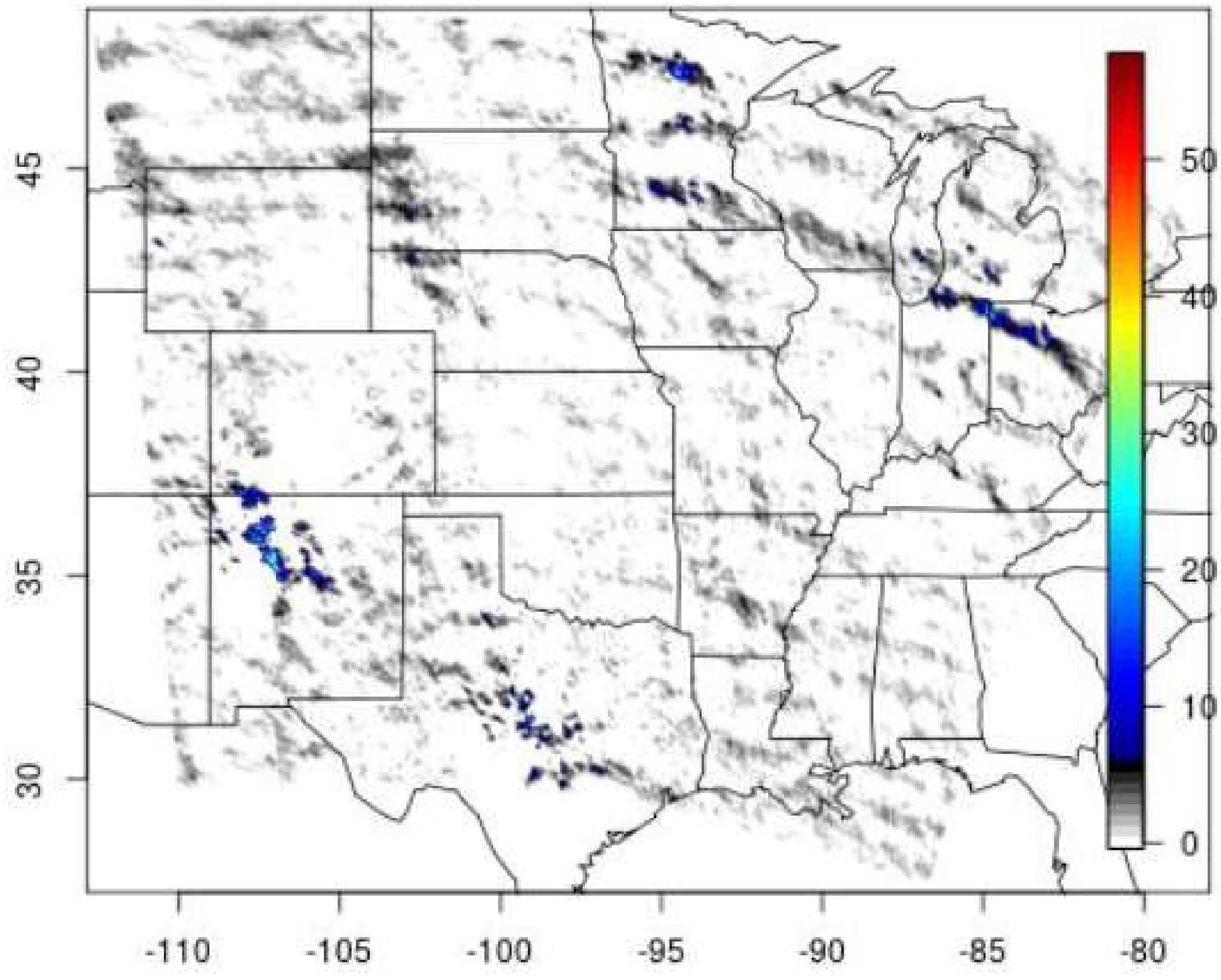


Point-to-grid?



24 June 2021

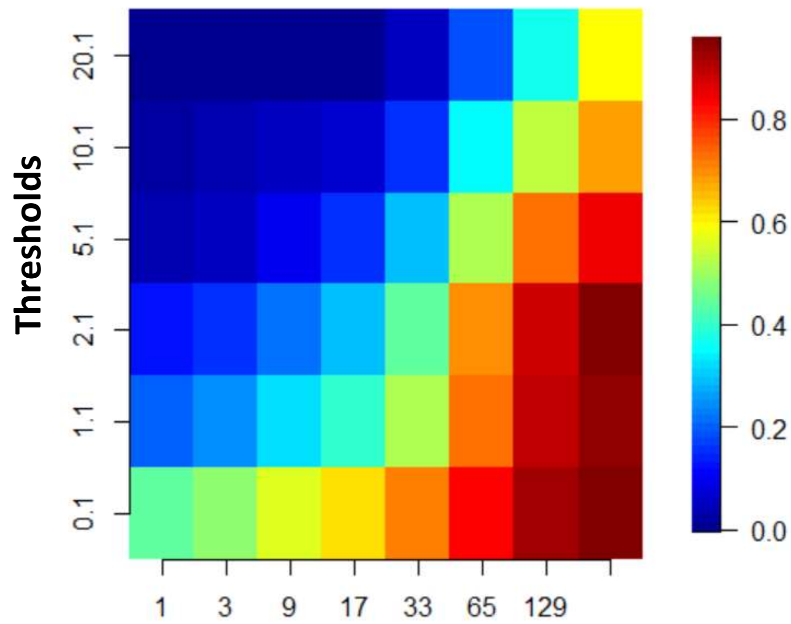
Uncertainty Estimation



Neighborhood Methods

Do not require an exact match between forecast and observations

- Identify unpredictable scales
- Identify scales of predictability
- See Ebert (2008, https://doi.org/10.1002/met.25) for a review of neighborhood methods.



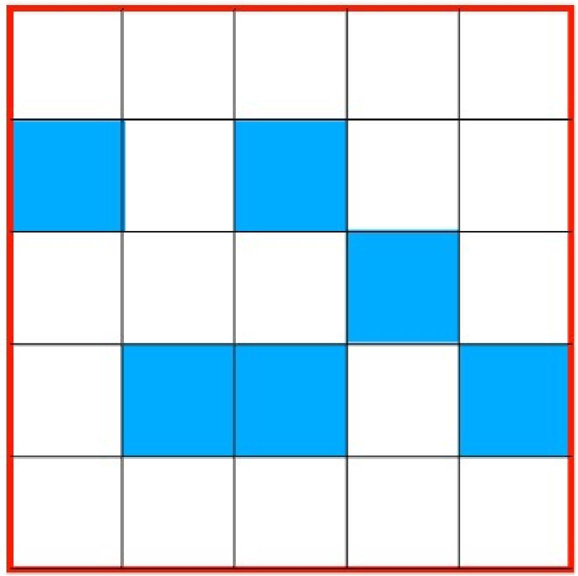
Neighborhood size (grid squares)

Neighborhood Methods

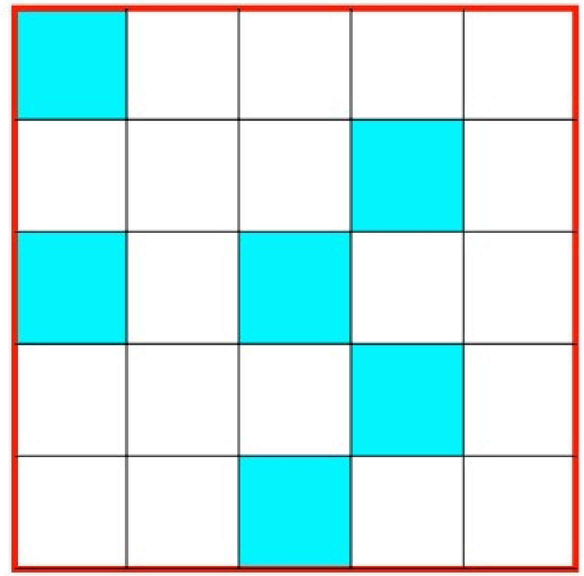
Fractions Skill Score

$$\text{FSS}=1-\frac{\text{MSE}(n)}{\text{MSE(reference, n)}}$$

$$=1-\frac{\frac{1}{N}\sum_i^N\left(\hat{f}_i-f_i\right)^2}{\frac{1}{N}\sum_i^N \hat{f}_i^2+\frac{1}{N}\sum_i^N f_i^2}$$



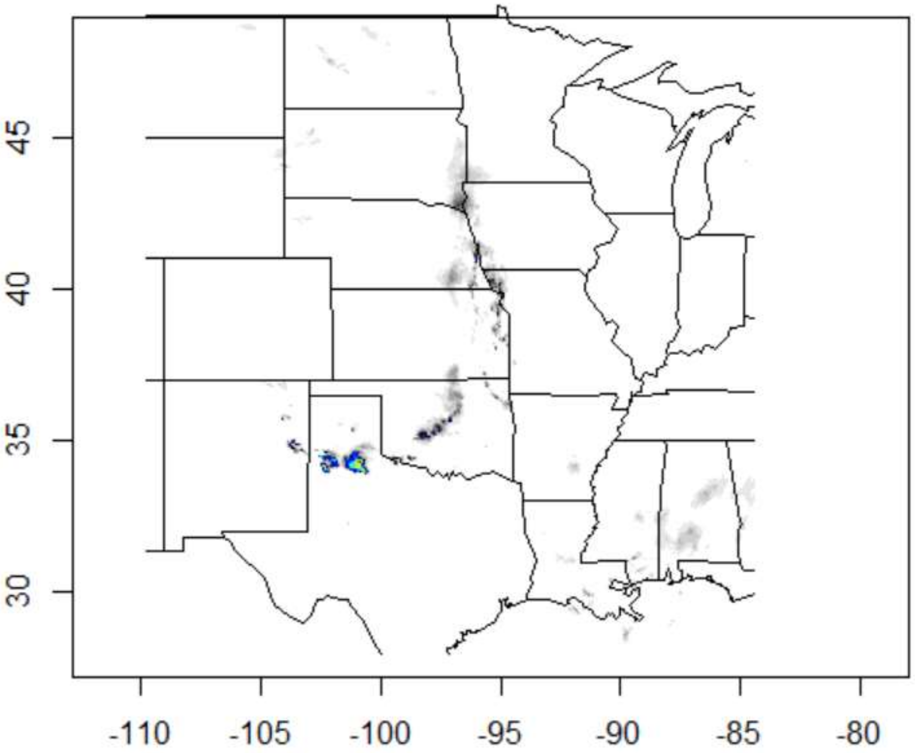
Fraction = 6/25 = 0.24



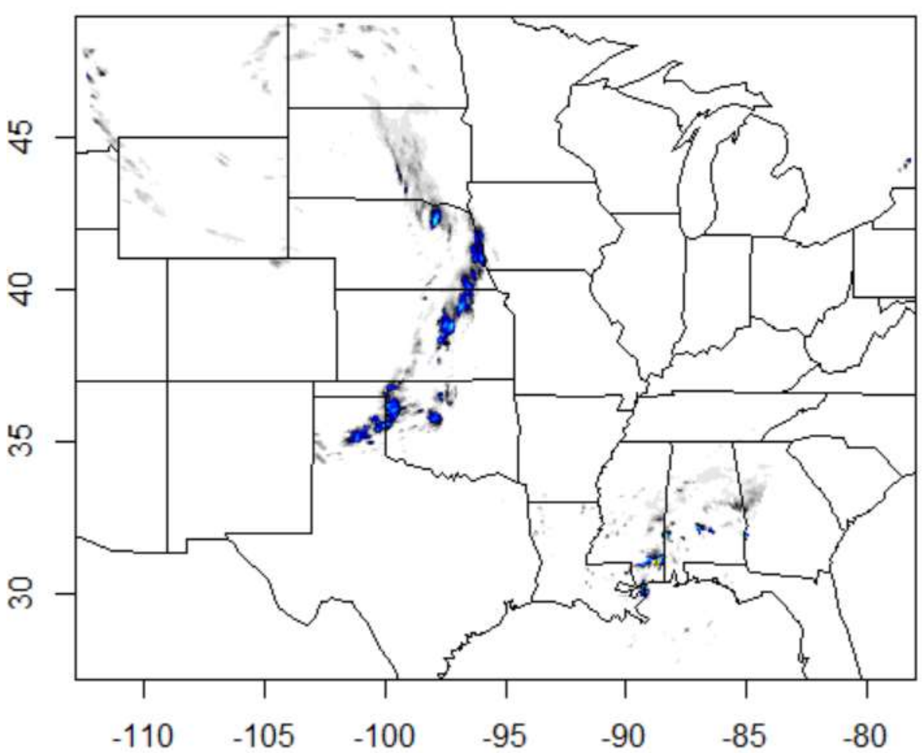
Fraction = 6/25 = 0.24

FSS Example

Observation

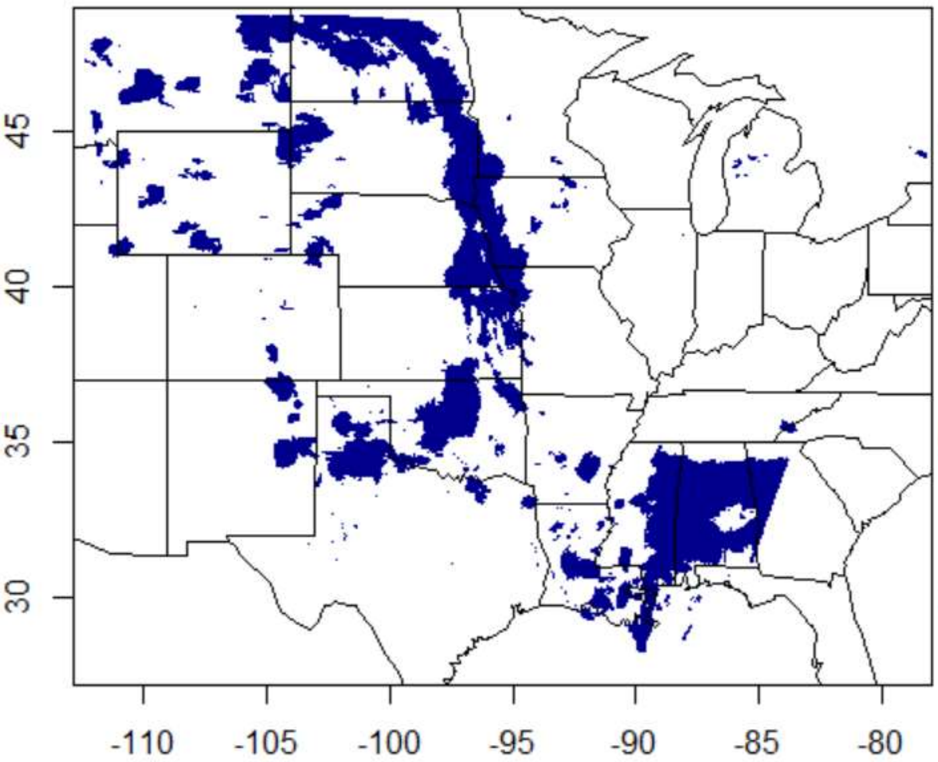


Forecast

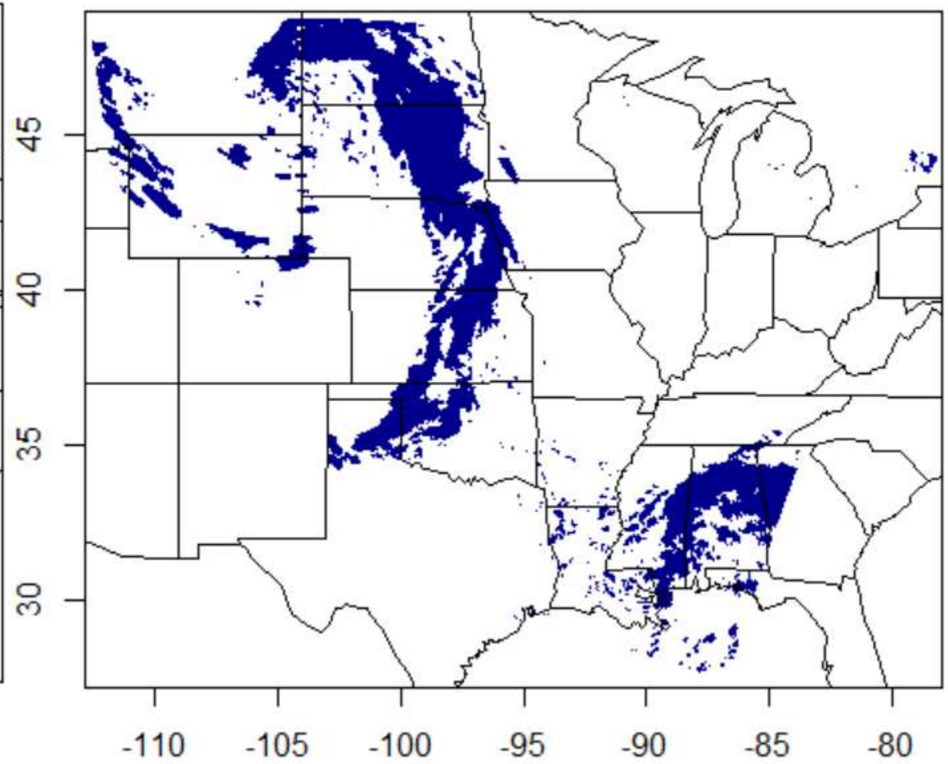


FSS Example

Observation

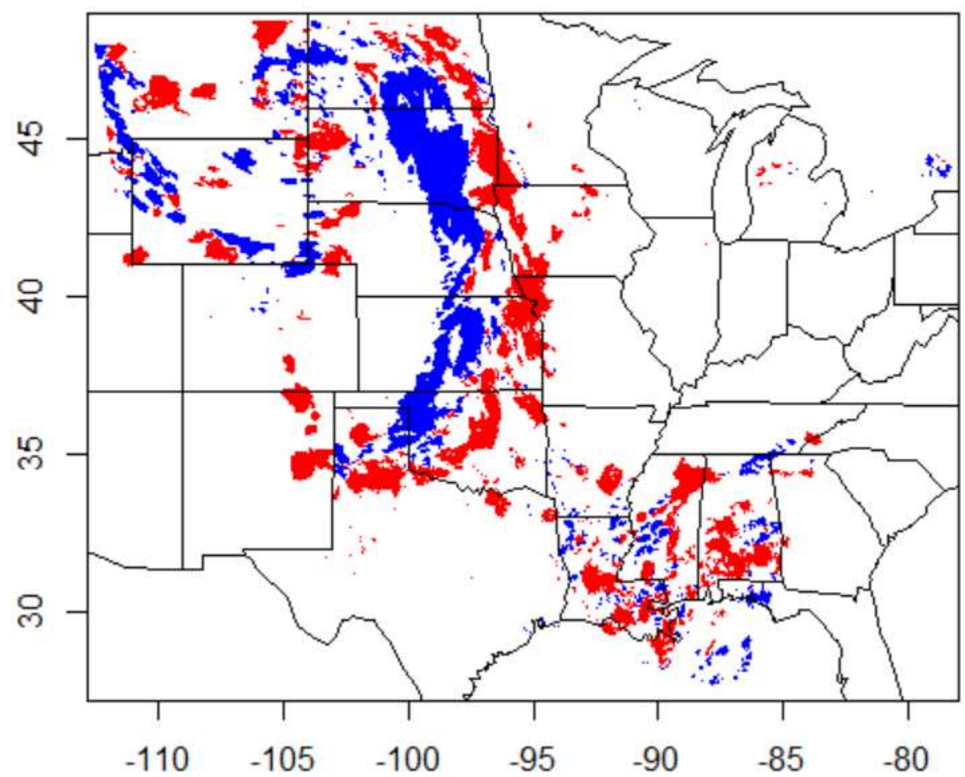


Forecast

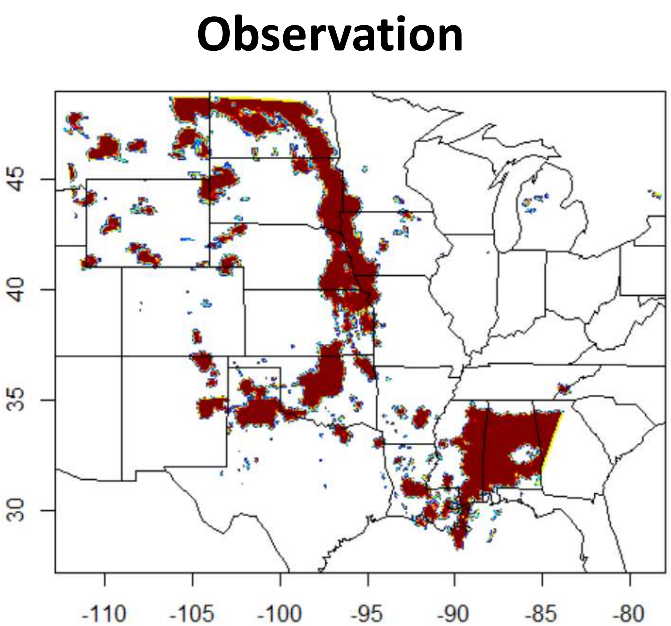


FSS Example

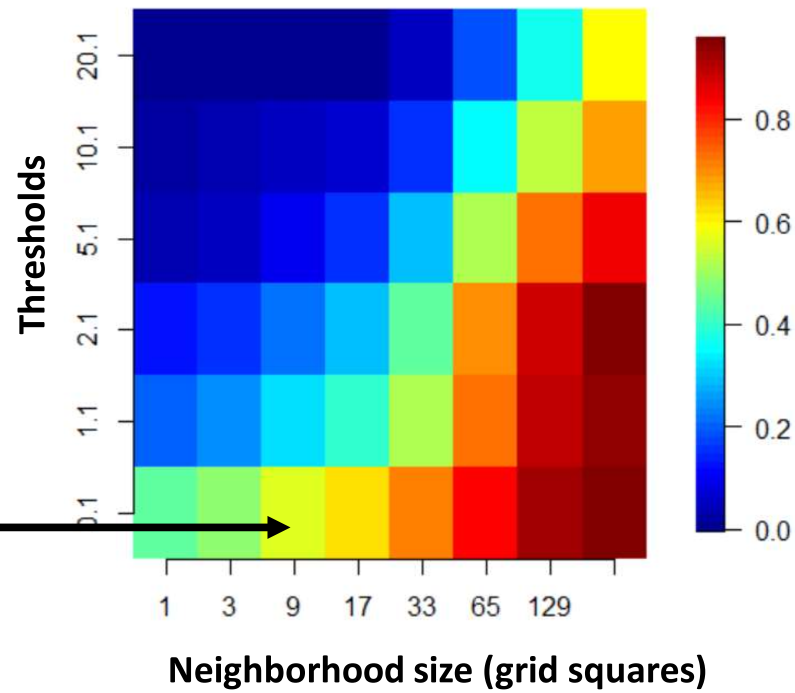
Binary Forecast — Binary Observation



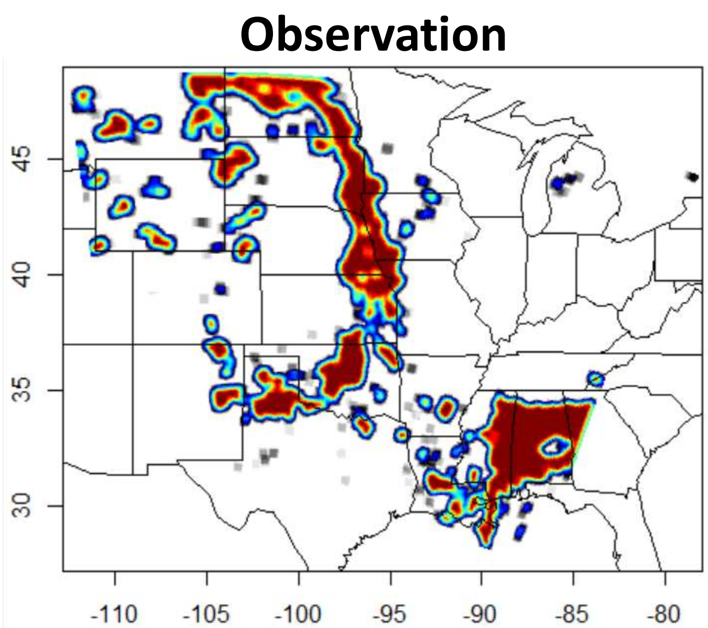
Neighborhood Methods



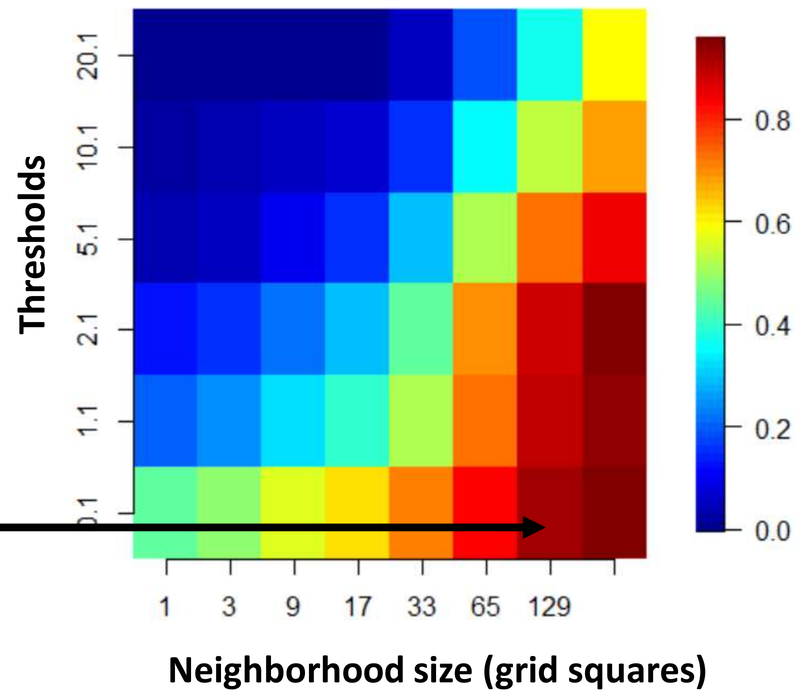
Binary field created using threshold of 0.1 mm/h and smoothed using a square neighborhood of size 9.



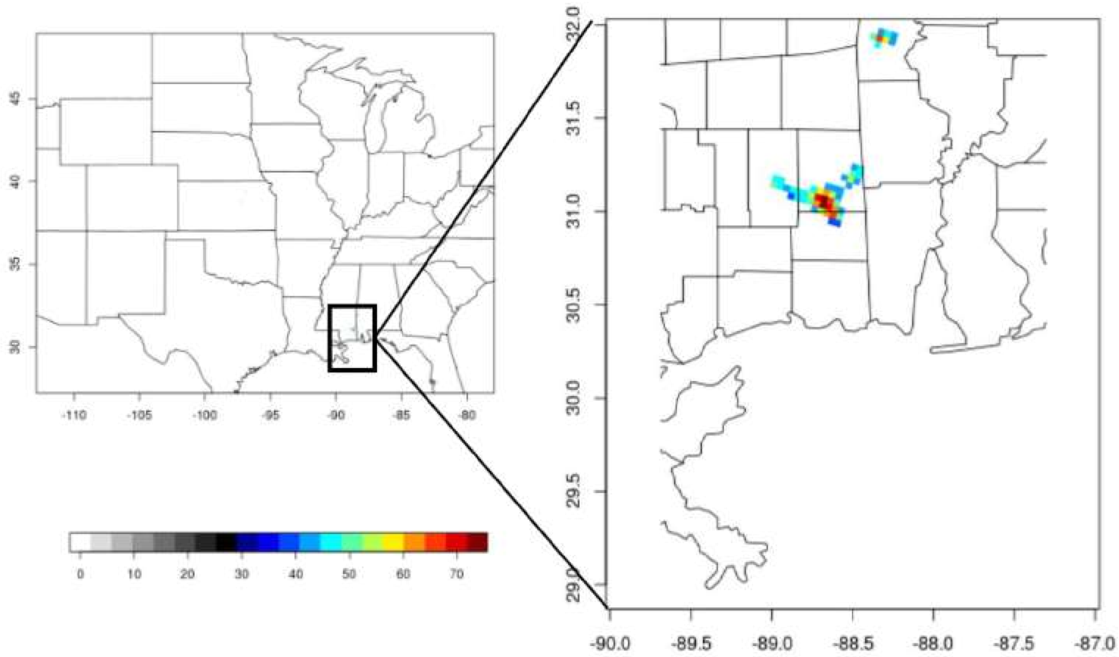
Neighborhood Methods



Binary field created using threshold of 0.1 mm/h and smoothed using a square neighborhood of size 129.

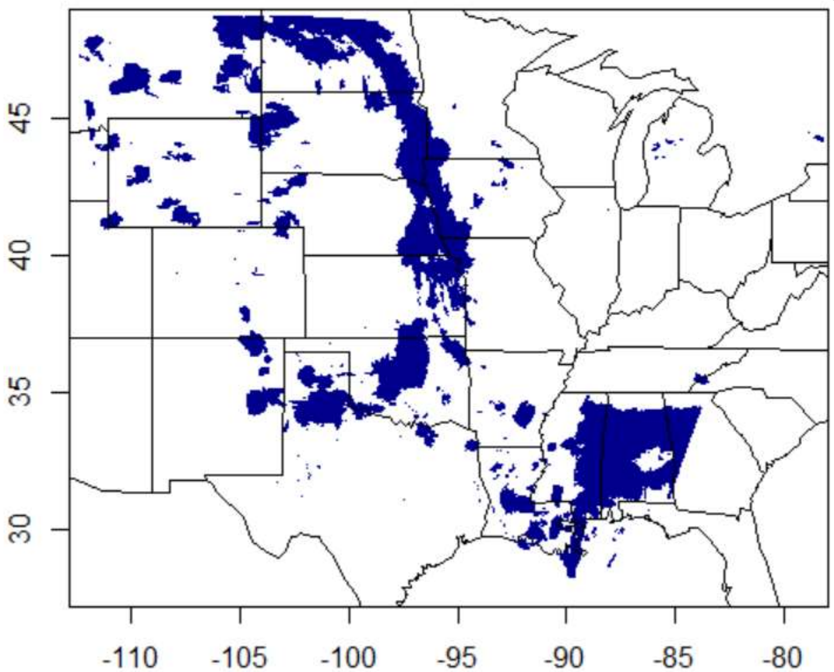


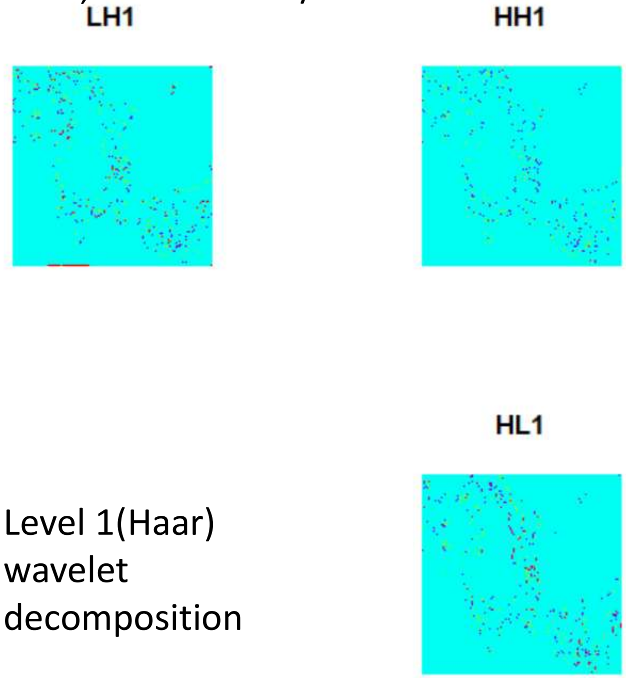
FSS Example


-110 -105 -100 -95 -90 -85 -80
30 35 40 45
0 10 20 30 40 50 60 70
32.0 31.5 31.0 30.5 30.0 29.5 29.0
-90.0 -89.5 -89.0 -88.5 -88.0 -87.5 -87.0

ISS Example (Scale Decomposition)

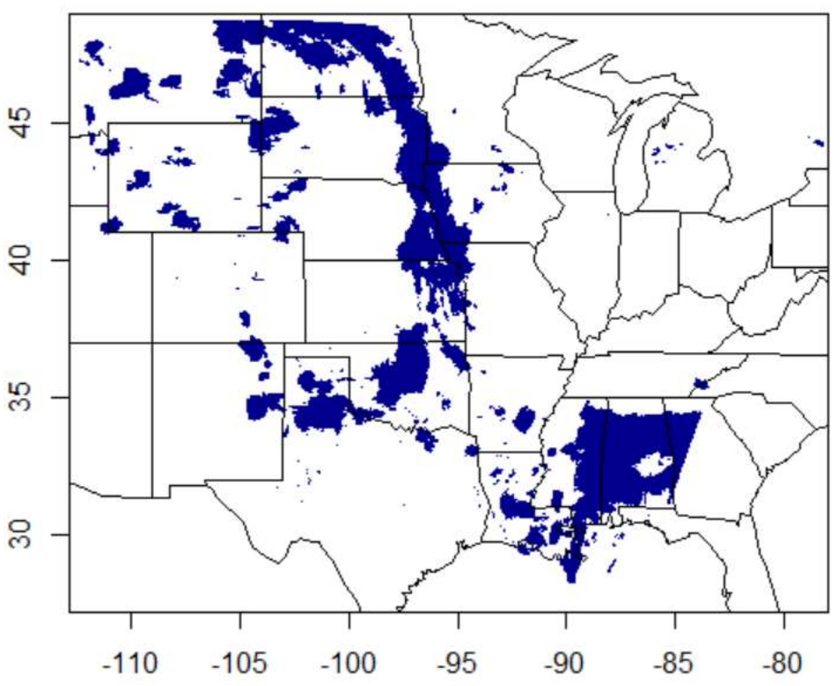
Casati et al. (2004, doi: 10.1017/S1350482704001239)
Casati (2010, doi: 10.1175/2009WAF2222257.1)

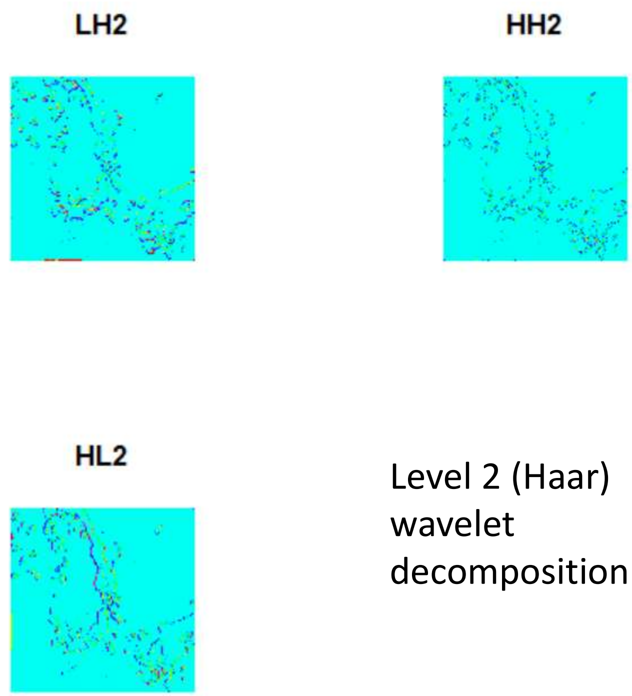




Level 1(Haar)
wavelet
decomposition

ISS Example (Scale Decomposition)





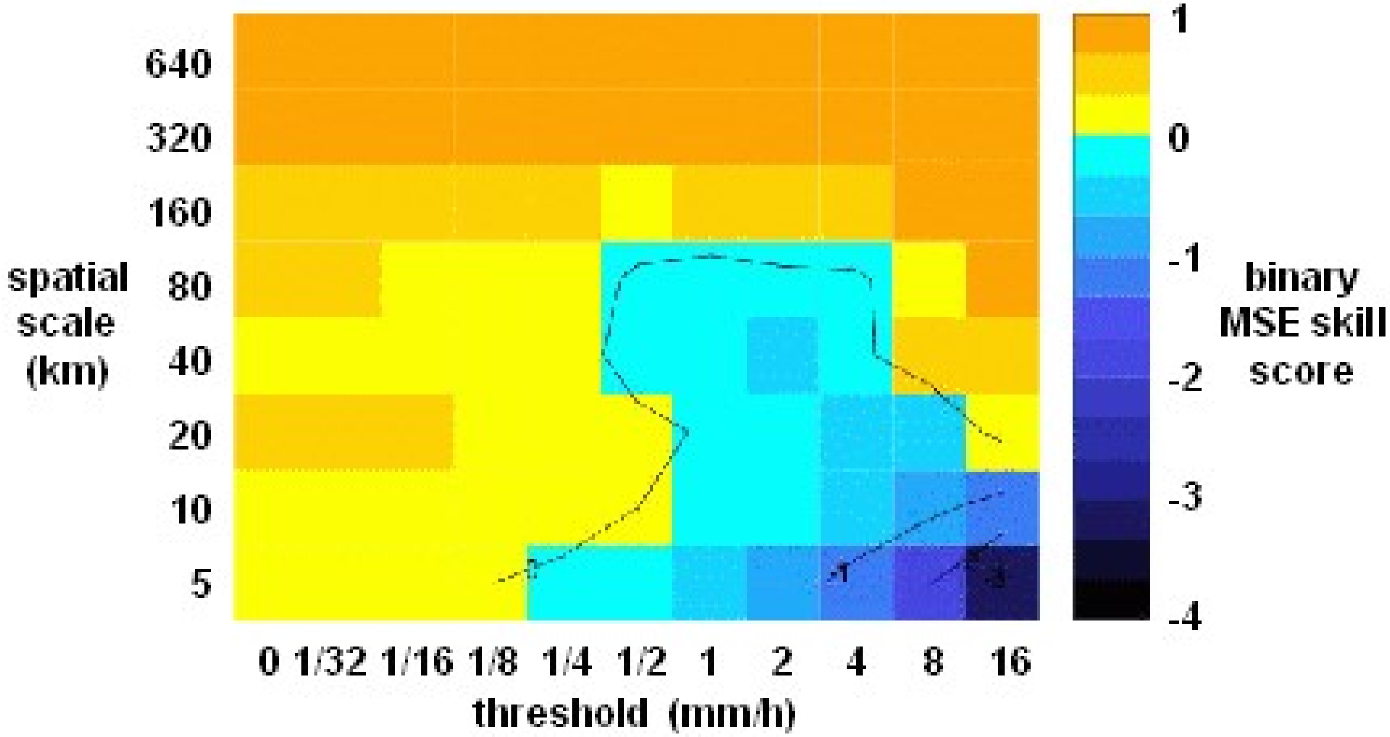
Level 2 (Haar) wavelet decomposition

24 June 2021

MSE skill score

$$SS_{u,j} = \frac{MSE_{u}, l - MSE_{u,l,random}}{MSE_{u,j,best} - MSE_{u,l,random}}$$

Sample climatology
(base rate)



Scale Decomposition

- Review of wavelet methods (Weniger et al., 2016, https://doi.org/10.1002/qj.2881).
- Alternative ways to do wavelet decomposition spatially
- Alternative multi-resolution analysis (MRA) that is naturally spatial
  - G. and Nychka (2009, http://n2t.net/ark:/85065/d7zg6r75) gives a simple introduction to the idea using thin-plate splines (but not practical for the sizes of data employed in operational forecasting systems)
  - Nychka et al. (2013, http://n2t.net/ark:/85065/d7w095cm) and Nychka et al. (2015, https://doi.org/10.1080/10618600.2014.914946) demonstrate how this type of MRA can be applied to large data sets.

Software

- Model Evaluation Tools (Brown et al. 2021, doi: 10.1175/BAMS-D-19-0093.1)
- SpatialVx (https://ral.ucar.edu/staff/ericg/Gilleland2021-SpatialVx.pdf)
- Others??
  - https://ral.ucar.edu/projects/icp/
  - https://projects.ral.ucar.edu/icp/

Summary

- Need to think spatially when verifying over space!
  - How do you obtain observations to compare against a gridded forecast?
  - How do you compare two forecasts that are on different grids?
  - If using an analysis for observations, is the model you used the same as the one for the forecast?
- Ideally one would re-align the forecast to better match the observations spatially before conducting any verification
  - Lots of difficulties in the implementation of these methods
- Feature-based methods provide a lot of useful information that allows for analyzing forecast performance in spatially meaningful ways.
- Spatial dissimilarity measures are simple and fast to calculate.
  - Need to consider how they inform about different types of errors.
  - Can be applied to entire fields or to features within a feature-based approach like MODE.
- Smoothing filter methods are much more limited than their popularity would suggest.
- Band-pass filters are potentially highly useful, particularly for uncertainty estimation, and can be potentially useful in conjunction with other methods.