# Spatial Forecast Verification

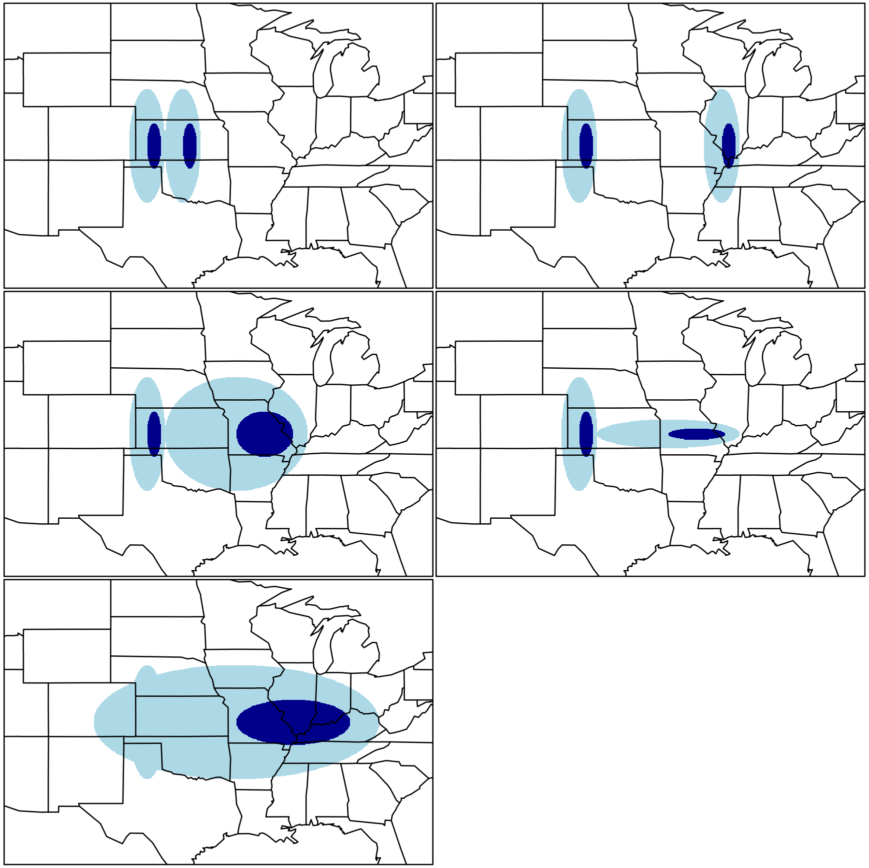
2010 ASP Colloquium

Eric Gilleland
E-mail: EricG @ ucar.edu
Research Applications Laboratory,
*National Center for Atmospheric Research*

Co-authors: D. Ahijevych, B.G. Brown, B. Casati, and E.E. Ebert

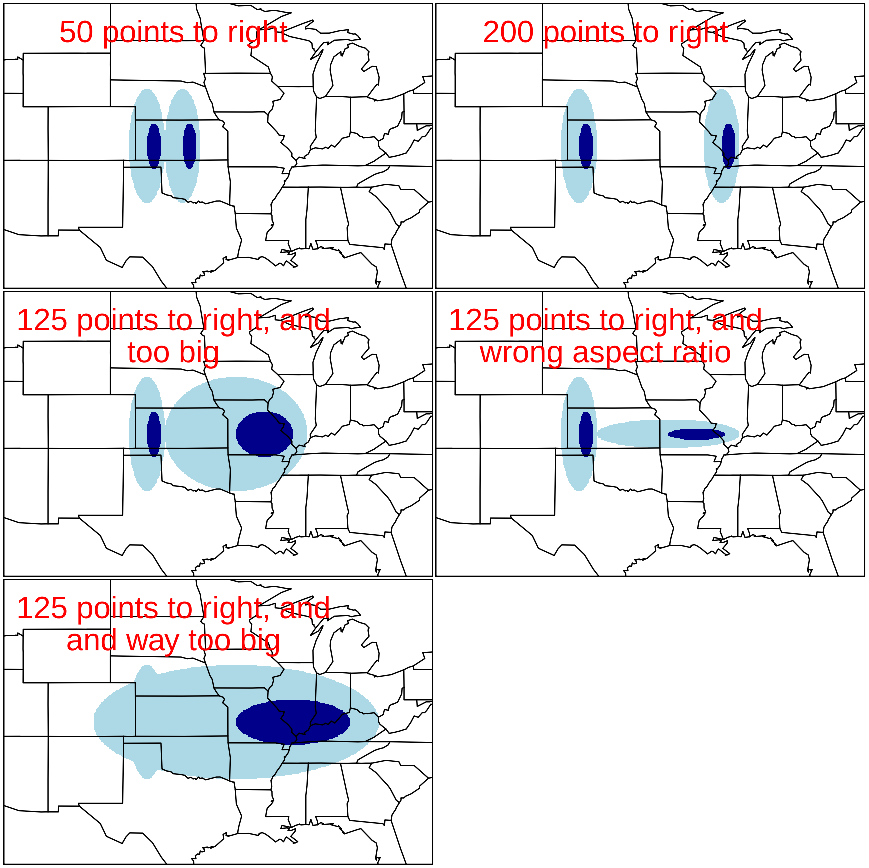
# Spatial Forecast Verification Methods

## Inter-Comparison Project (ICP)



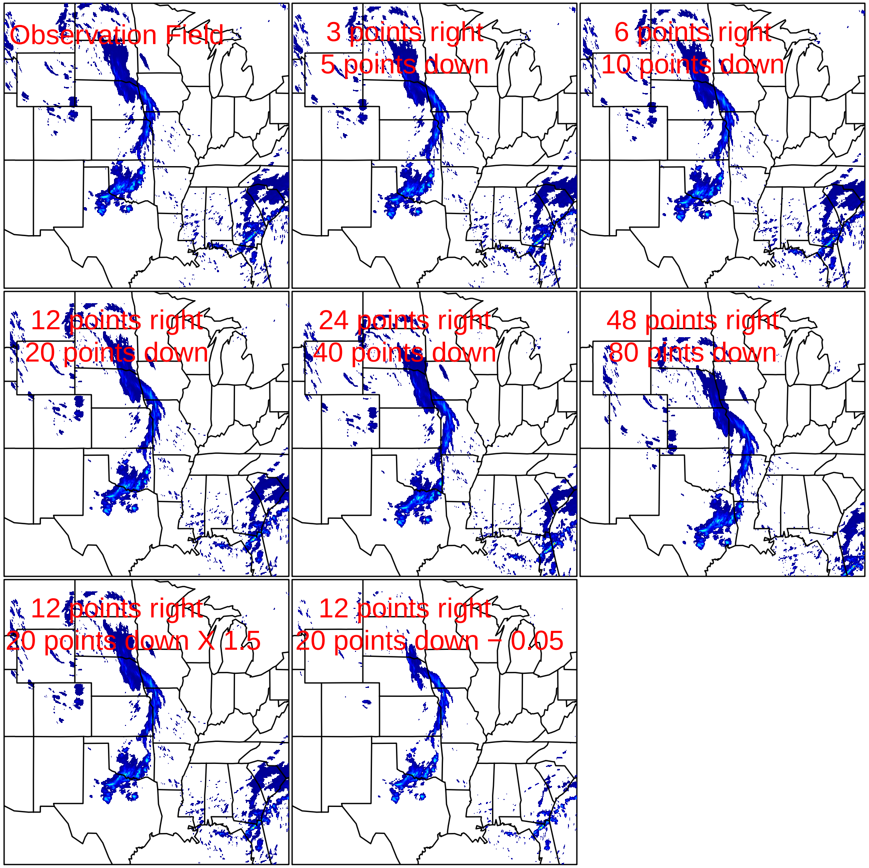
# Spatial Forecast Verification Methods

## Inter-Comparison Project (ICP)



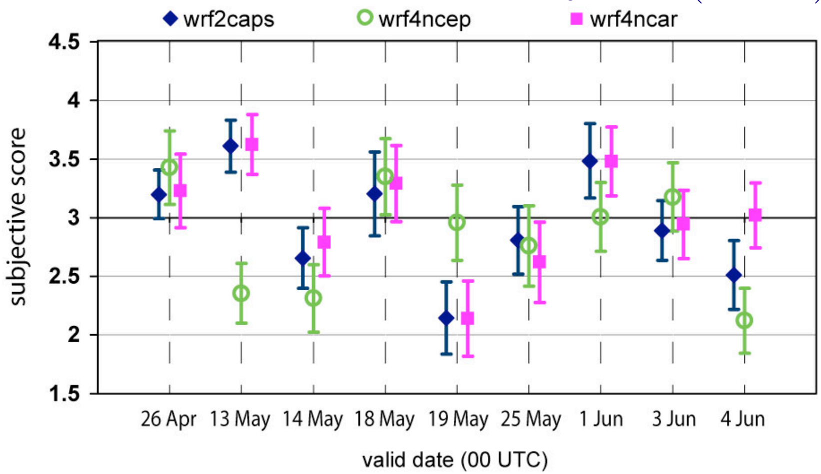
# Spatial Forecast Verification Methods

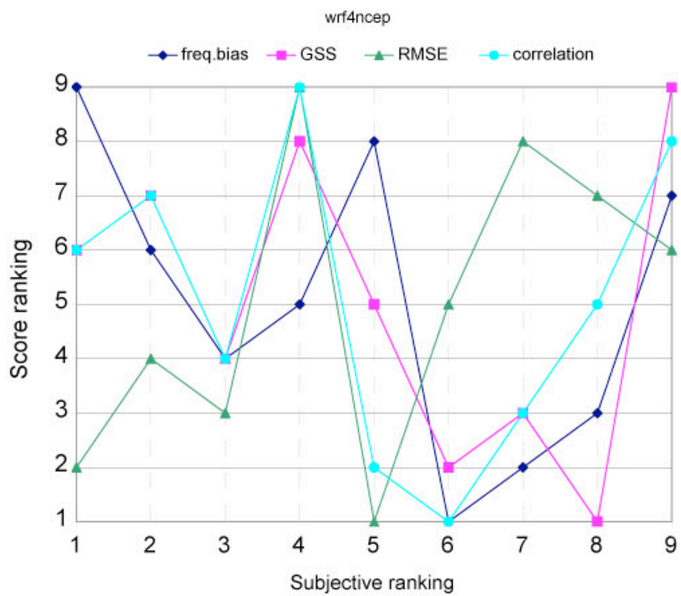
## Inter-Comparison Project (ICP)



# Spatial Forecast Verification Methods

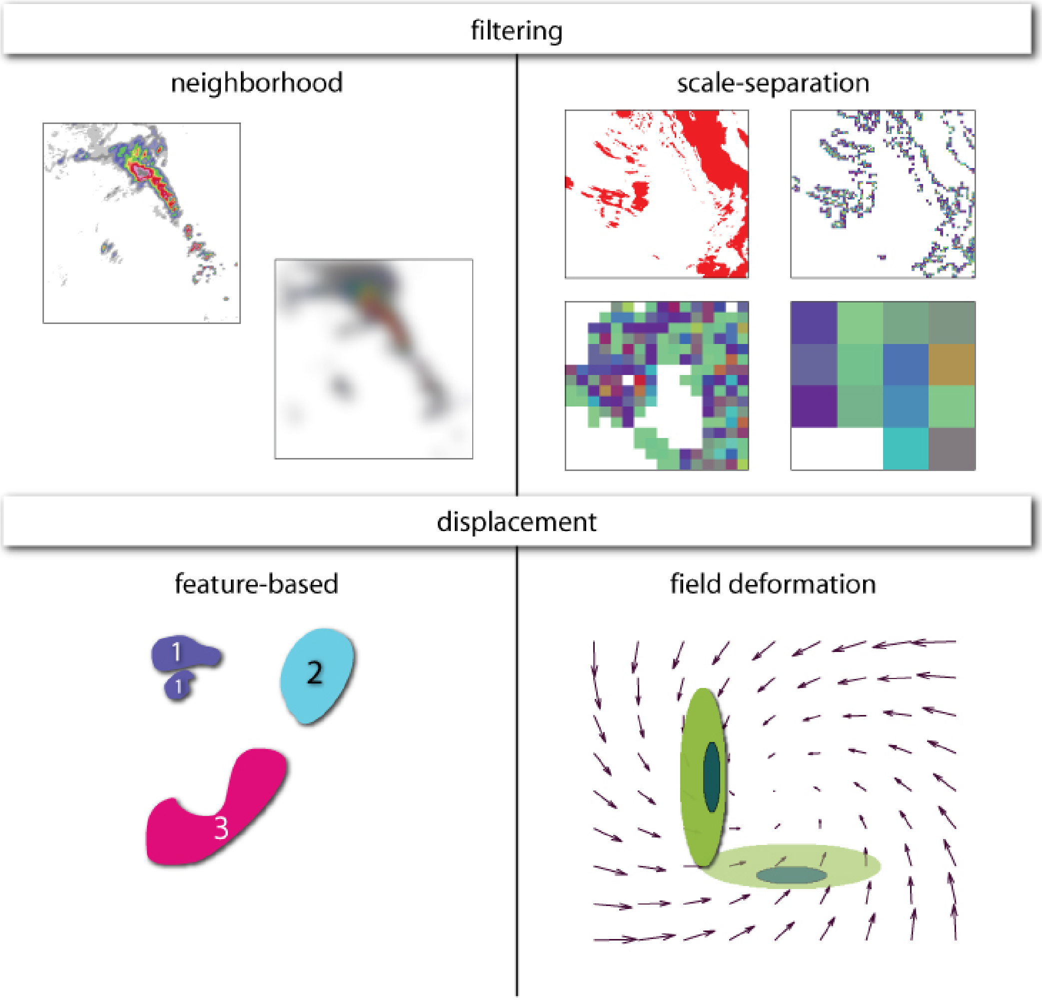
## Inter-Comparison Project (ICP)



# Spatial Forecast Verification Methods

## Inter-Comparison Project (ICP)



# Filter Methods : Neighborhood

Model: $M_j(O(x,y)) = B_j(F(x,y)) + \varepsilon_j$

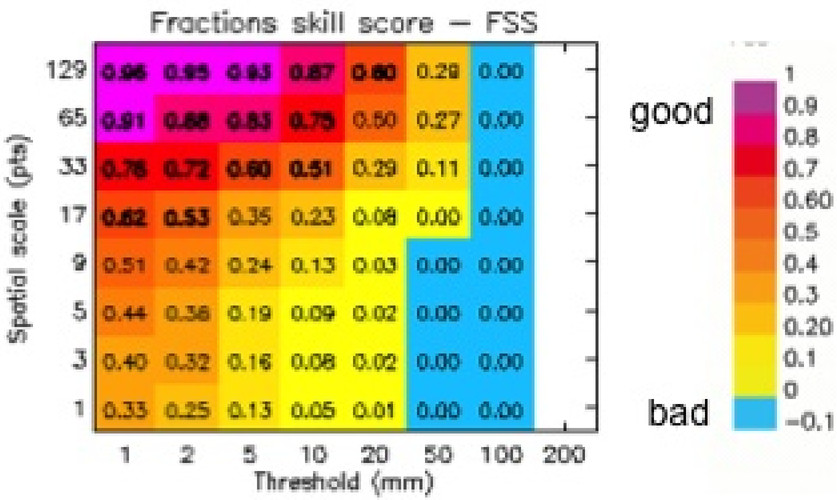
Goal: Examine forecast performance in a region without requiring exact grid-point to grid-point matches.

Advantages: Generally straightforward; Provide useful information about forecast performance. Less sensitive to small localized errors. Physical interpretations possible (e.g., scales where forecasts have skill).

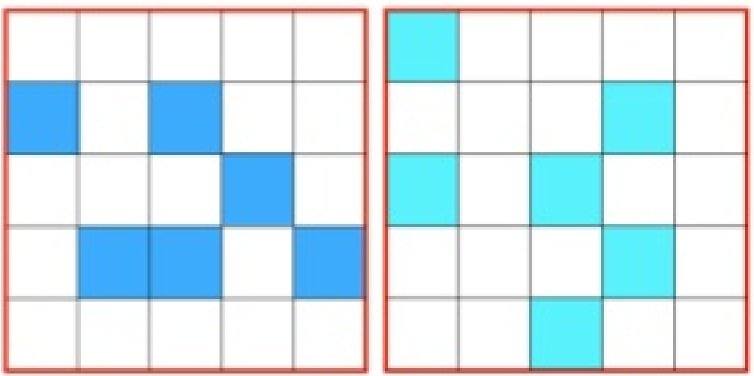
Disadvantages: Limited diagnostic information. Do not inform about specific error types, but may be sensitive to them. Do not inform about spatial structure errors.

Examples: Simplest example is *upscaling*. Many such methods have been proposed (Ebert, 2008 gives a nice review).

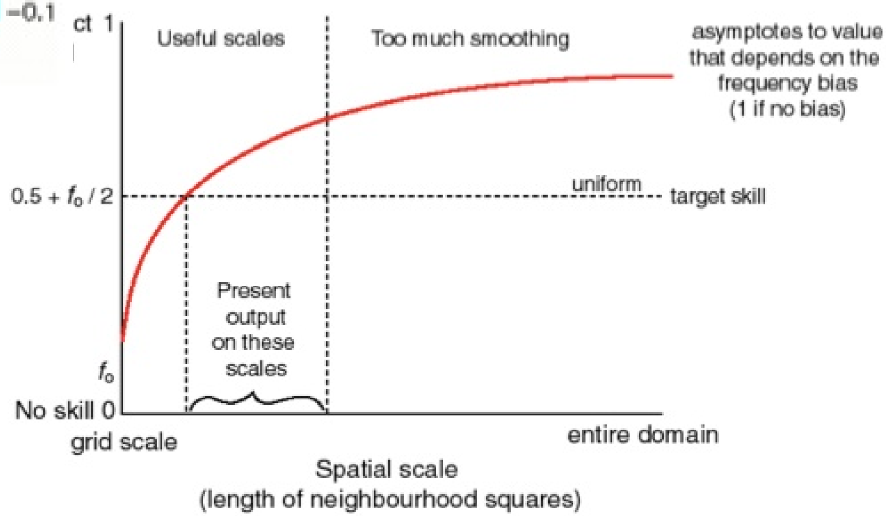
# Filter Methods: Neighborhood




$$\mathrm{FSS} = 1 - \frac{\sum_{i=1}^{N} (P_{\mathrm{fcst}} - P_{\mathrm{obs}})^2}{\sum_{i=1}^{N} P_{\mathrm{fcst}}^2 + \sum_{i=1}^{N} P_{\mathrm{obs}}}$$



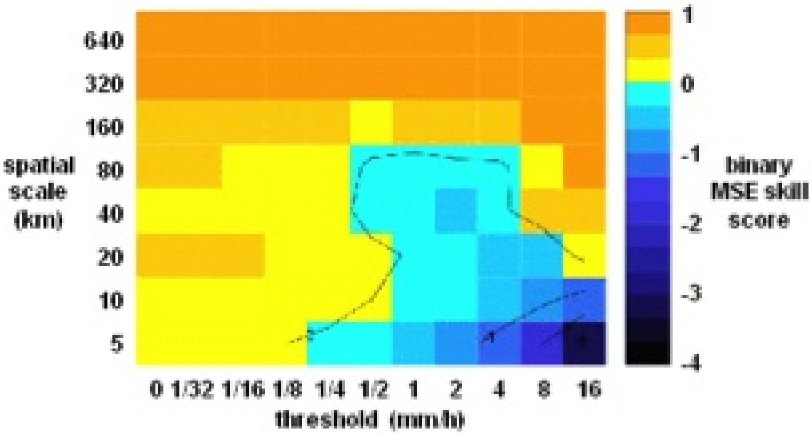
Fraction = 6/25 = 0.24 Fraction = 6/25 = 0.24




Fractions Skill Score (Robert and Lean, 2008)

# Filter Methods: Scale Separation

Similar to neighborhood methods, can inform about scale, but now scales are *independent*.




Examples of filters: Fourier decomposition, Wavelets, etc. Variograms (Marzban and Sandgathe, 2009)

Power spectra (Harris *et al.*, 2001)

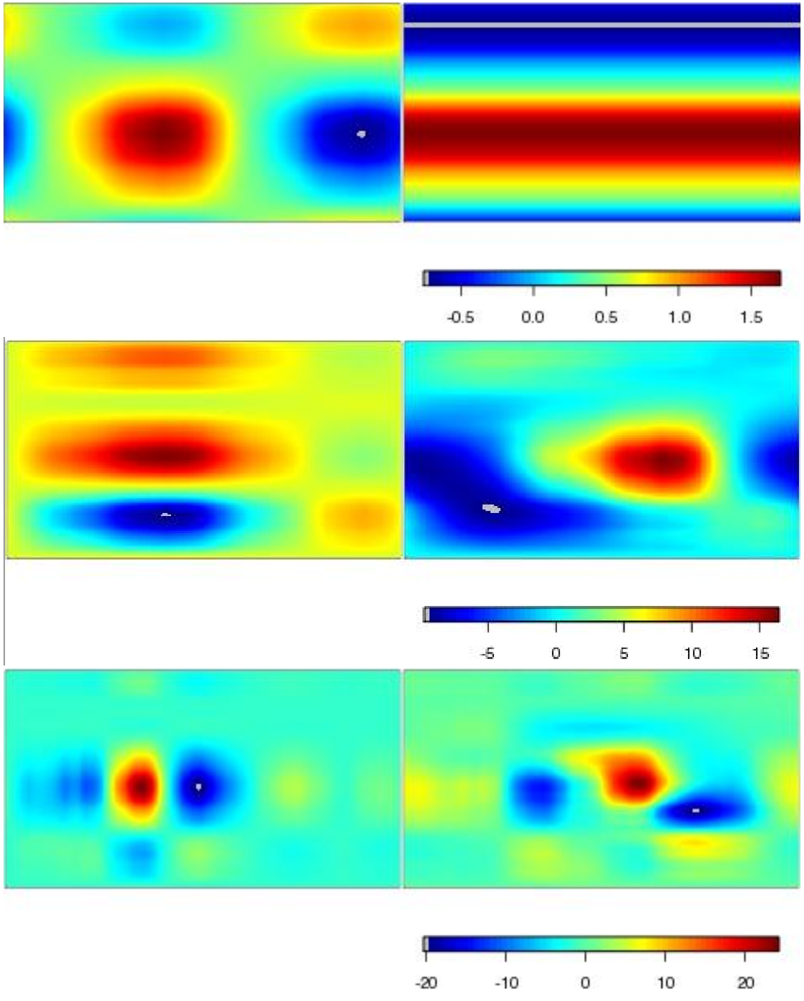
Wavelets (Briggs and Levine, 1996)

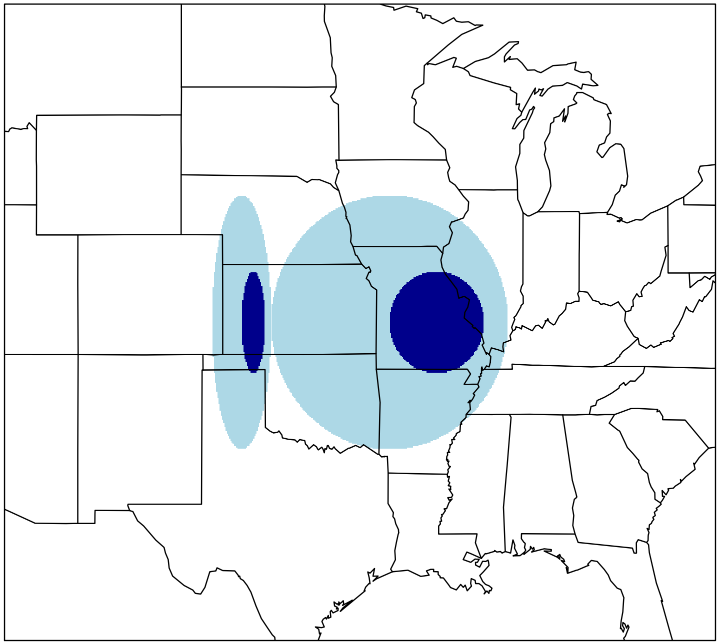
Intensity Scale (IS): (Casati *et al.*, 2004) (wavelets applied to binary event fields)

Multi-scale variability (Zapeda-Arce *et al.*, 2000; Harris *et al.*, 2001; Mittermaier 2006)

# Filter Methods: Scale Separation

## Wavelets



# Filter Methods: Scale Separation

Advantages: Informs about skill at *specific* (possibly physically meaningful) scales. Like neighborhood methods, typically use traditional scores at different scales and thresholds. Lots of potential to use in concert with displacement methods (e.g., Lack *et al.*, 2009).

Disadvantages: Do not directly inform about location errors or spatial structure. Wavelet decomposition maintains location information of each field at each scale, but still need to use another method to inform about location and spatial structure errors. For some filters, the scales may not be physically meaningful.

# Displacement Methods: Field deformation

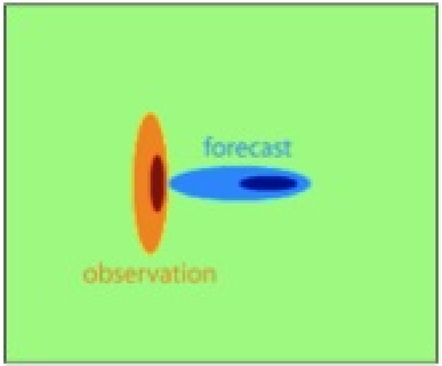
Model: $O(x, y) = F(\mathbf{\Phi}(x, y)) + \varepsilon$

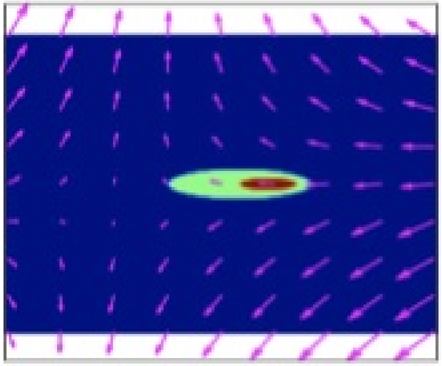
Goal: Inform about how well the forecast capture spatial extent/patterns.

Examples:

- Binary Image Metrics (Gilleland *et al.*, 2008; Zhu *et al.*, submitted to *Wea. Forecasting*)
- Forecast Quality Index (Venugopal *et al.*, 2005)
- Optical Flow (e.g., Keil and Craig, 2008, 2009)
- Image Warping (e.g., Alexander *et al.*, 1998; Gilleland, Lindström and Lindgren, 2010)
- Distortion representation (e.g., Hoffman *et al.*, 1995)
- Gaussian mixtures (Lakshmanan and Kain, 2009)

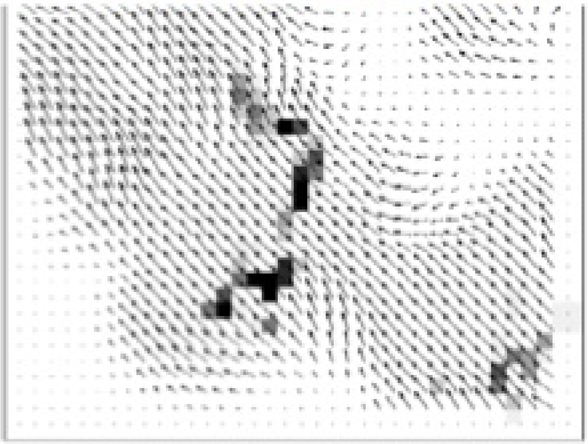
# Displacement Methods: Field deformation



Advantages: Provide information about location errors, and certain structure errors. Vector fields provide diagnostic information. Physically meaningful.

From Keil and Craig 2008



Directly inform about small localized errors and larger-scale errors.

Disadvantages: Do not inform about individual features, only an entire field (or sub-field) at once.

# Displacement Methods: Field Deformation

## Binary Image Metrics

Definition: A metric, $M$, between two sets of pixels $A$ and $B$ contained in a raster of pixels (i.e., domain), $\boldsymbol{\mathcal{X}}$, satisfies:

1. Positivity: $M(A, B) = 0$ if and only if $A = B$,
2. Symmetry: $M(A, B) = M(B, A)$, and
3. Triangle Inequality: $M(A, C) + M(C, B) \geq M(A, B)$.

Analogous definition for a metric between to pixels $\boldsymbol{x}$ and $\boldsymbol{y}$.

# Displacement Methods: Field Deformation

## Binary Image Metrics

- Metrics are useful for comparing two forecasts (against the same observation);
- Perfect score only when forecast is identical to observation (for binary image metrics, in terms of location of events);
- Symmetry ensures that answer does not depend on order of comparison;
- Triangle inequality ensures that results are not overly sensitive $M(O, F_1)$ much better than $M(O, F_2) \implies$ $M(F_1, F_2)$ is appropriately large.
- **Mathematically sound, but does answer make physical sense?**

# Displacement Methods: Field Deformation

## Binary Image Metrics: Hausdorff Metric

Definition: For two binary fields, $F$ and $O$, with sets $A$ and $B$ representing events (e.g., precipitation over a specified threshold) for the forecast and observed fields, resp. (i.e., $A$ and $B$ are everywhere 1, and the field is 0 elsewhere), the Hausdorff metric can be written as:

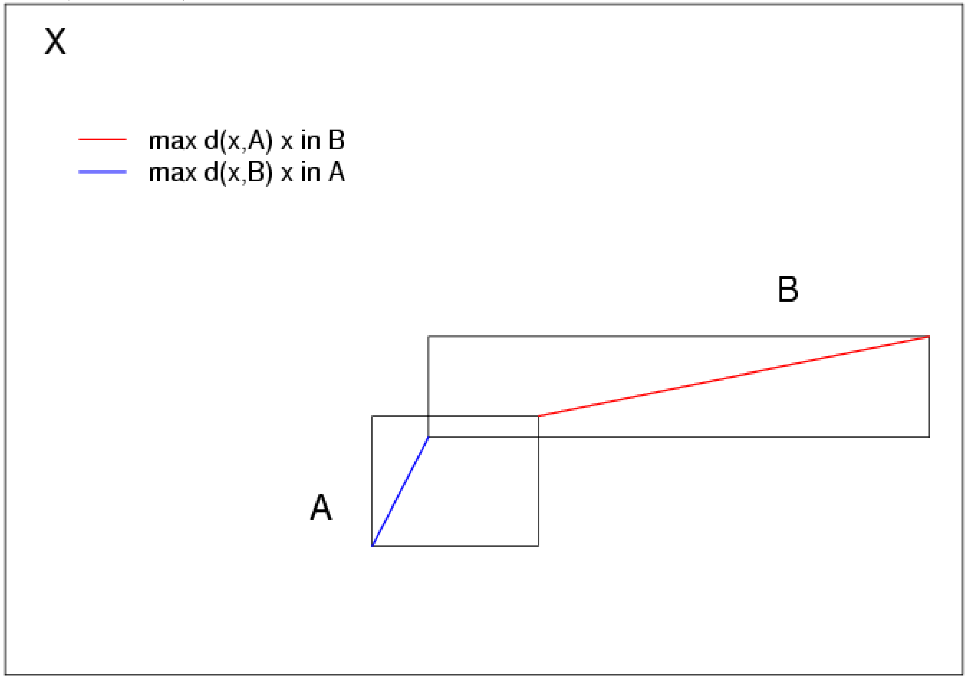
$$H(A, B) = \max_{\boldsymbol{x} \in \mathcal{X}} |d(\boldsymbol{x}, A) - d(\boldsymbol{x}, B)|,$$

where $d(\boldsymbol{x}, A)$ is the shortest distance between the pixel $\boldsymbol{x}$ and the set $A$.

# Displacement Methods: Field Deformation

## Binary Image Metrics: Hausdorff Metric

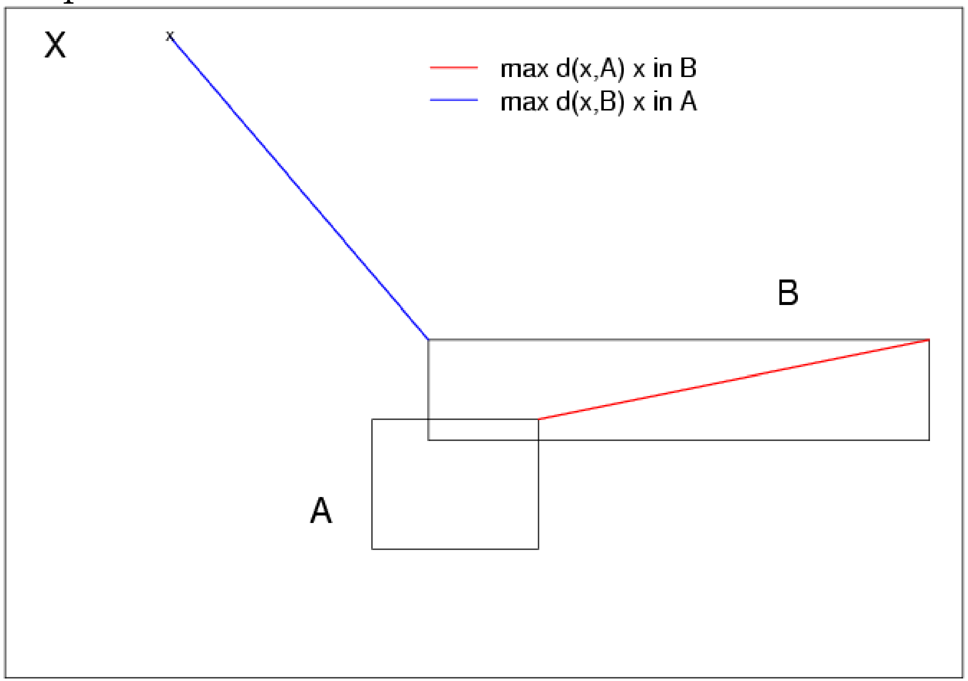
$H(A, B)$ is the length of the red line here.



# Displacement Methods: Field Deformation

## Binary Image Metrics: Hausdorff Metric

$H(A, B)$ has an extreme sensitivity to changes in even a small number of pixels.



# Displacement Methods: Field Deformation

## Binary Image Metrics: Baddeley's $\Delta$ Metric

Replace the maxima in $H(A, B)$ with an $L_p$ norm (and transform $d(\boldsymbol{x}, \cdot)$):

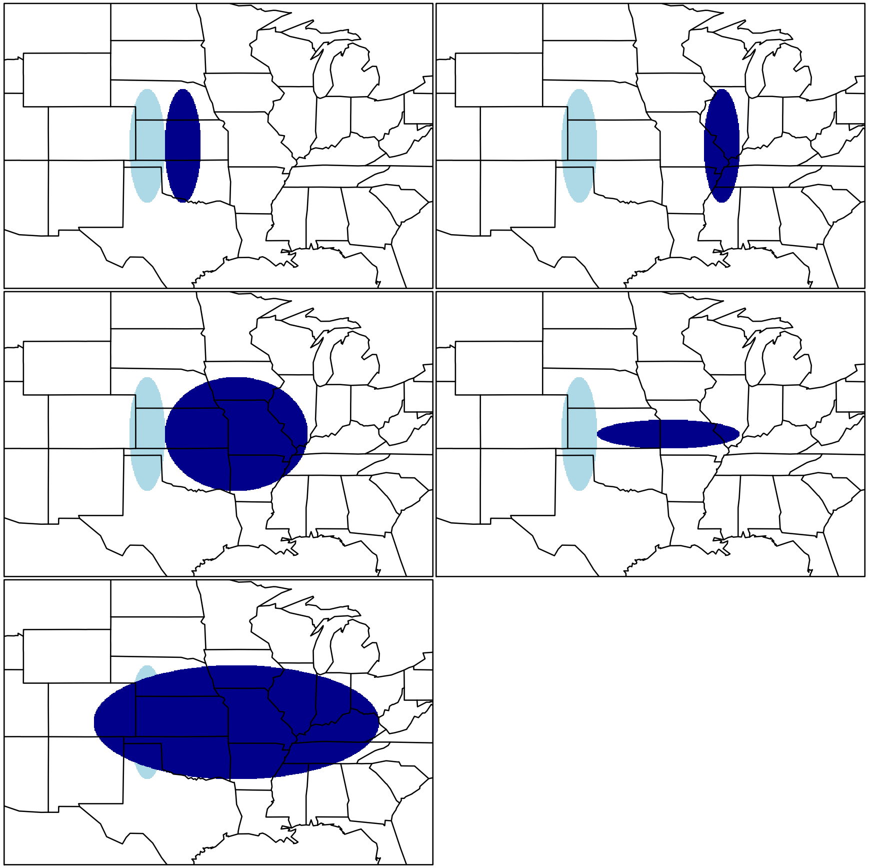
$$\Delta(A, B) = \left[ \frac{1}{n(\mathcal{X})} \sum_{\boldsymbol{x} \in \mathcal{X}} |w(d(\boldsymbol{x}, A)) - w(d(\boldsymbol{x}, B))|^p \right]^{1/p},$$

usually take $p = 2$ and $w(z) = \min\{z, c\}$, $c$ some constant chosen by trial-and-error.

Baddeley, A.J., 1992; *Nieuw Archief voor Wiskunde*, **10**, 157–183.

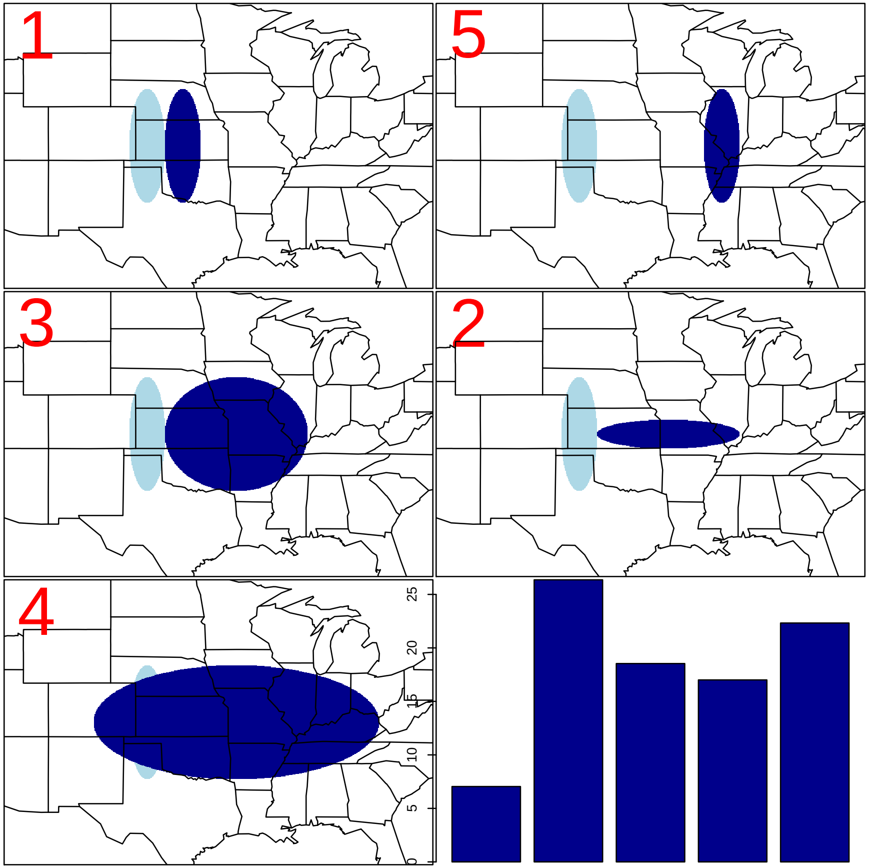
# Displacement Methods: Field Deformation

## Binary Image Metrics: Baddeley's $\Delta$ Metric



# Displacement Methods: Field Deformation

## Binary Image Metrics: Baddeley's $\Delta$ Metric



# Displacement Methods: Field Deformation

## Forecast Quality Index (FQI)

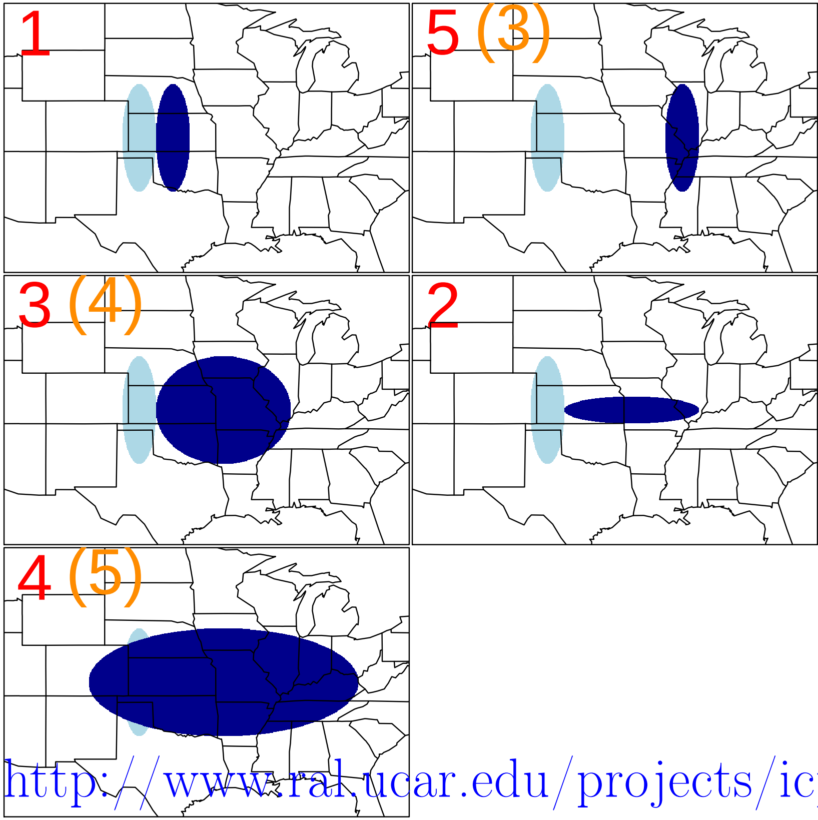
Replace the maximum in $H(A, B)$ with the $k$-th quantile (known as the Partial Hausdorff Distance, PHD), normalize it with surrogate fields (otherwise PHD is not a metric), and also incorporate **intensity** information.

$$\mathrm{FQI}(A, B) = \frac{\dfrac{\mathrm{PHD}_{k(A,B)}}{\frac{1}{m}\sum_{i=1}^{m}\mathrm{PHD}_k\left(A, \hat{A}_i\right)}}{\frac{2\mu_A\mu_B}{\mu_A^2+\mu_B^2}\frac{2\sigma_A\sigma_B}{\sigma_A^2+\sigma_B^2}},$$

where $\hat{A}_i$ is the $i$-th surrogate field for field $A$, $\mu$ and $\sigma$ represent the means and standard deviations (of **intensity**).

# Displacement Methods: Field Deformation

## Forecast Quality Index (FQI)




Basu and Foufoula-Georgiou, 2008. Presentation, available at

http://www.ral.ucar.edu/projects/icp/Results/FeaturesBased/FQI/ICP-FQI-Results.pdf

# Displacement Methods: Field Deformation

## Forecast Quality Index (FQI)

Perturbed cases (using a threshold of 1 mm):

| Case (Description) | FQI | Rank |
|---|---|---|
| pert001 (3 points right and 5 points down) | 0.03 | 1 |
| pert002 (6 points right and 10 points down) | 0.07 | 2 |
| pert003 (12 points right and 20 points down) | 0.13 | 3 (tie) |
| pert004 (24 points right and 40 points down) | 0.24 | 6 |
| pert005 (48 points right and 80 points down) | 0.64 | 7 |
| pert006 (12 points right, 20 points down, and $\times 1.5$) | 0.13 | 3 (tie) |
| pert007 (12 points right, 20 points down, and $-0.05$ in.) | 0.23 | 5 |

http://www.ral.ucar.edu/projects/icp/Results/FeaturesBased/FQI/ICP-FQI-Results.pdf

(Basu and Foufoula-Georgiou, 2008 presentation)

# Displacement Methods: Field Deformation

## Field Deformation Methods: Optical Flow

Keil and Craig (2009) use a pyramidal matching algorithm to find the optimal vector field describing the deformation at successively finer scales within a fixed search environment (seek to minimize an amplitude-based quantity).

**More from Caren on this approach**

# Field Deformation Methods: Image Warping

$$O(x, y) = F(W_x(x, y), W_y(x, y)) + \varepsilon$$

- $W$ is a warping function that acts on both coordinates $x$ and $y$ of an image, and is applied to both coordinates;
- Many choices for $W$, e.g.,
  - polynomials (e.g., Alexander *et al.*, 1999; Dickinson and Brown, 1996)
  - B-splines (e.g., Engel *in prep*?)
  - Thin-plate splines (e.g., Gilleland, Lindström and Lindgren, 2010)
- Find optimal warp by optimizing a likelihood function.

# Displacement Methods: Field Deformation

## Field Deformation Methods: Image Warping

Can warp all pixels in an image, but usually choose a subset (control points). Entire deformation is determined by these points, but is applied to all points. Optimize (log) likelihood:

$$\ell(\boldsymbol{p}^F | O, F, \boldsymbol{p}^O) = \log \mathrm{p}(O | F, \boldsymbol{p}^F, \boldsymbol{p}^O) + \log \mathrm{p}(\boldsymbol{p}^F | \boldsymbol{p}^O)$$

# Displacement Methods: Features based

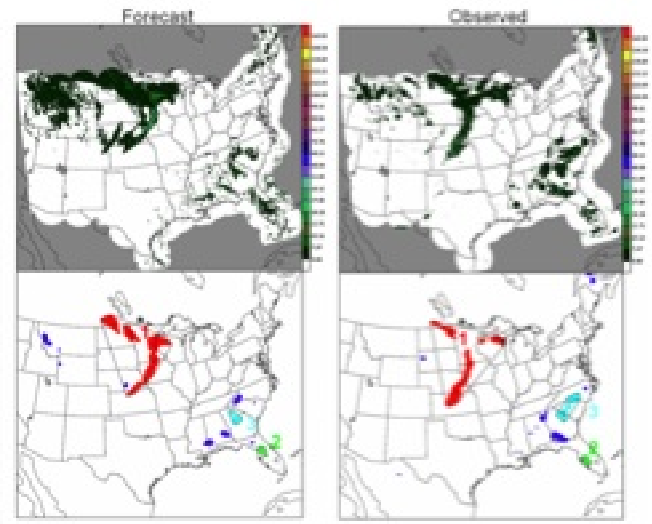
Model: $O_A(x, y) = F_B(x, y) + \varepsilon$

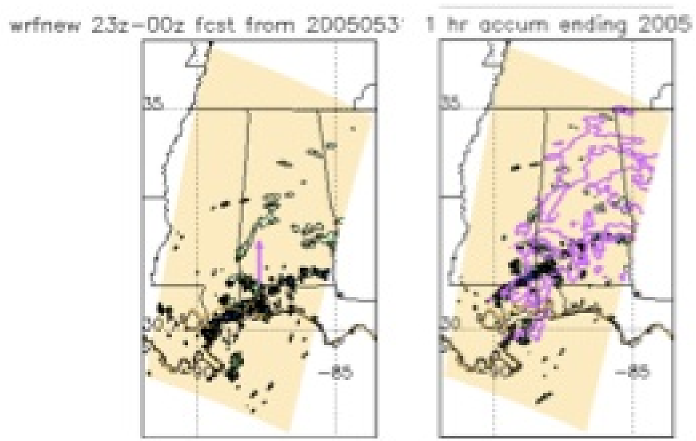
Goal: Measure and compare user-relevant features in the forecast and observed fields.

Examples:

- CRA (e.g., Ebert and McBride, 2000; Ebert and Gallus, 2009)
- MODE (e.g., Davis *et al.*, 2006, 2009)
- Procrustes (Lack *et al.*, 2009)
- Cluster Analysis (e.g., Marzban and Sandgathe, 2006; Marzban *et al.*, 2008, 2009)
- SAL (e.g., Wernli *et al.*, 2008, 2009)
- Composite (e.g., Nachamkin, 2006, 2009)

# Displacement Methods: Features based






MODE example 2008 CRA: Ebert and Gallus 2009

Advantages: Provide information about location errors, and certain structure errors. Vector fields provide diagnostic information. Physically meaningful. Directly inform about small localized errors and larger-scale errors. Informs on individual features. Identify hits, misses and false alarms.

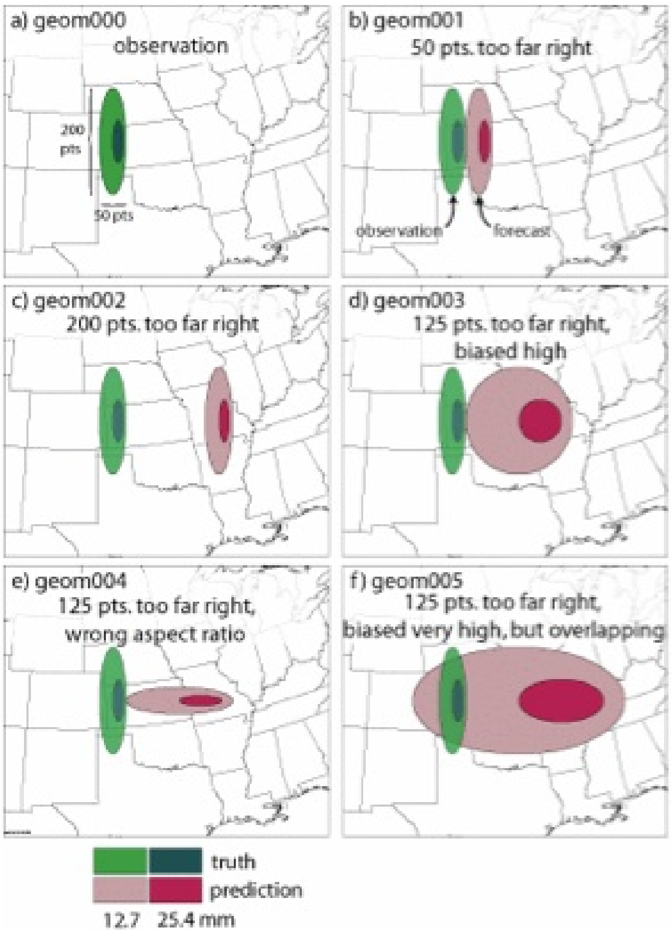
Disadvantages: Often need to merge and match *features*, which can be tricky.

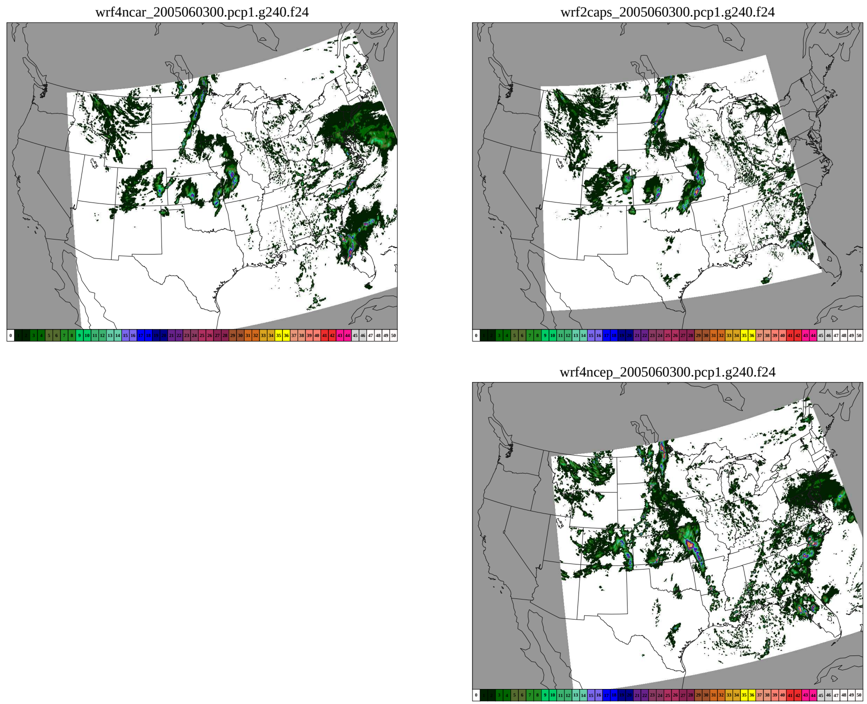
# Summary

| **Category** | **Scales with skill** | **Location errors** | **Intensity errors** | **Structure errors** | **Occurrence (hits, misses and false alarms)** |
|---|---|---|---|---|---|
| Neighborhood | Yes | No | Yes | No | Yes |
| Scale-separation | Yes | No | Yes | No | Yes |
| Features-based | No | Yes | Yes | Yes | Yes |
| Deformation | No | Yes | Yes | No | No |

# Spatial Forecast Verification Methods Inter-Comparison Project (ICP)

http://www.ral.ucar.edu/projects/icp



# Final Remarks

http://www.ral.ucar.edu/projects/icp

- See ICP web page under *References* and *Special Collection* for full references from these slides.
- Special collection of *Weather and Forecasting* for ICP.
- So far, *geometric*, *perturbed* and *real* test cases have focused on QPF fields over the central and eastern United States. Need to look at other regions and other field types (e.g., wind, pressure, etc.).
- Participation in the ICP is encouraged. Sign up to receive emails at the web site.
- Expand ICP to other verification issues (e.g., ensembles, spatial-temporal)?