Spatial Verification Methods

Barbara Brown (bgb@ucar.edu)

National Center for Atmospheric Research (NCAR)
Boulder, Colorado

Collaborators: Randy Bullock, John Halley Gotway, David Ahijevych, Eric Gilleland, Beth Ebert, Barbara Casati

4th International Verification Methods Workshop
Helsinki, Finland
June 2009

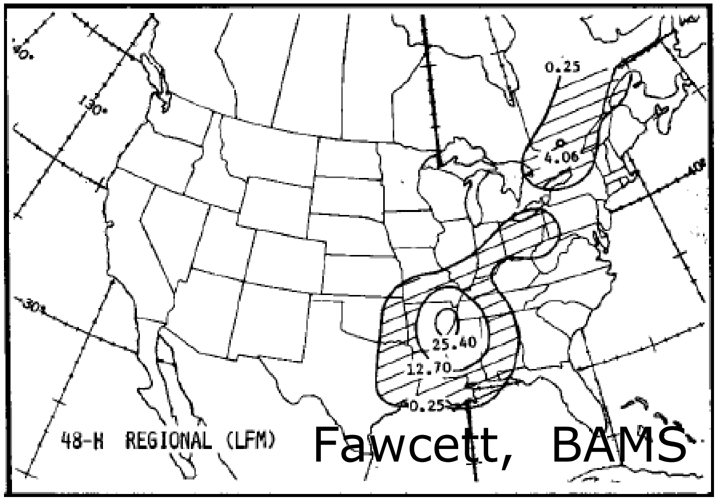
Goals

- Briefly describe new approaches for evaluation of spatial (gridded) forecasts that
  - Provide meaningful information about performance
  - Overcome some of the insensitivity of traditional approaches
- Present results from the spatial method intercomparison project (ICP)

Challenge of High Resolution

Examples of 12-h accumulated precip

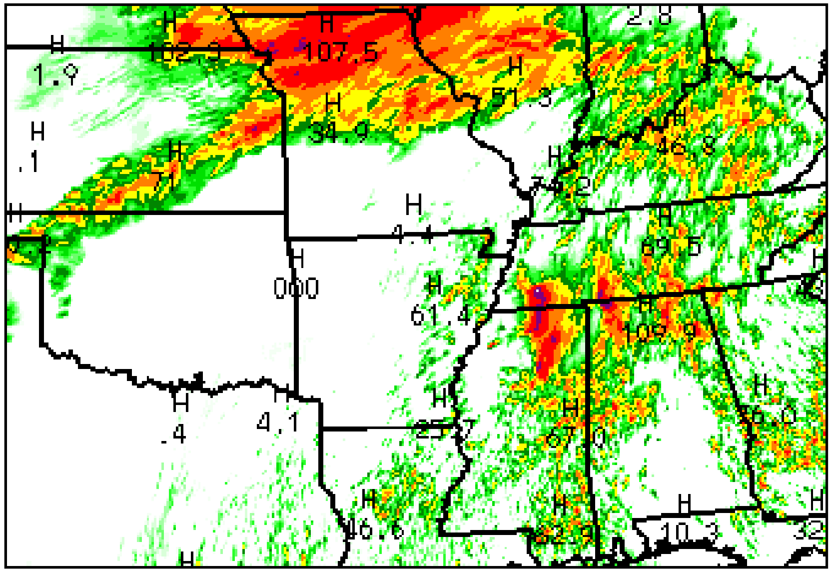
THEN



Fawcett, BAMS

190-km LFM, 1977

NOW

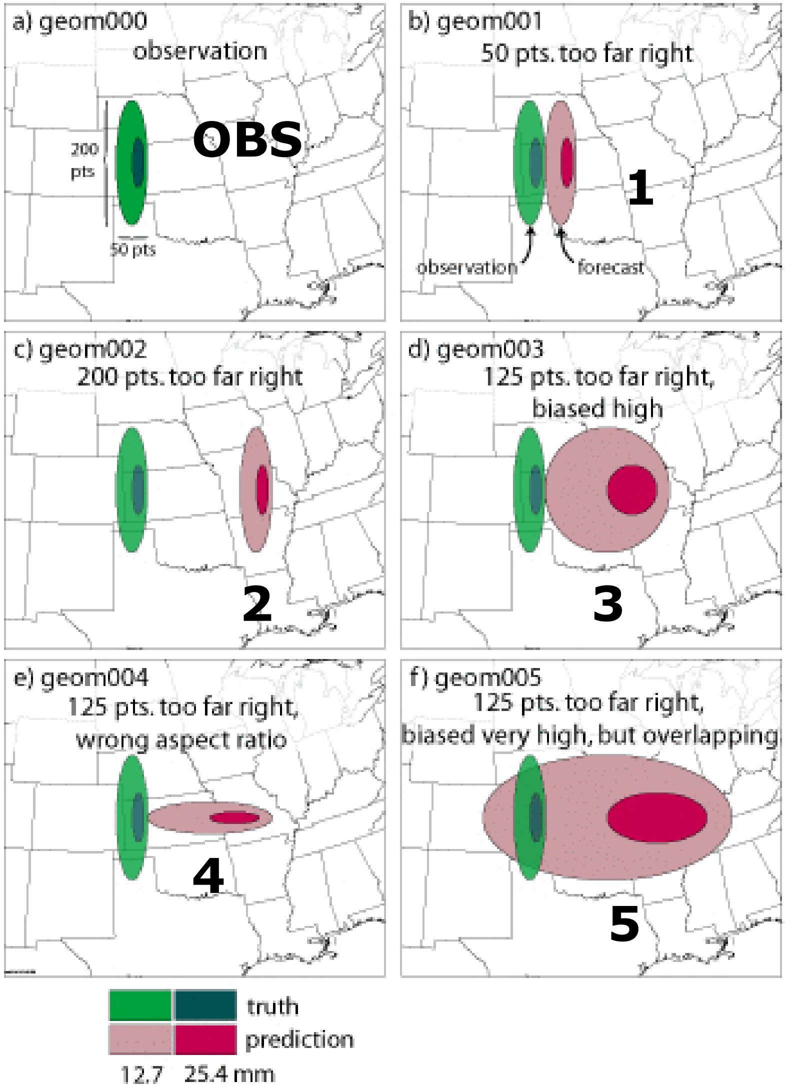


3-km WRF, 2009

Traditional approach

Consider gridded forecasts and observations of precipitation

Which is better?



Traditional approach

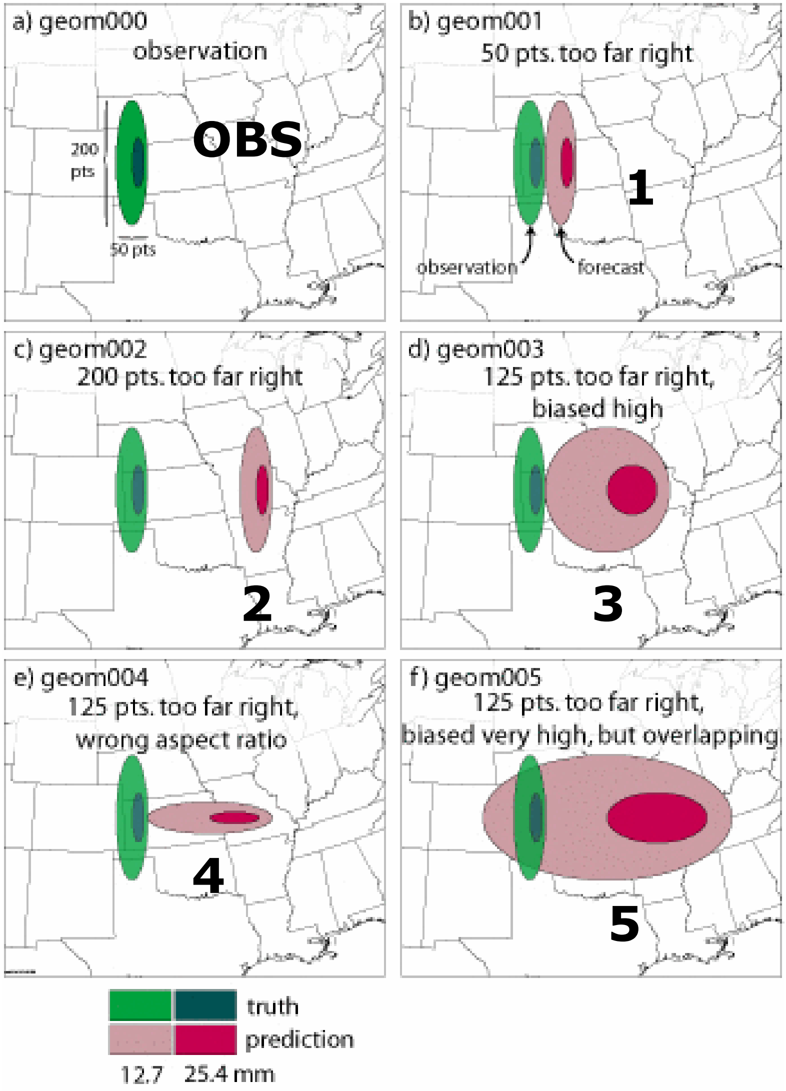
Scores for Examples 1-4:

Correlation Coefficient = -0.02
Probability of Detection = 0.00
False Alarm Ratio = 1.00
Hanssen-Kuipers = -0.03
Gilbert Skill Score (ETS) = -0.01

Scores for Example 5:

Correlation Coefficient = 0.2
Probability of Detection = 0.88
False Alarm Ratio = 0.89
Hanssen-Kuipers = 0.69
Gilbert Skill Score (ETS) = 0.08

Forecast 5 is "Best"

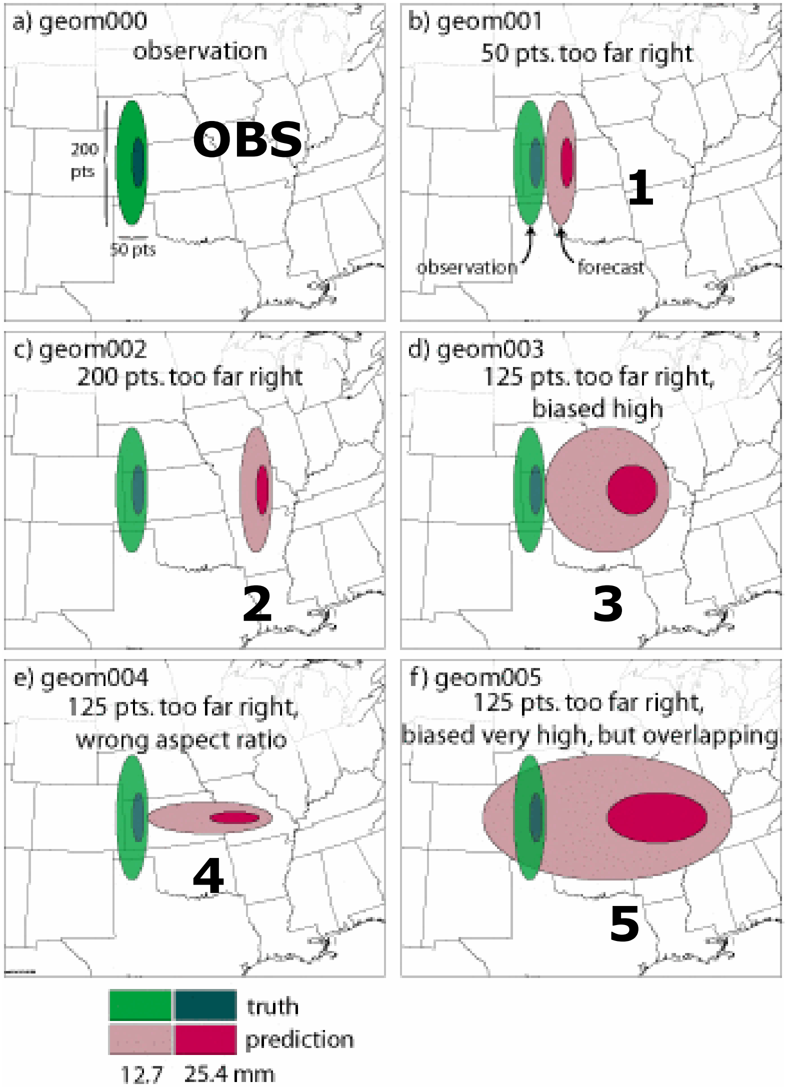


Traditional approach

Some problems with the traditional approach:

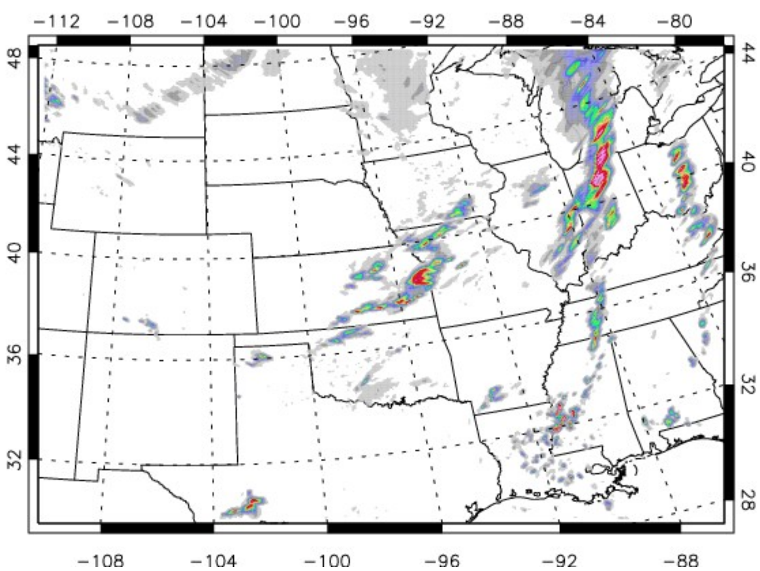
(1) ***Non-diagnostic*** *– doesn't tell us <u>what</u> was wrong with the forecast – or what was right*

(2) ***Utra-sensitive*** *to small errors in simulation of localized phenomena*



Spatial forecasts

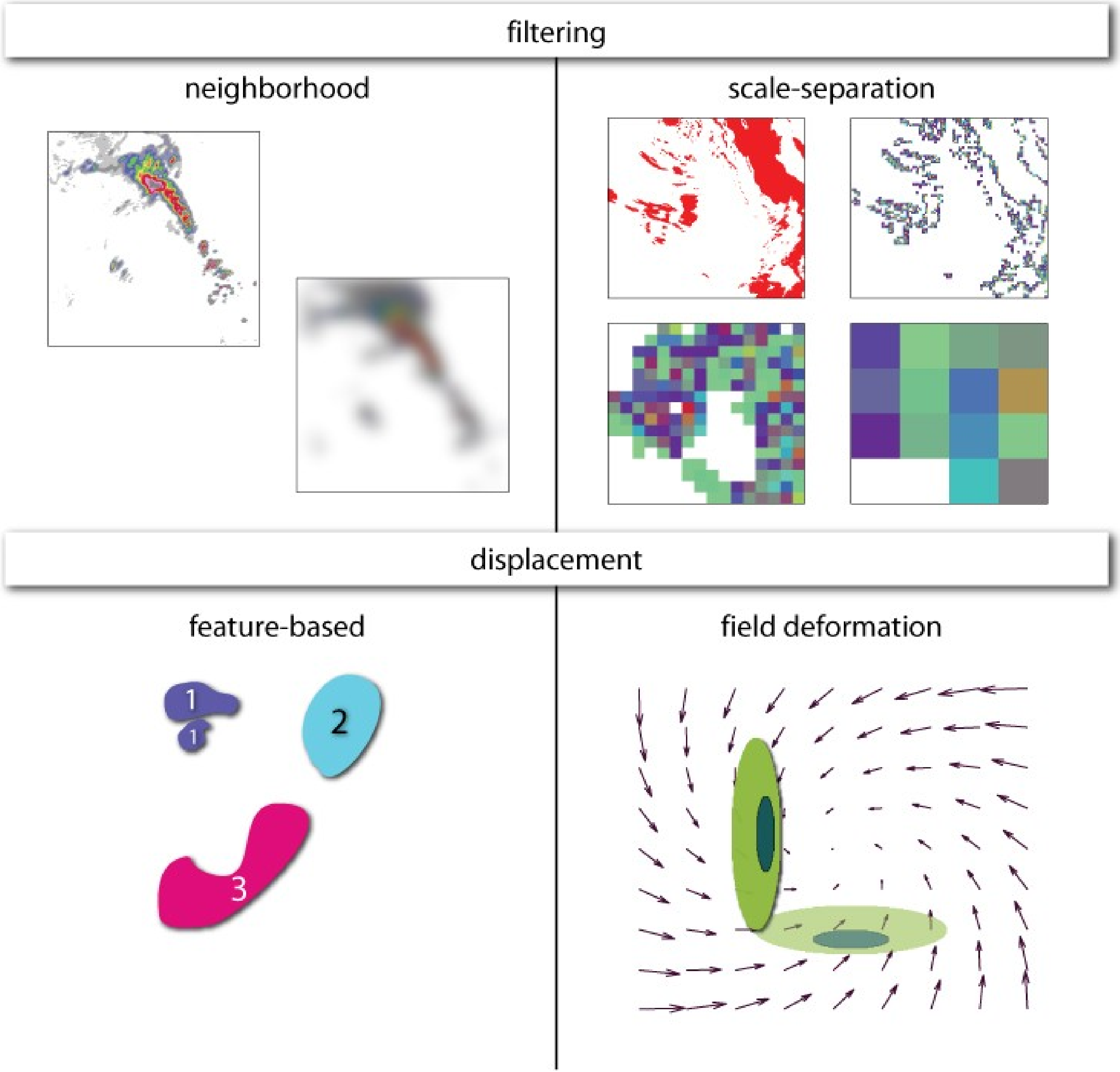
Weather variables (e.g., precipitation) defined over spatial domains have **coherent structure and features**



Spatial methods aim to:

- Account for uncertainties in timing and location
- Account for spatial structure
- Provide information on error in physical terms
- Provide information that is
  - Diagnostic
  - Meaningful to forecast users

Spatial Method Categories



New spatial verification approaches

Neighborhood

Give credit to "close" forecasts

Scale separation

Measure scale-dependent error

Field deformation

Measure distortion and

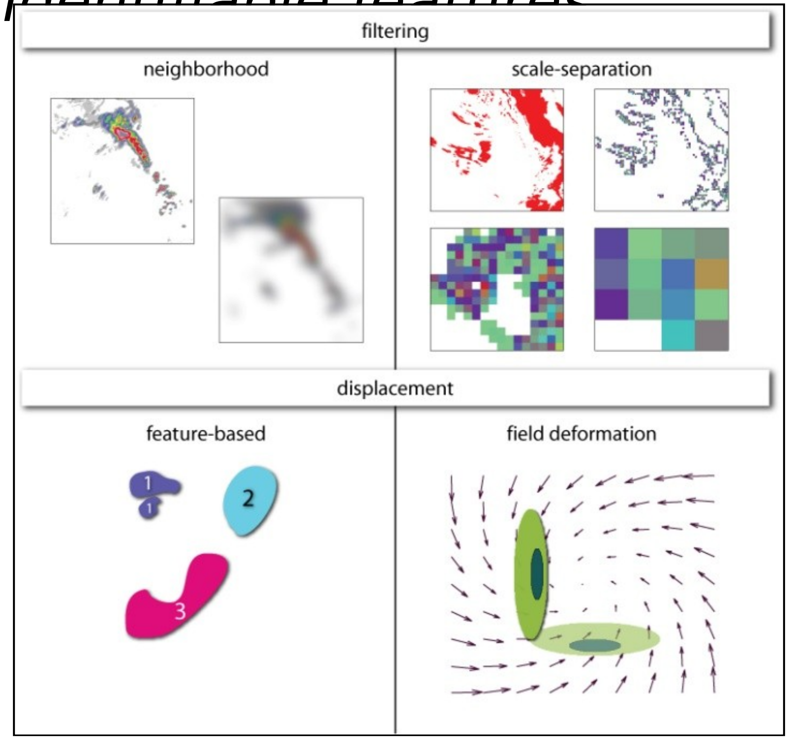
displacement (phase error) for

whole field

How should the forecast be adjusted to make the best match

Object- and feature-based

Evaluate attributes of identifiable features



Method Intercomparison Project (ICP)

Goals:

- Compare information provided by various newly proposed spatial verification methods
- Investigate strengths and weaknesses

Activities:

- Workshops (2007,2008)

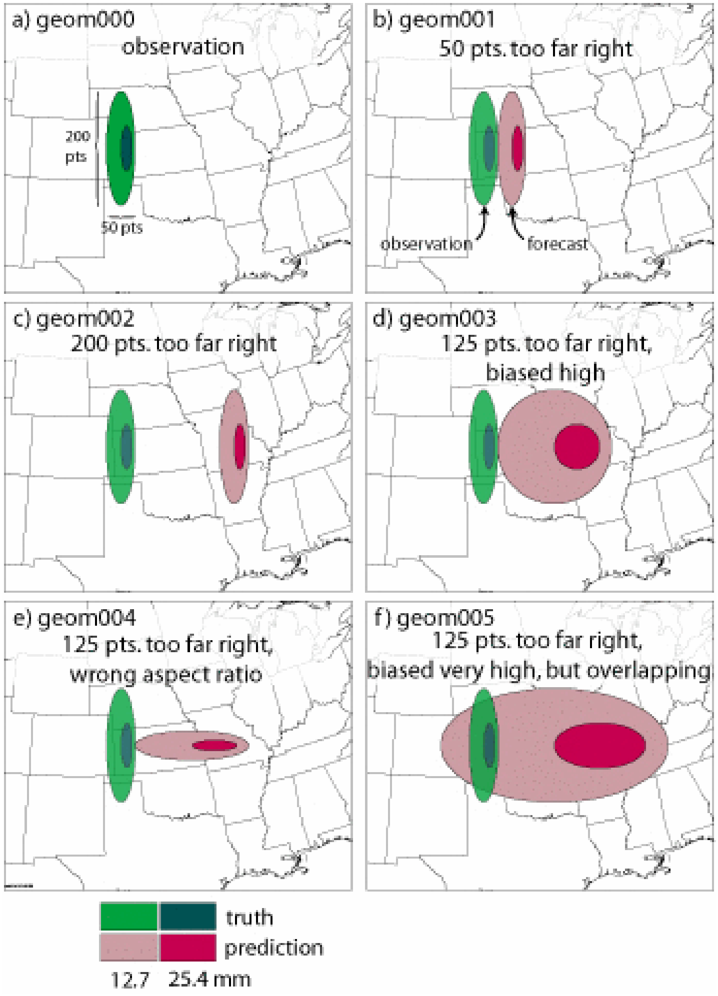
Web site: http://www.ral.ucar.edu/projects/icp/

Datasets:

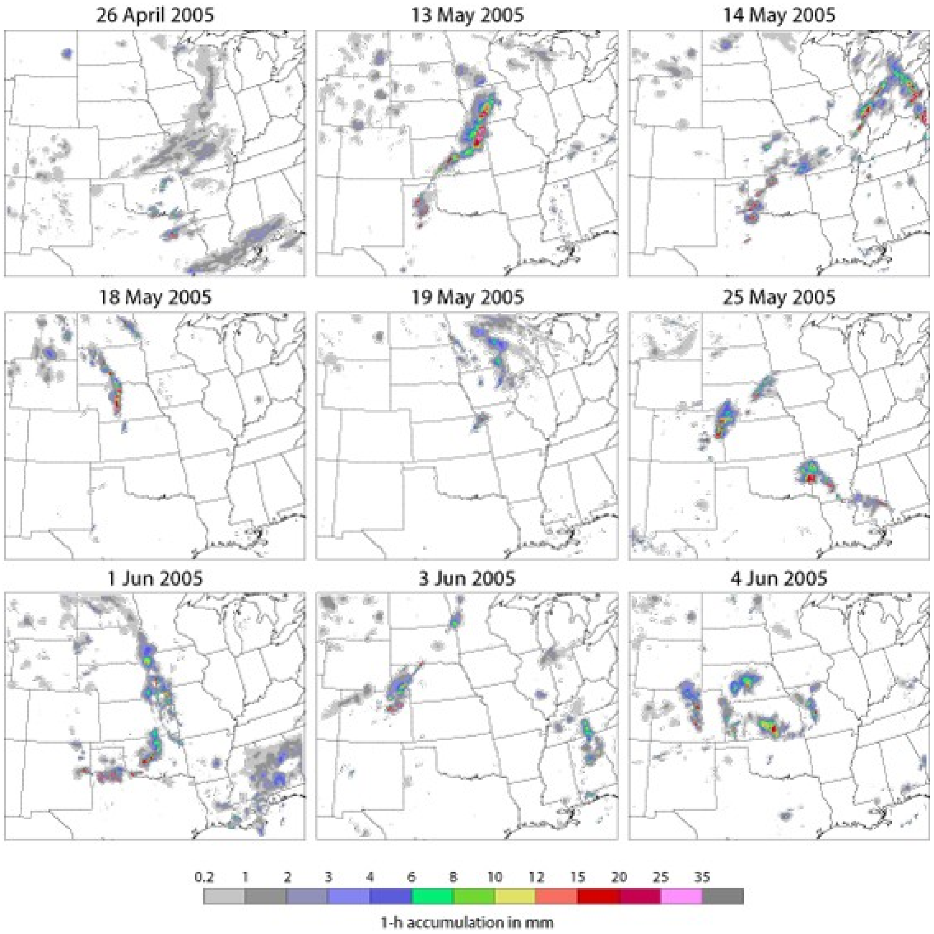
- Geometric cases
- Actual precipitation forecasts and obs
  - WRF precipitation forecasts (4 km)
  - Stage IV precipitation analysis
  - Resolution: 4 km
  - Domain: Central U.S.
  - Time period: May-Jun 2005 (9 focus cases)
- Perturbed cases

Cases used in the ICP

Geometric cases



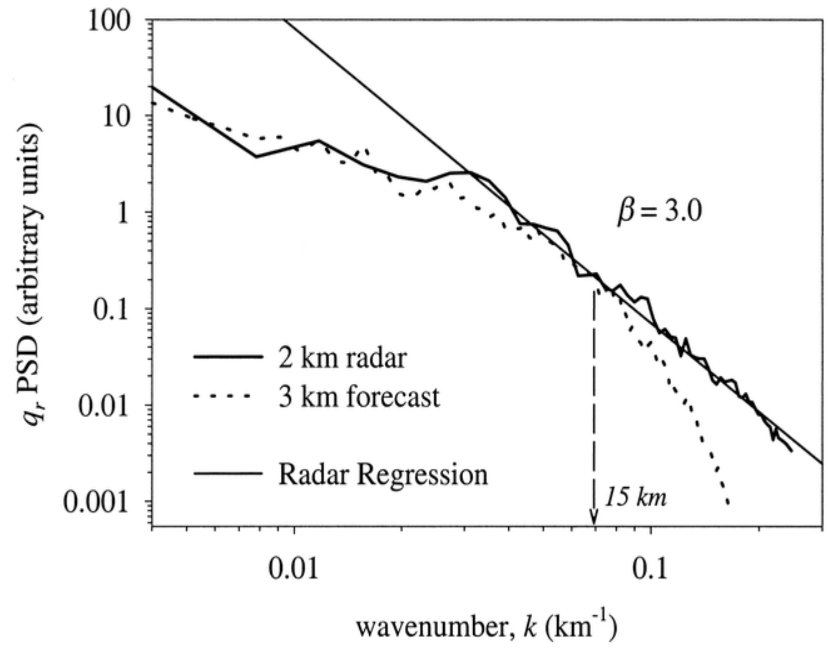
"Real" cases



Scale separation methods

- Goal:
  Examine performance as a function of spatial scale

- Example: Power spectra
  - Does it look real?
  - Harris et al. (2001) compared multi-scale statistics for model and radar data

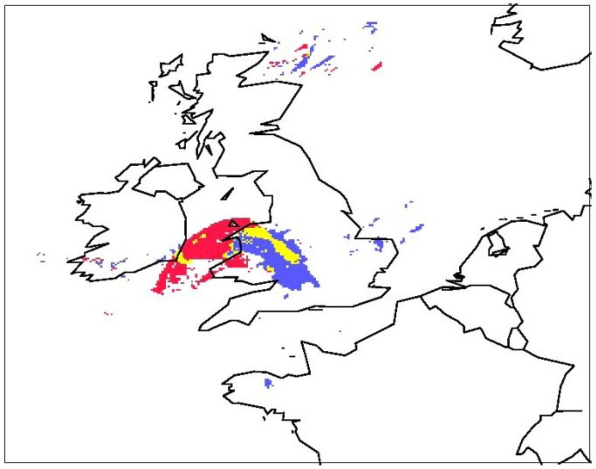


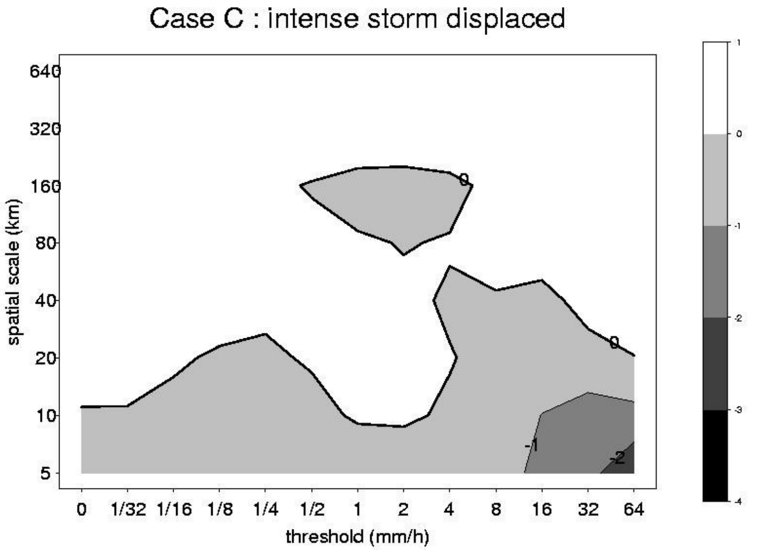
From Harris et al. 2001

Scale separation methods

Example methods :

- Intensity-scale (Casati et al. 2004)
- Multi-scale variability (Zapeda-Arce *et al.* 2000; Harris *et al.* 2001; Mittermaier 2006)
- Variogram (Marzban and Sandgathe 2009)



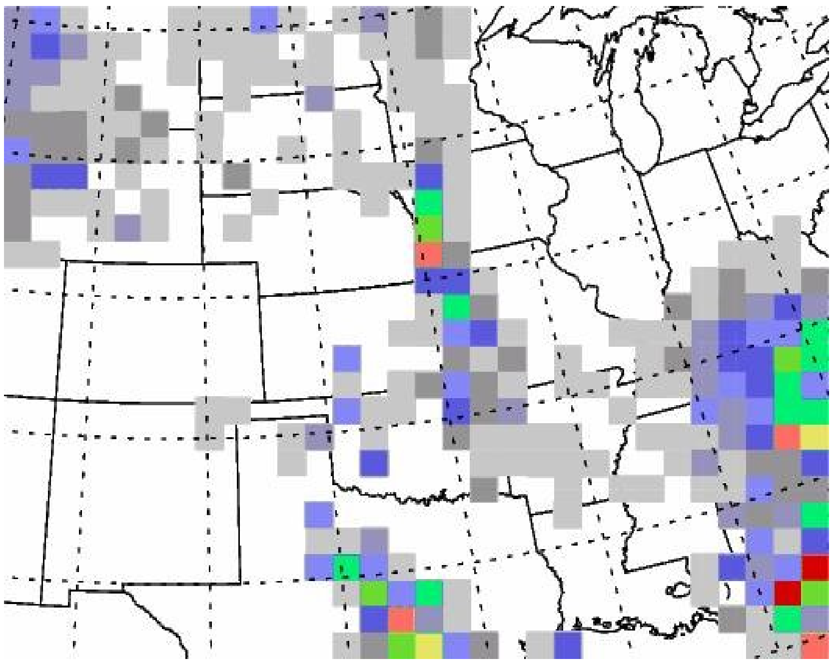


Neighborhood verification

Goal:

Examine forecast performance in a region; don't require exact matched

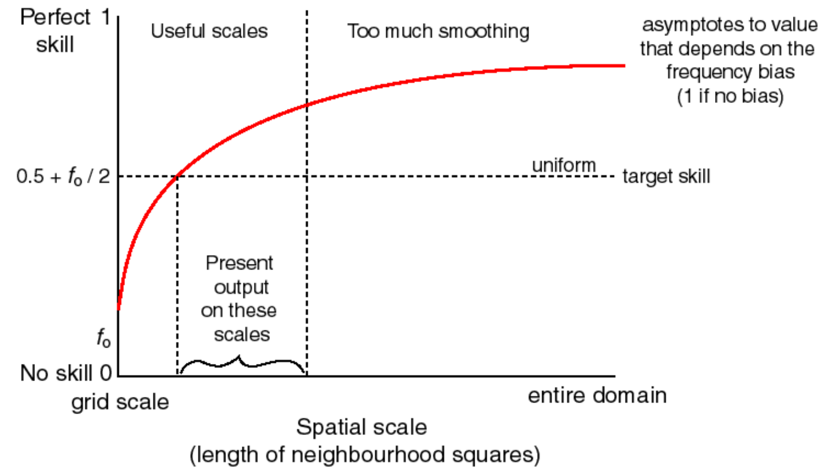
- Also called "fuzzy" verification
- Example: Upscaling
  - Put observations and/or forecast on coarser grid
  - Calculate traditional metrics
- Provide information about scales where the forecasts have skill

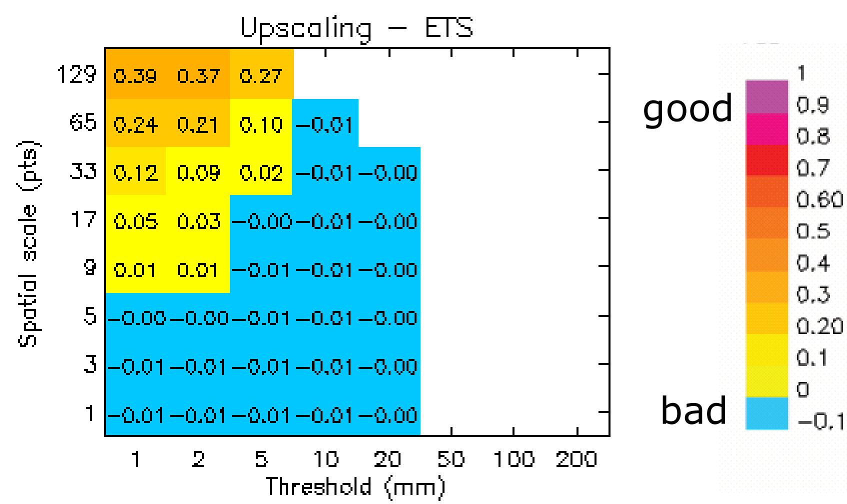


Neighborhood methods

Example methods :

- Distribution approach (Marsigli)
- Fractions Skill Score (Roberts 2005; Roberts and Lean 2008; Mittermaier and Roberts 2009)
- Multiple approaches (Ebert 2008, 2009) (e.g., Upscaling, Multi-event cont. table, Practically perfect)

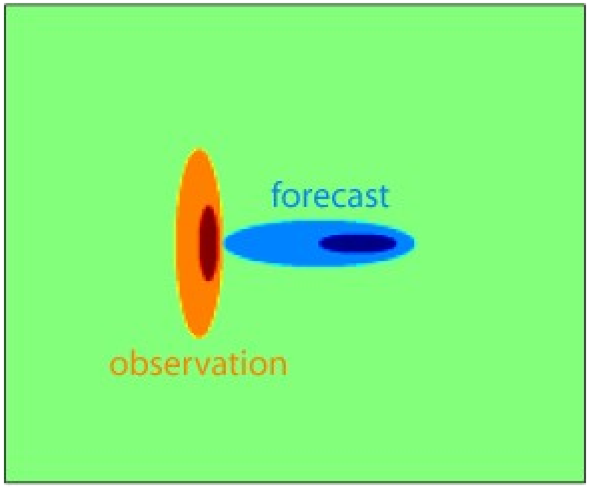


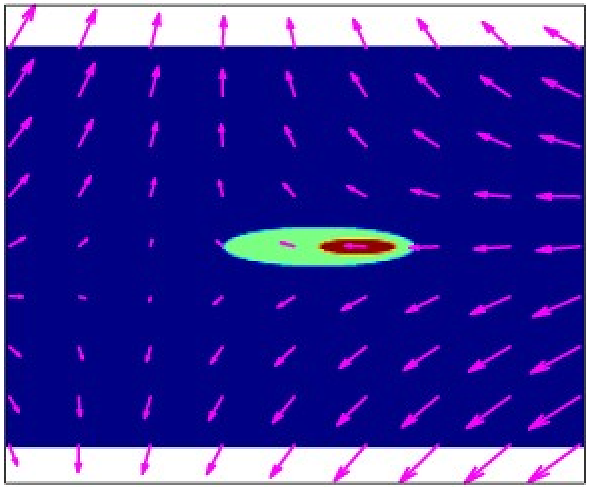


Field deformation

Goal:

Examine how much a forecast field needs to be transformed in order to match the observed field

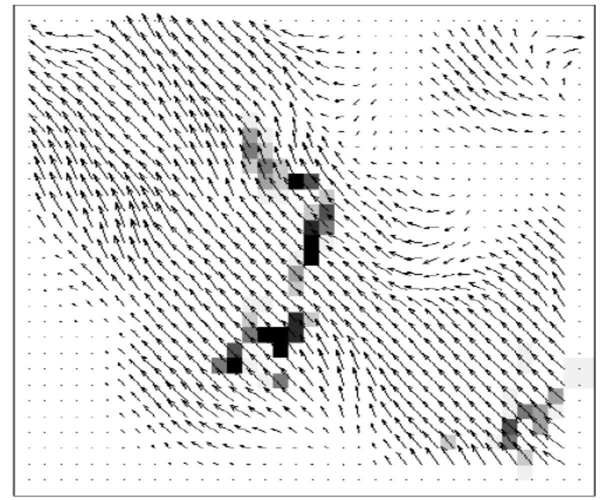




Field deformation methods

Example methods :

- Forecast Quality Index (Venugopal *et al*. 2005)
- Forecast Quality Measure/Displacement Amplitude Score (Keil and Craig 2007, 2009)
- Image Warping (Gilleland et al. 2009; Lindström *et*

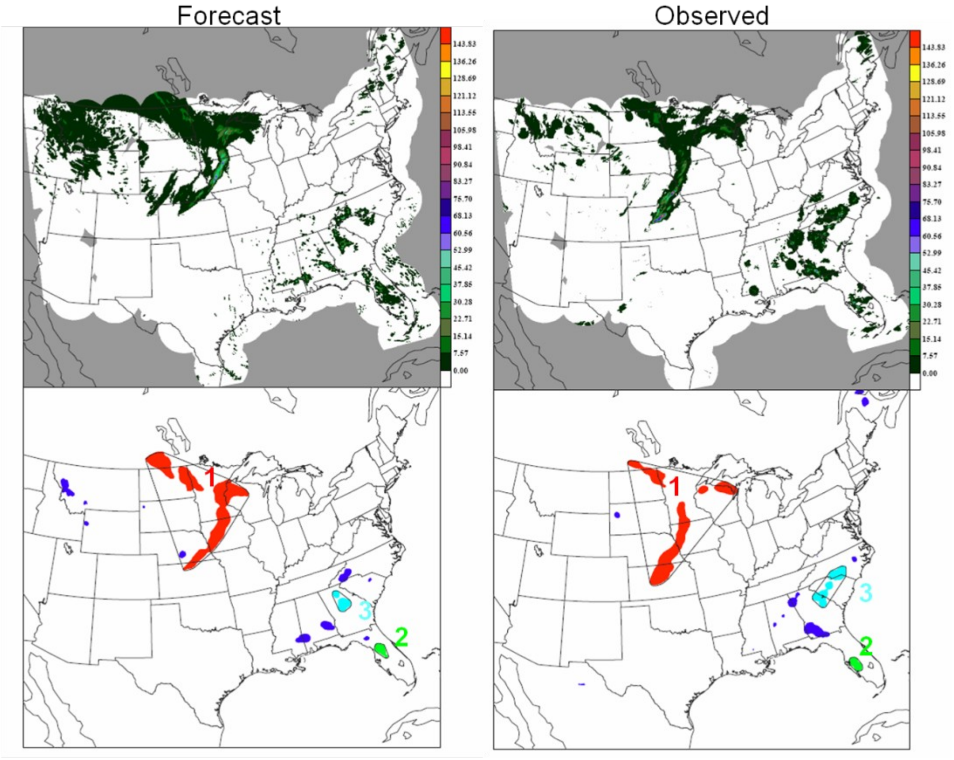


From Keil and Craig 2008

Object/Feature-based

Goals:

2. Identify relevant features in the forecast and observed fields
3. Compare attributes of the forecast and observed features



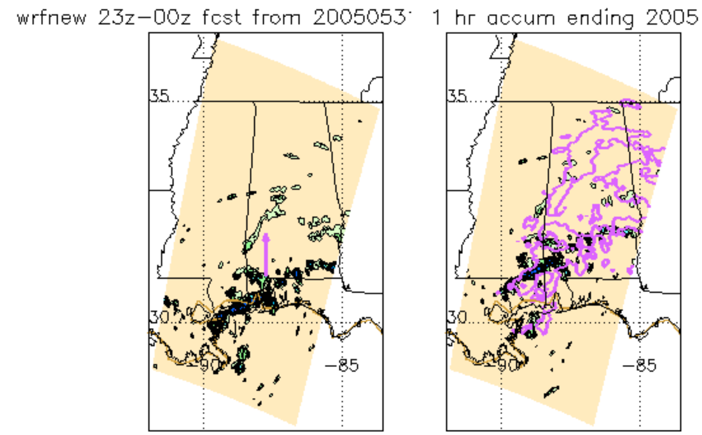
MODE example 2008

Object/Feature-based

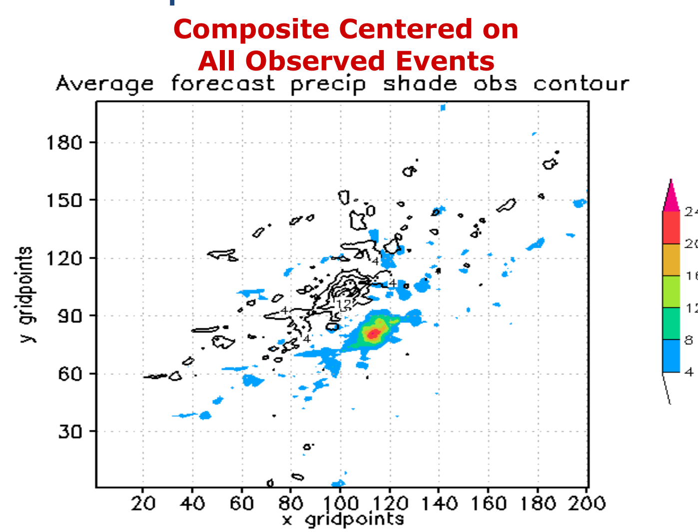
Example methods:

- Cluster analysis (Marzban and Sandgathe 2006a,b)
- Composite (Nachamkin 2005, 2009)
- Contiguous Rain Area (CRA) (Ebert and McBride 2000; Ebert and Gallus 2009)
- MODE (Davis *et al*. 2006, 2009)
- Procrustes (Micheas *et al*. 2007, Lack *et al*. 2009)

CRA example (Ebert and Gallus)

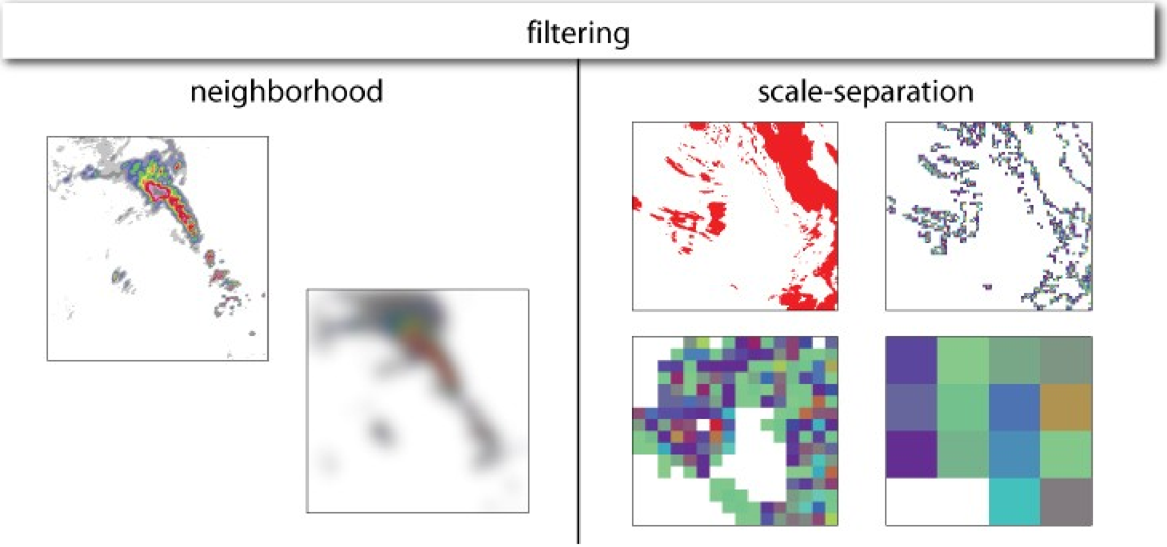


Composite: Nachamkin



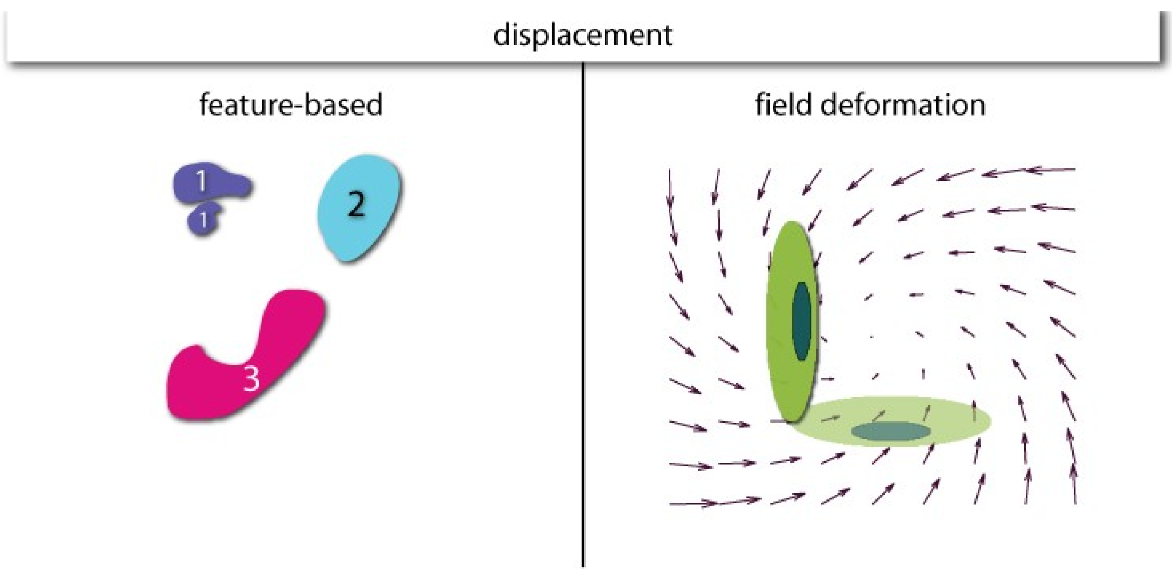
Limitations: Filtering (Neighborhood and Scale separation)

Does not clearly isolate specific errors (e.g., displacement, amplitude, structure)



Limitations: Displacement methods (features-based, field deformation)

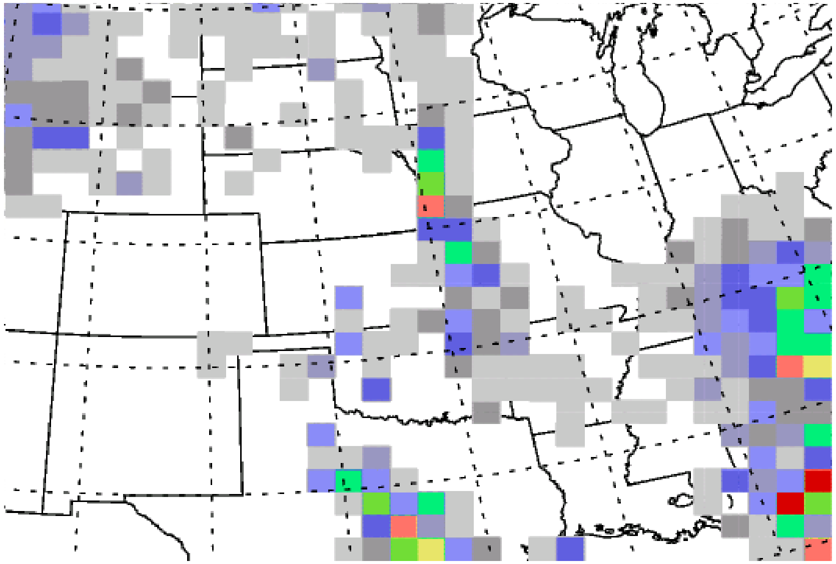
- May have somewhat arbitrary matching criteria
- Often many parameters to be defined
- More research needed on diagnosing mesoscale structure



Strengths – Filtering

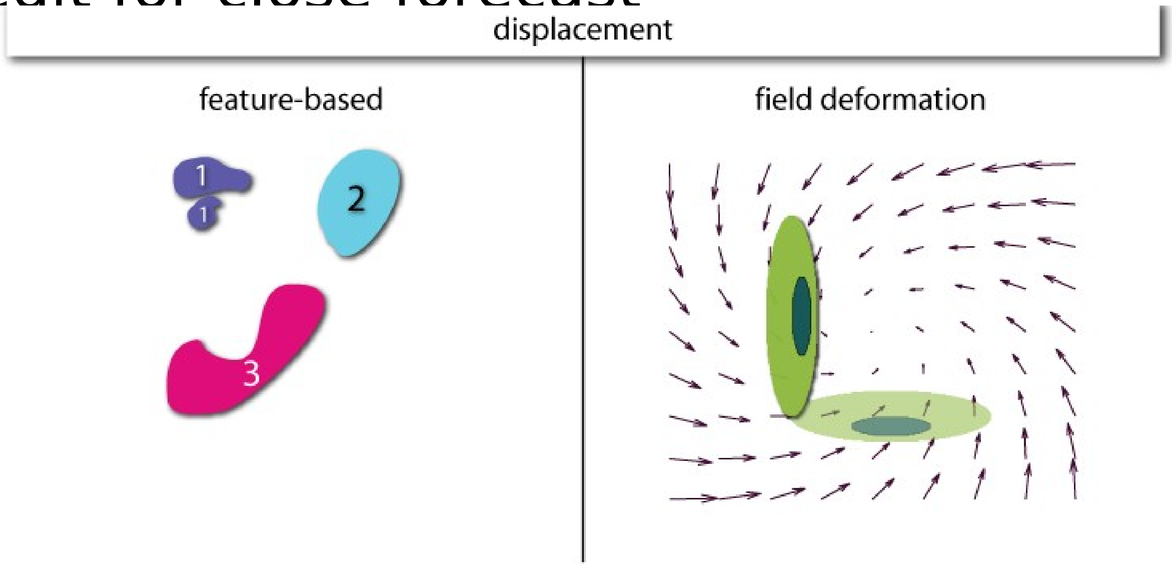
Neighborhood & Scale-Separation

- Accounts for
  - Unpredictable scales
  - Uncertainty in observations
- Simple – ready-to-go
- Evaluates different aspects of a forecast (e.g., texture)
- Scale dependent



Strengths - Displacement

- Features-based
  - credit for close forecast
  - measures displacement, structure
- Field-deformation
  - Distinguish aspect ratio and orientation angle error
  - credit for close forecast



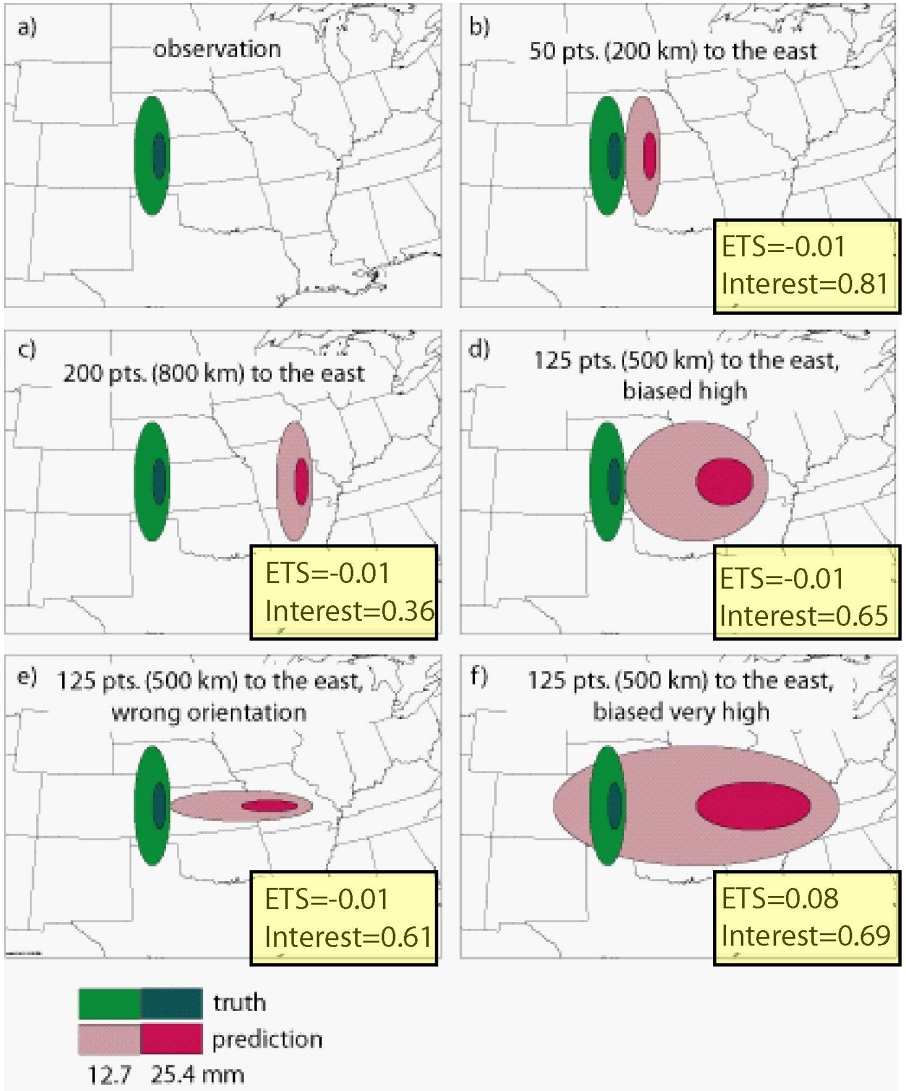
What do the new methods measure?

Attribute	*Traditional*	*Feature-based*	*Neighbor-hood*	*Scale*	*Field Defor-mation*
Perf at different scales	Indirectly	Indirectly	**Yes**	**Yes**	No
Location errors	No	**Yes**	Indirectly	Indirectly	**Yes**
Intensity errors	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Structure errors	No	**Yes**	No	No	**Yes**
Hits, etc.	**Yes**	**Yes**	**Yes**	Indirectly	**Yes**

Back to the original example…
What can the new methods tell us?

Example:

- MODE "Interest" measures overall ability of forecasts to match obs
- Interest values provide more intuitive estimates of performance than the traditional measure (ETS)
- But note: Even for spatial methods, Single measures don't tell the whole story!



Conclusion

- New spatial methods provide great opportunities for more meaningful evaluation of precipitation forecasts
  - Feed back into forecast development
  - Provide information to users
- Each method is useful for particular types of situations and for answering particular types of questions
- ICP may represent a model for future "Verification Testbed" activities

For more information, see http://www.rap.ucar.edu/projects/icp/index.html

Spatial Verification ICP summary

- Weather and Forecasting special issue
  - Literature review – Gilleland et al.
  - Results overview – Ahijevych et al.
- Individual Methods
  - Casati; Davis et al.; Ebert; Ebert and Gallus; Gilleland et al.; Keil and Craig; Lack et al.; Lindstrom et al.; Marzban and Sandgathe; Marzban et al.; Mittermaier and Roberts; Nachamkin; Wernli et al.