Spatial Forecast Verification

Eric Gilleland
Research Applications Laboratory,
National Center for Atmospheric Research,
Boulder, CO, USA
(EricG @ ucar.edu)

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

9th EMS Annual Meeting
28 September – 2 October 2009
Toulouse, France

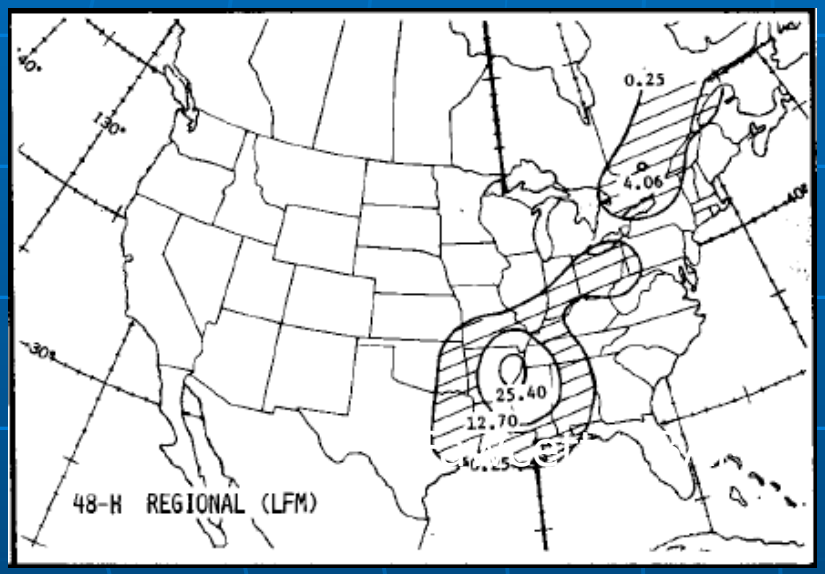
JWGFVR

NCAR

Challenge of High Resolution

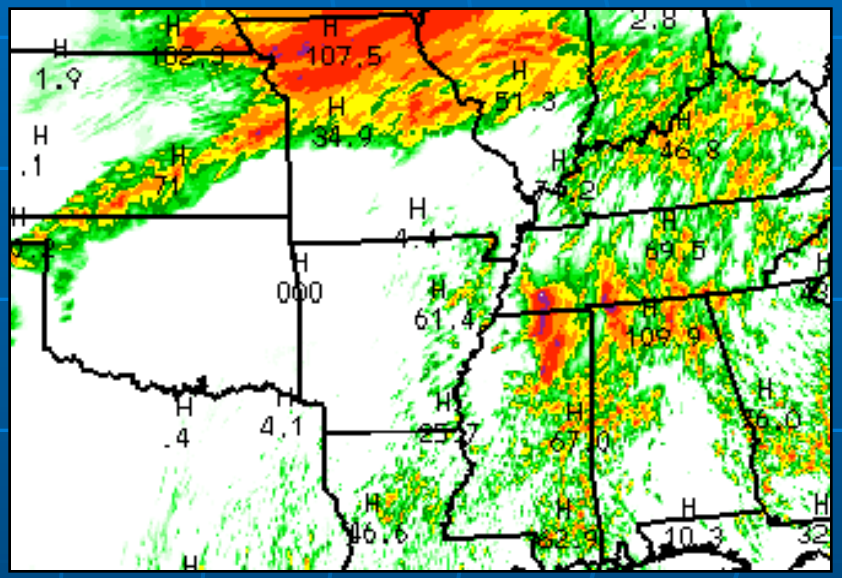
Examples of 12-h accumulated precipitation

THEN



190-km LFM, 1977

NOW



3-km WRF, 2009

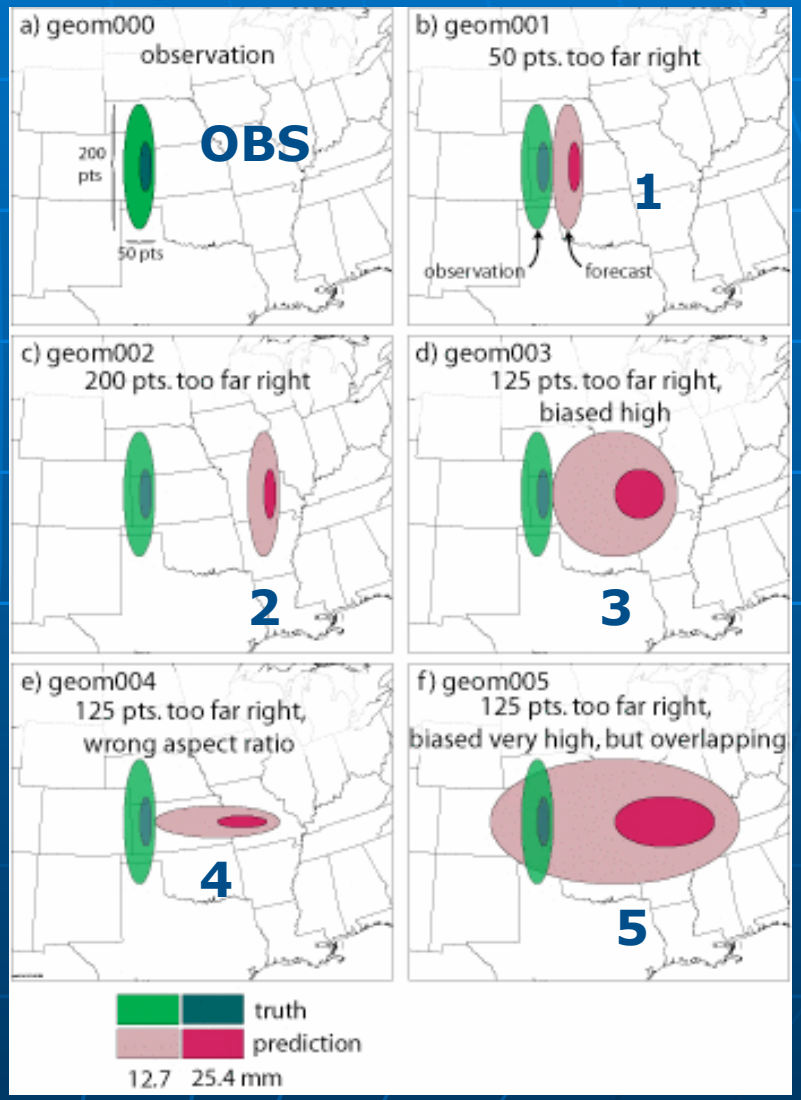
From Ebert 2009

Consider gridded forecasts and observations of precipitation…

Traditional approach: "What is the skill score?"

Based on comparing overlapping grid points

Which is better?



Traditional approach

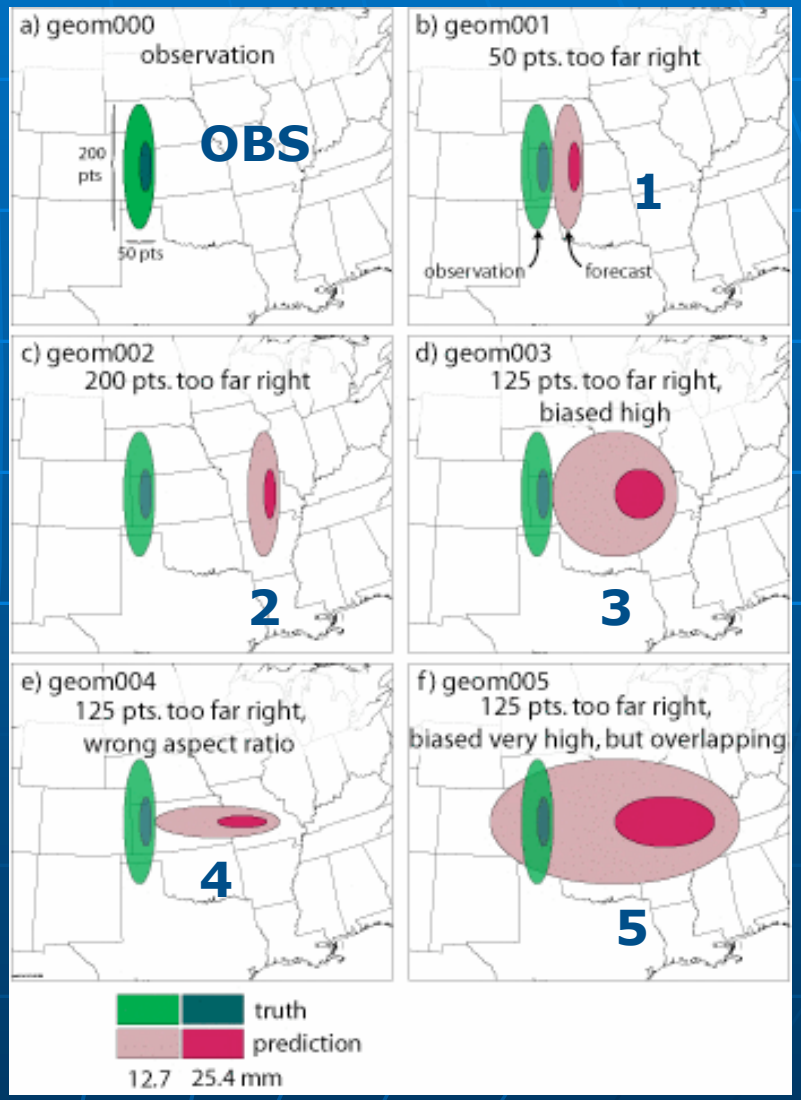
Scores for Examples 1-4:

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

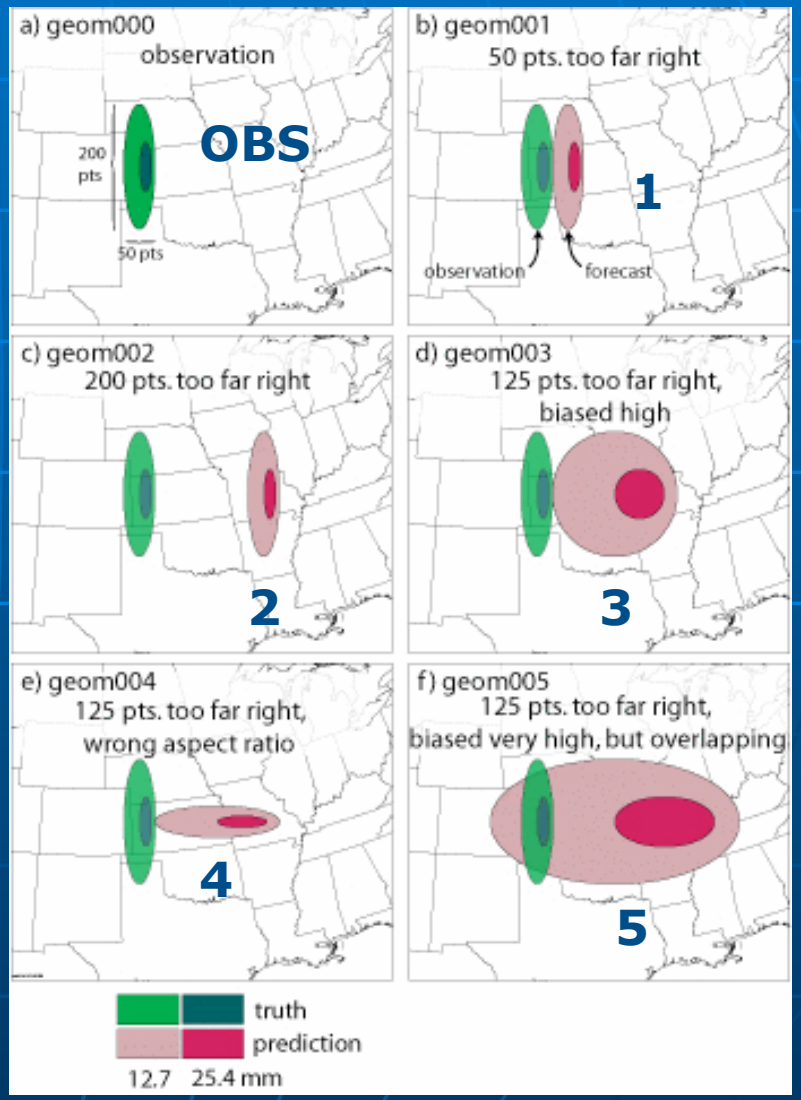
Some problems with the traditional approach:

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

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

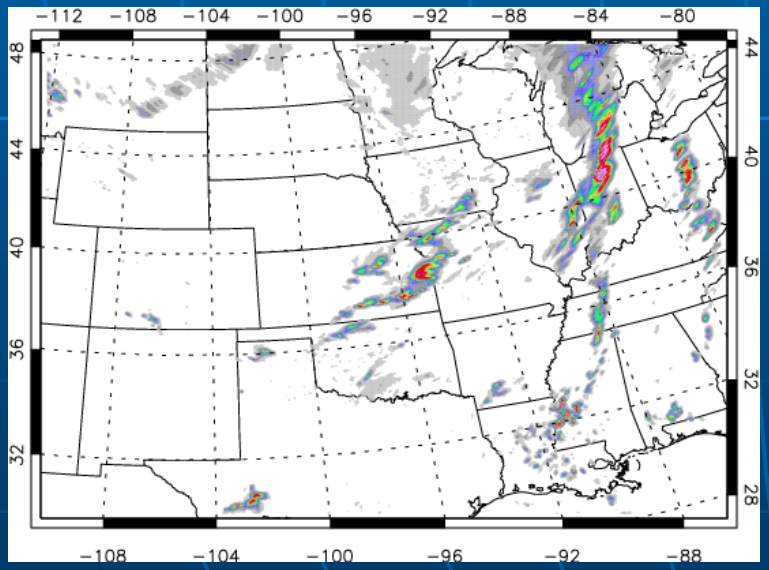
(3) No user-relevant information - how do I use the forecast to make a decision?

(4) Subjective assessments often disagree with scores



Spatial forecasts

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



New spatial verification techniques aim to:

- Account for
  - Uncertainties about location
  - Spatial structure
- Provide information that represents error in physical terms

New spatial verification approaches

Neighborhood

Give credit to "close" forecasts

Features-based

Evaluate attributes of identifiable features

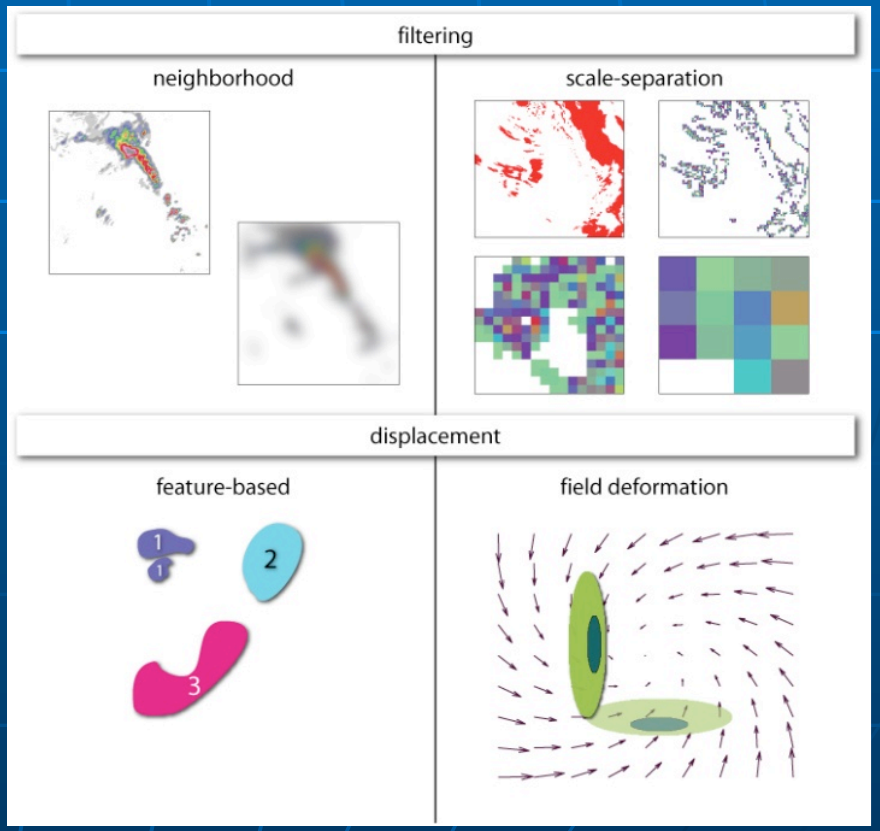
Scale-separation

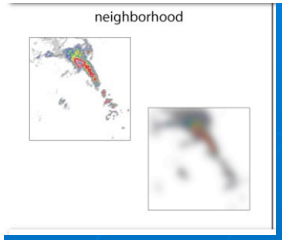
Measure scale-dependent error

Field deformation

Measure distortion and displacement (phase error) for whole field

How should the forecast be adjusted to make the best match with the observed field?

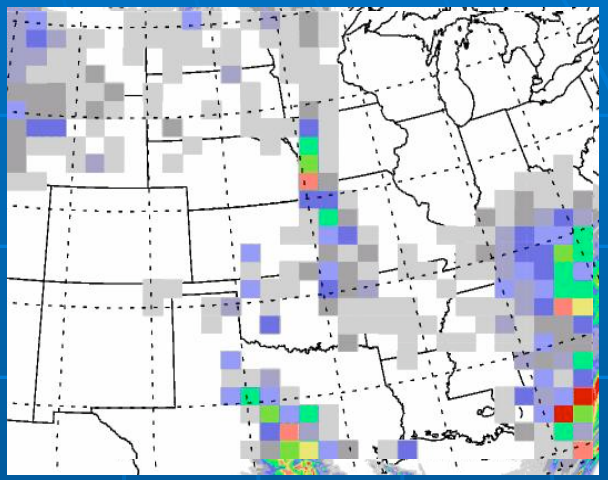




Neighborhood methods

Goal: Examine forecast performance in a region; don't require exact matches

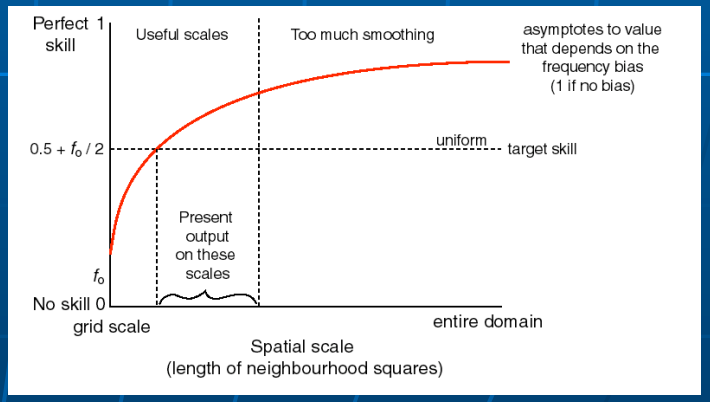
Provide information about scales where the forecasts have skill



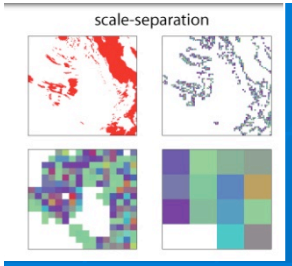
Examples:

- Upscaling
  - Put observations and/or forecast on coarser grid
  - Calculate traditional metrics
- Fractions Skill Score (FSS) (Roberts and Lean 2008, Mittermaier and Roberts 2009)
- Many Others

(e.g., Ebert (2008); Atger (2001); Marsigli et al. (2006))



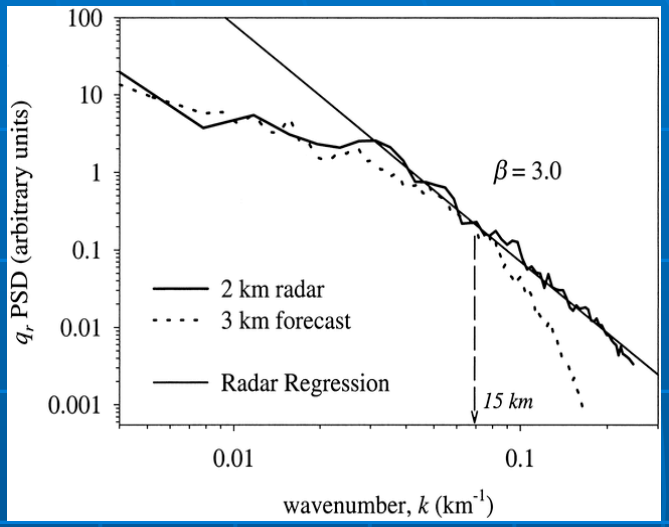
From Mittermaier 2008



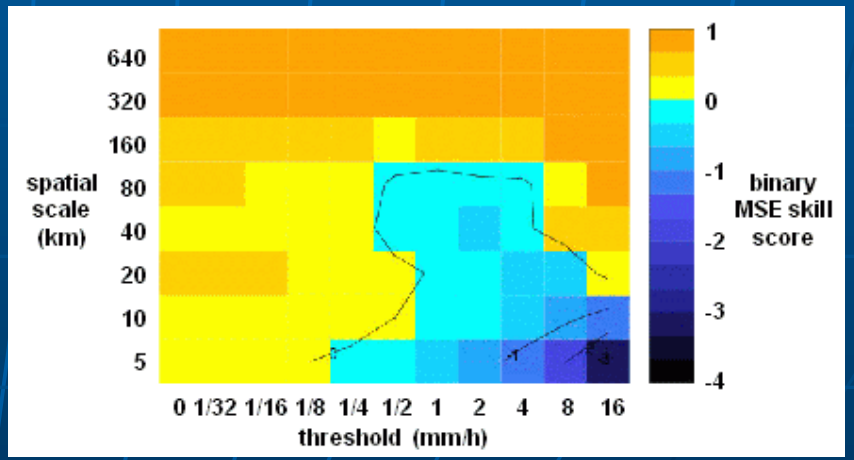
Scale-separation methods

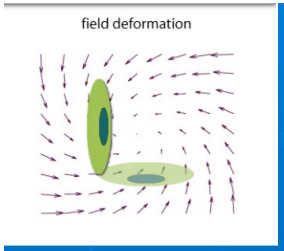
- Goal:
  Examine performance as a function of spatial scale

- Examples:
  - Power spectra
    - Does it look real?
    - Harris et al. (2001)
  - Wavelets
    - Briggs and Levine (1996)
  - 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)

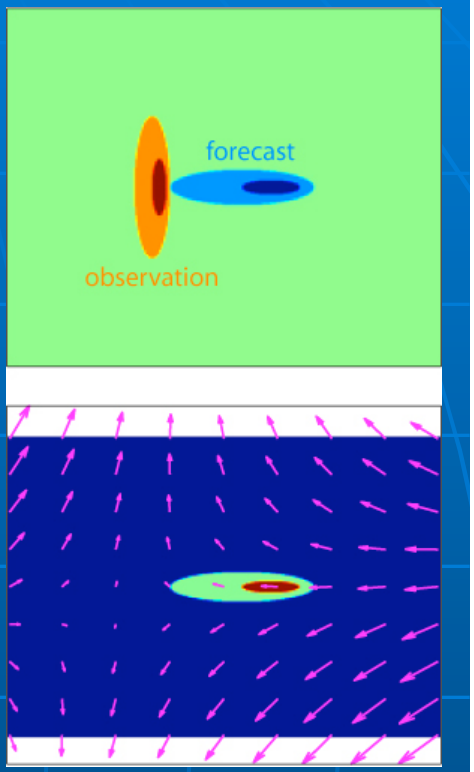


From Harris et al. 2001





Field deformation

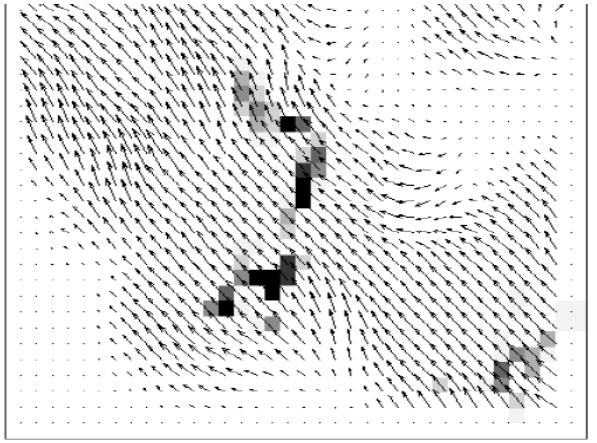


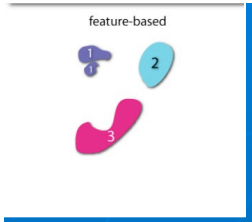
Goal: Examine how much a forecast field needs to be transformed in order to match the observed field

Examples:

- Optical Flow
  (Keil and Craig 2008, 2009, Marzban et al. 2009)
- Image Warping
  (Gilleland et al. 2009; Lindström *et al.* 2009; Engel 2009)
- Gaussian mixtures (New)
  (Lakshmanan and Kain, 2009)
- Forecast Quality Index
  (Venugopal *et al*. 2005)

From Keil and Craig 2008



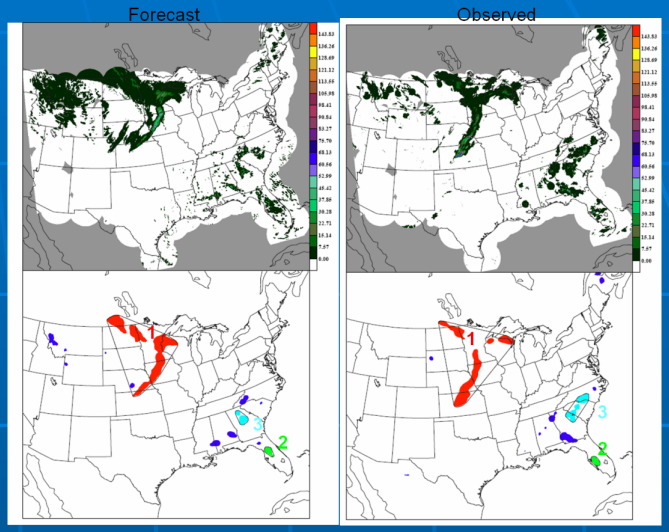


Features-based

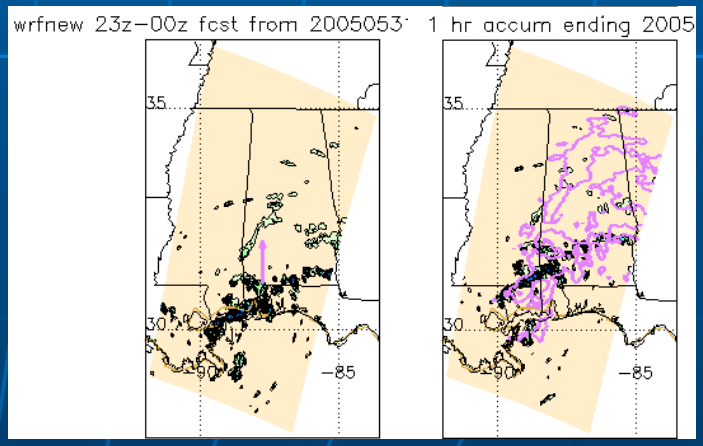
Goals: Measure and compare (user-) relevant features in the forecast and observed fields

Examples:

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



MODE example 2008



CRA: Ebert and Gallus 2009

So many choices: How do you choose what to use?

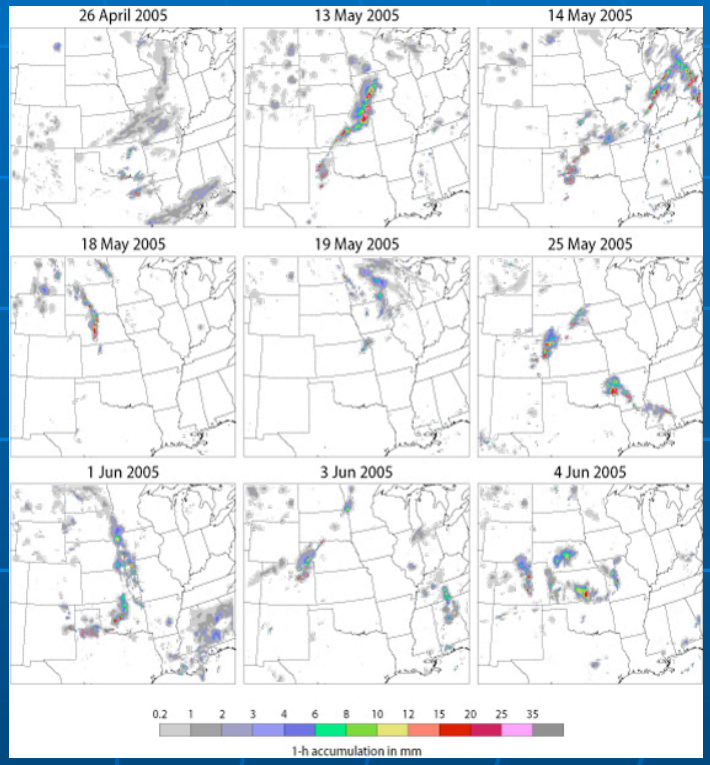
Spatial verification method intercomparison project

- Goals:
  - Assess how methods work with the same datasets
  - Understand characteristics of methods and results
  - Assess strengths and weaknesses
  - Identify information provided by each method
- International collaborative project
  - Many contributors

ICP Test Cases

- Geometric cases
- "Real" cases
  - Actual precipitation forecasts and analysis
  - – Central U.S. WRF forecasts (~4 km)
  - – Stage II precipitation analysis
  - – May-Jun 2005
    (9 focus cases)
  - *Subjective evaluation*
- Perturbed cases

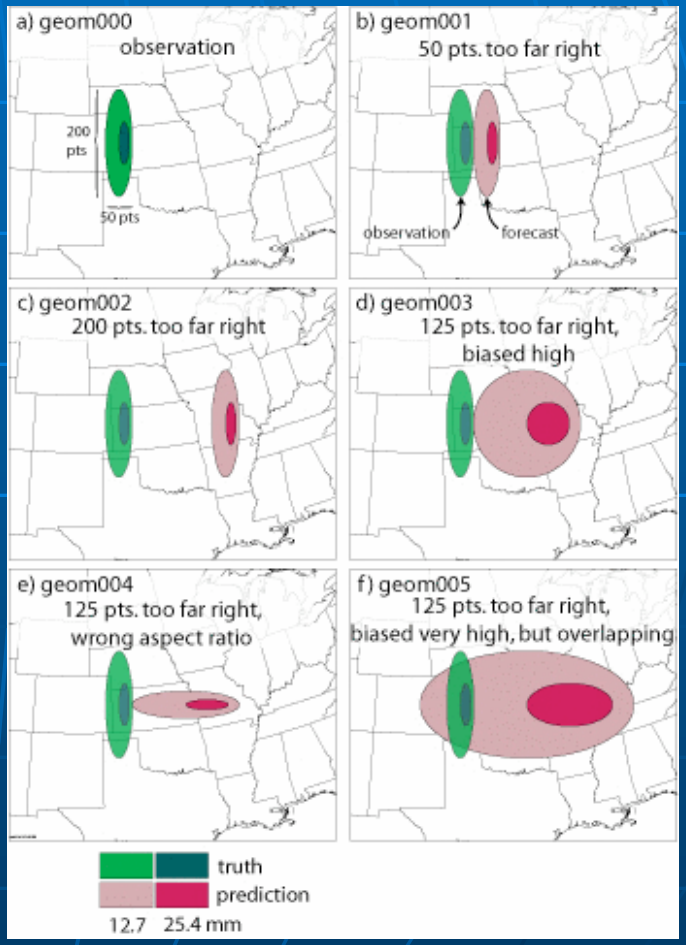
"Real" test cases



Geometric results

Error type	Method Category			
	Neigh-bor-hood	Scale-separation	Feature based	Field defor-mation
Displace-ment *(geom001 geom002)*	No	No	Yes	**Yes**
Frequency bias *(geom003 geom005)*	Yes	Yes	Yes	**Yes**
Aspect ratio *(geom004)*	**No**	**No**	**No**	**Yes**

Geometric cases



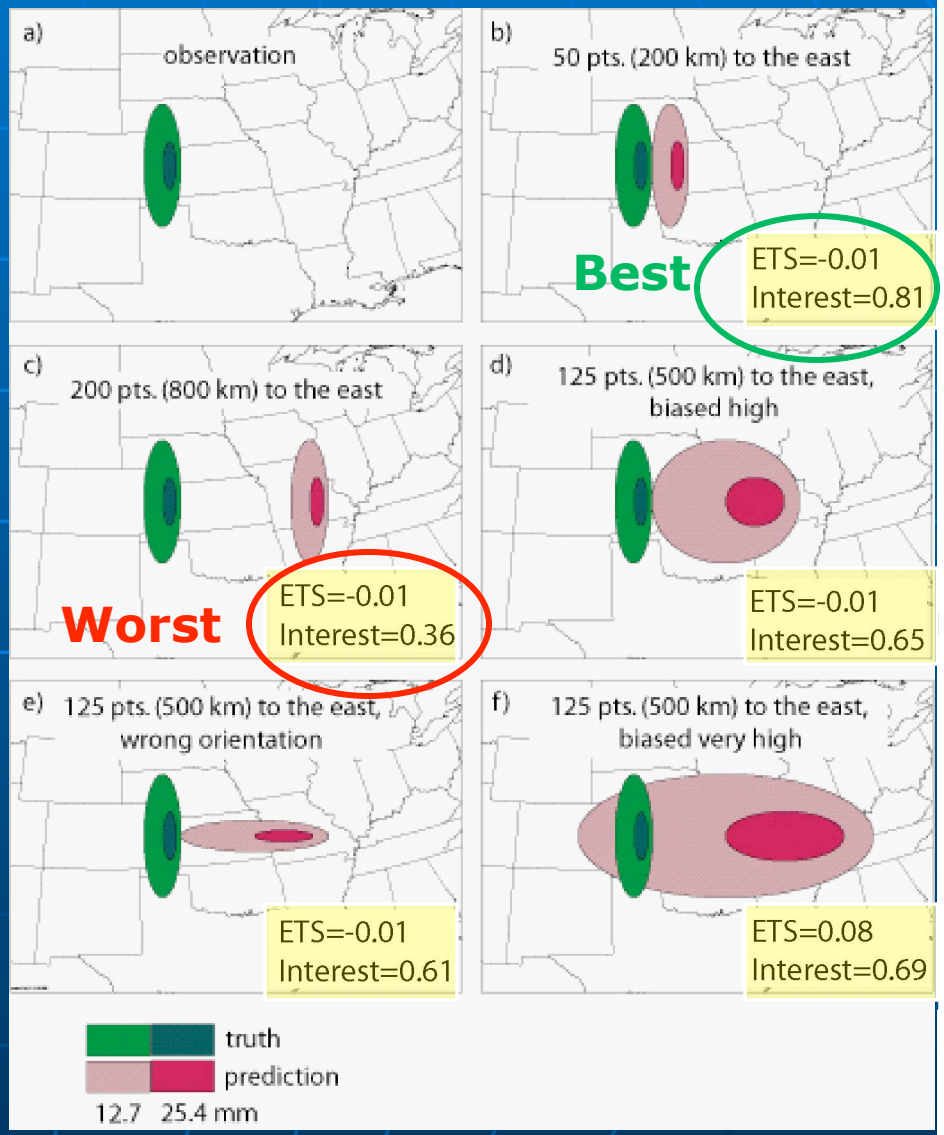
What do the new methods measure?

	Traditional	*Features-based*	*Neighbor-hood*	*Scale-sep*	*Field Defor-mation*
Scale	Indirectly	Indirectly	**Yes**	**Yes**	Yes*
Location	No	**Yes**	Indirectly	Indirectly	**Yes**
Intensity	**Yes**	**Yes**	**Yes**	**Yes**	**Yes**
Structure	No	**Yes**	No	No	**Yes**
Hits, etc.	**Yes**	**Yes**	**Yes**	Indirectly	**Yes**

* Not so much as has been proposed already, but it is theoretically possibly to do.

Back to the original 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 – even for spatial methods, Single measures don't tell the whole story!



Summary and Poursuivre

- ICP results provide guidance on interpretation and applicability of many of the newly proposed (user-relevant) spatial verification methods
- Methods provide information that can
  - Feed back into forecast development
  - Provide information to users
- Next steps
  - Additional datasets
  - Time domain
  - New variables: clouds, wind
  - Verification testbed?

Spatial Methods and the ICP

For more information (and full references from this talk), see
http://www.ral.ucar.edu/projects/icp

Upcoming special collection of *Weather and Forecasting*

- *Two overview papers (Ahijevych et al 2009; Gilleland et al 2009)*
- *12 Papers on individual methods*

Some spatial methods are available as part of the Model Evaluation Tools (MET) at
http://www.dtcenter.org/met/users