Gridded Forecast Verification

2011 Joint Statistical Meetings, July 30 - August 4, 2011, Miami Beach, Florida, Miami Beach Convention Center

Eric Gilleland

Research Applications Laboratory,

National Center for Atmospheric Research

Co-authors: D. Ahijevych, B. G. Brown, B. Casati, L. Chen, M. DePersio, G. Do, E. E. Ebert, K. Eilertson, Y. Jin, E.L. Kang, F. Lindgren, J. Lindström, R. L. Smith, and C. Xia

Background and Motivation

Forecast Verification: High vs. Low Resolution

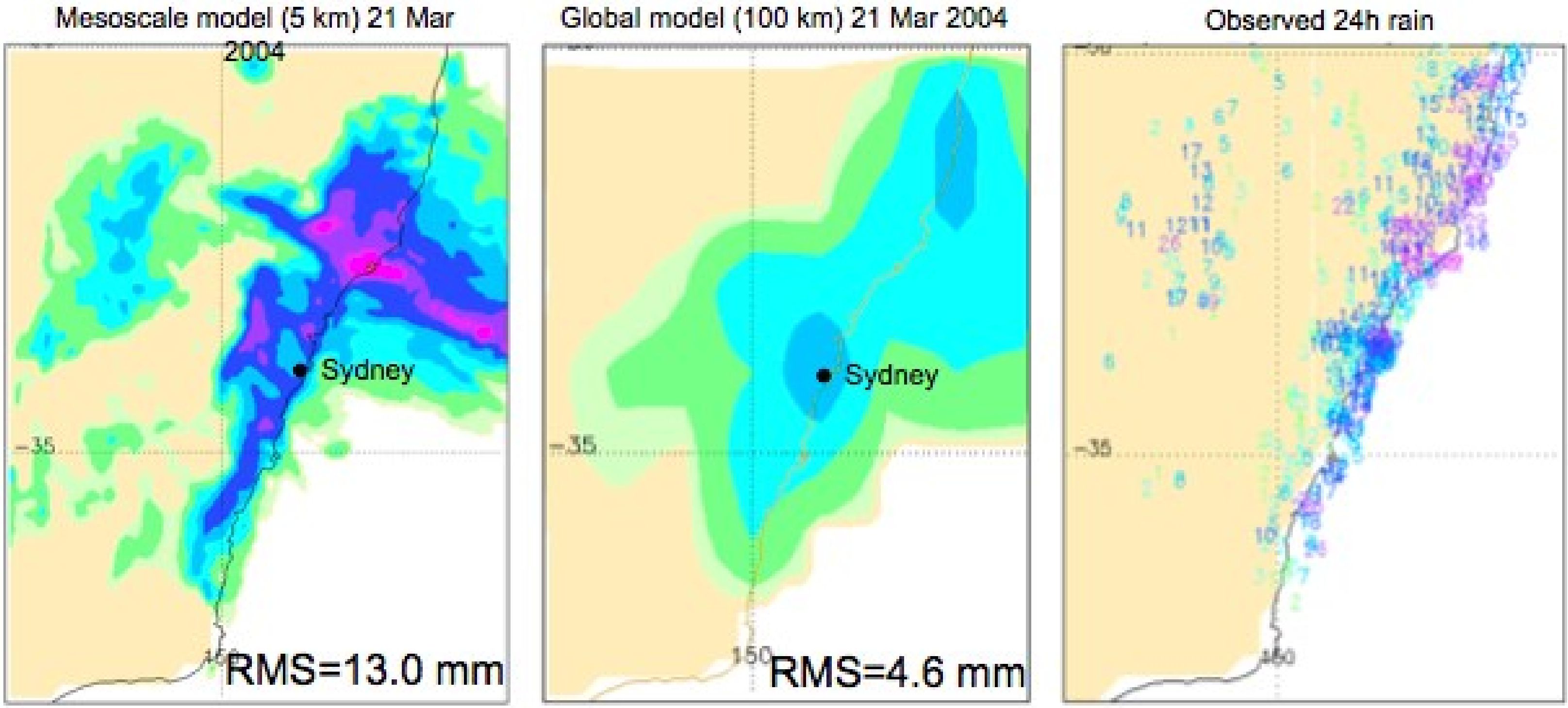
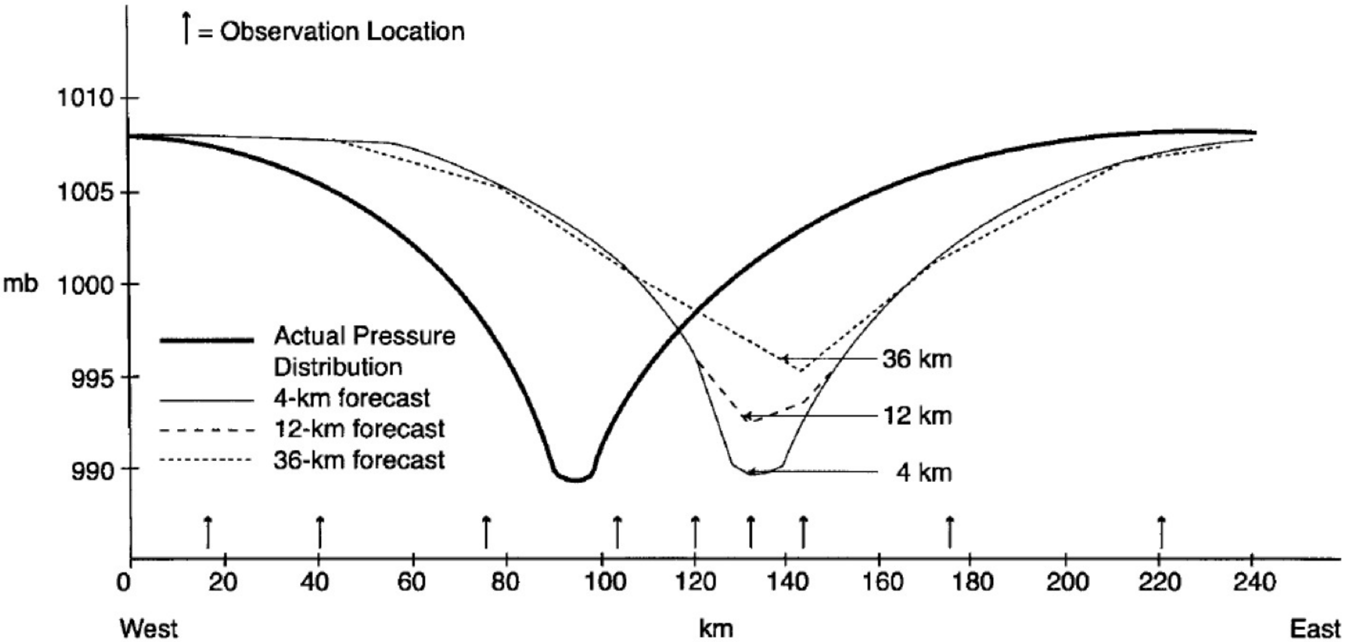


Fig. from E.E. Ebert

Which forecast would you prefer to use?

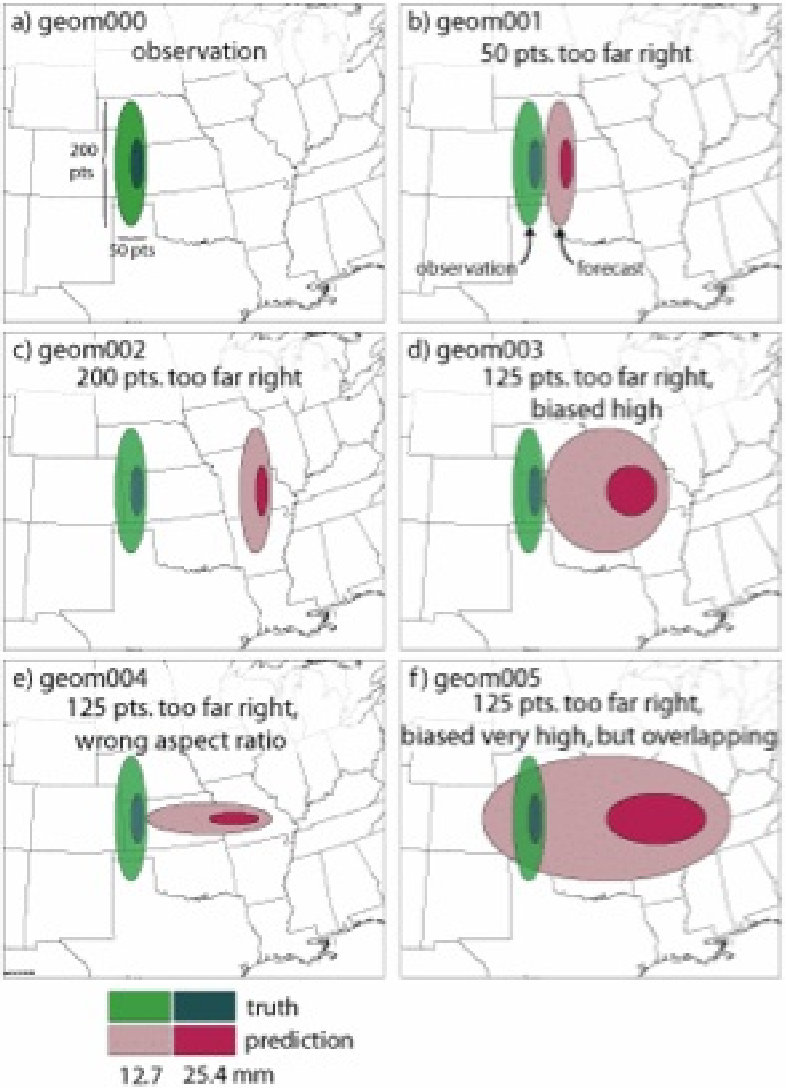
Forecast Verification

Mass *et al.*, 2002: Bull. Amer. Meteorol. Soc., **83**, 407–430.



Timing error: Traditional ***grid-point to grid-point*** verification yields RMSE of 4.19-, 4.81- and 5.25- mb for 36-, 12- and 4-km, resp.

Background and Motivation



Traditional Approach

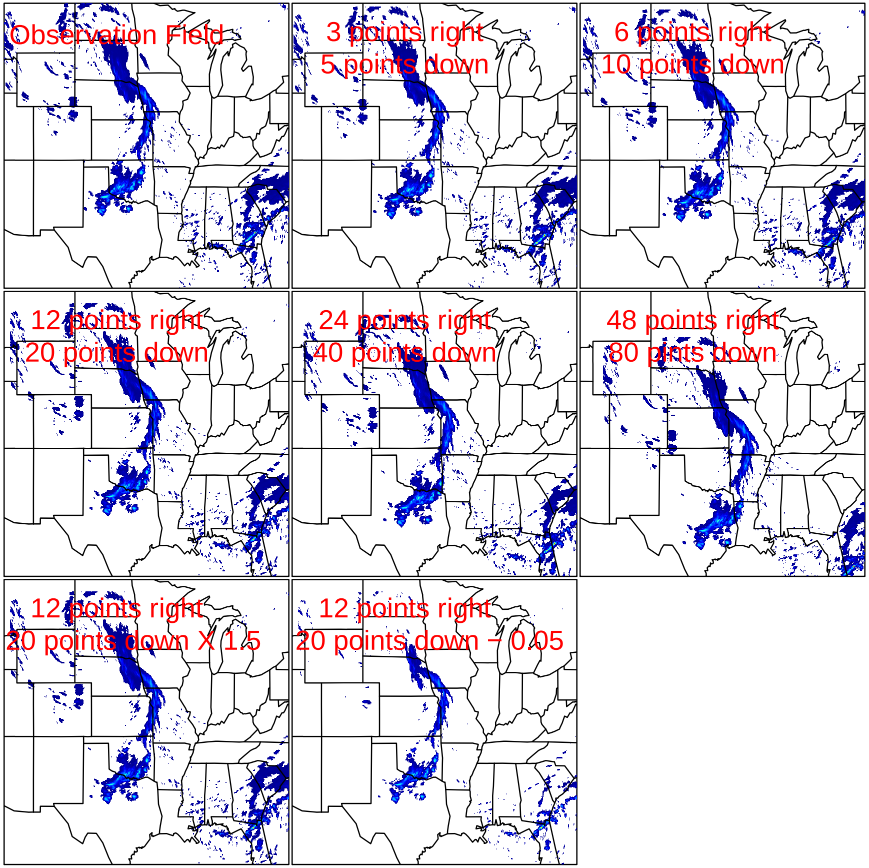
Based on comparing overlapping grid points...

Score	1-4	5
Correlation Coefficient	-0.02	0.2
Probability of Detection	0.00	0.88
False Alarm Ratio	1.00	0.89
Hanssen-Kuipers	-0.03	0.69
Gilbert Skill Score	-0.01	0.08

Forecast 5 is “best"

Spatial Forecast Verification Methods

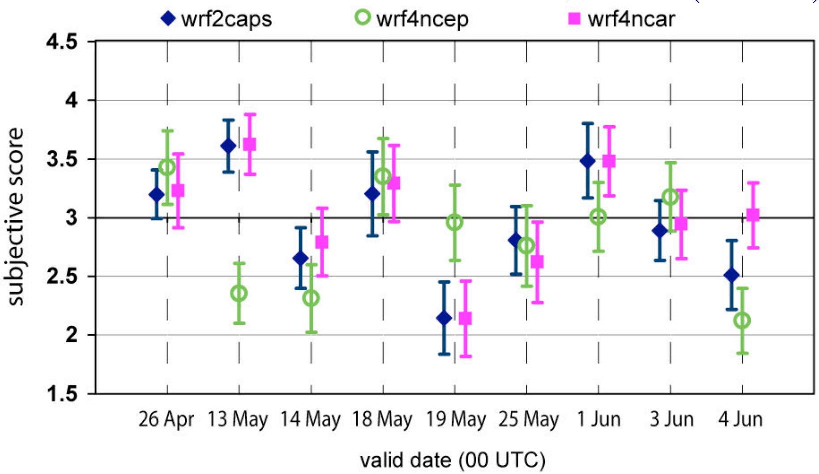
Inter-Comparison Project (ICP)



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

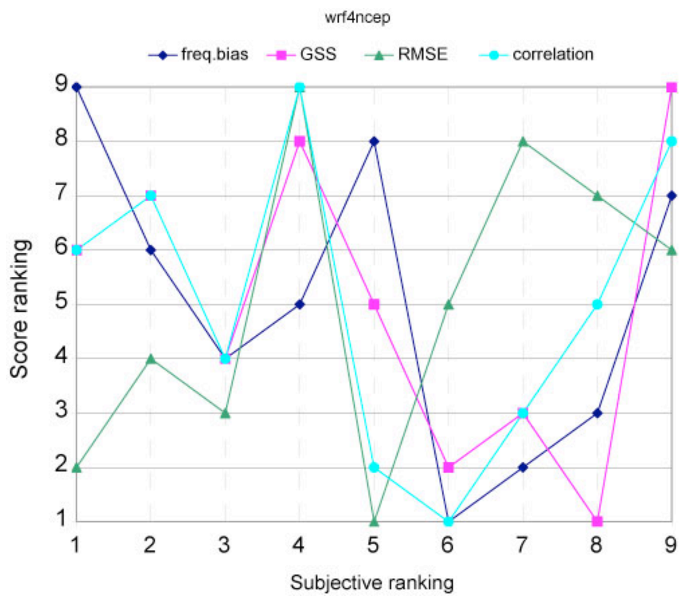
Spatial Forecast Verification Methods

Inter-Comparison Project (ICP)



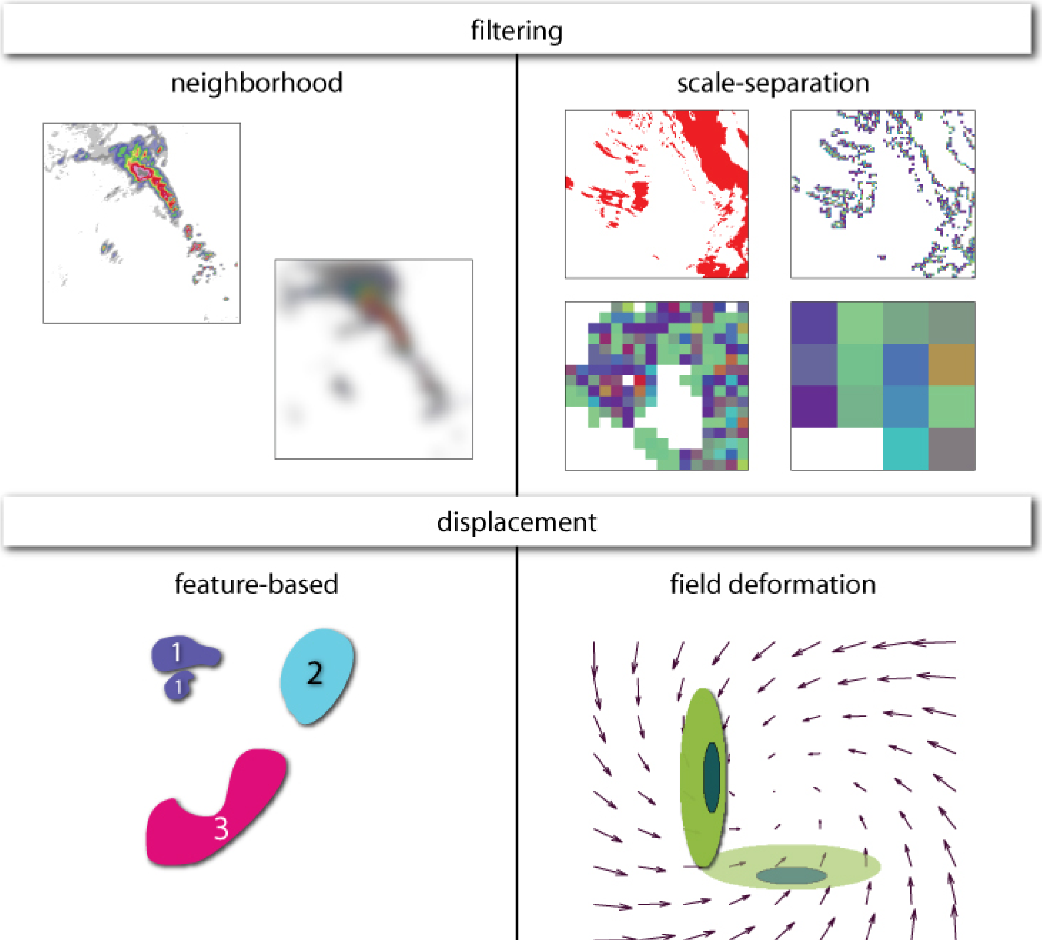
Ahijevych *et al.* (2009);
G. *et al*, (2009, 2010)

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



Spatial Forecast Verification Methods

Inter-Comparison Project (ICP)



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

Filter Methods : Neighborhood

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

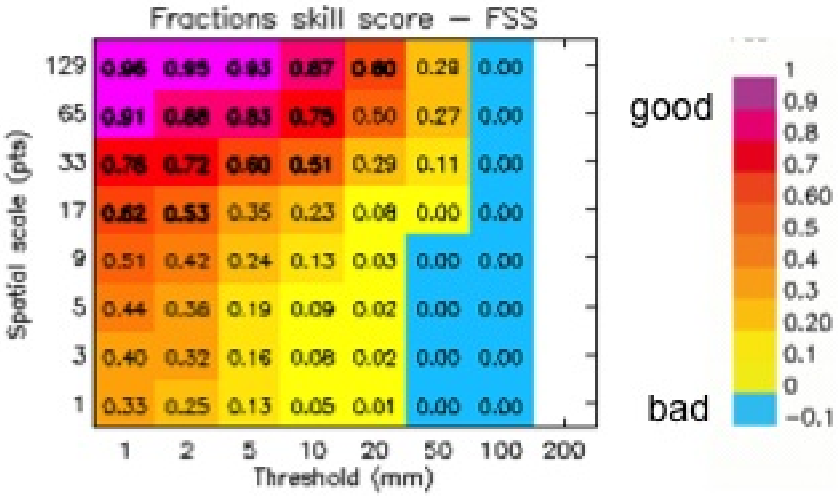
Goal: Examine forecast performance in a region without requiring exact grid-point to grid-point matches.

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).

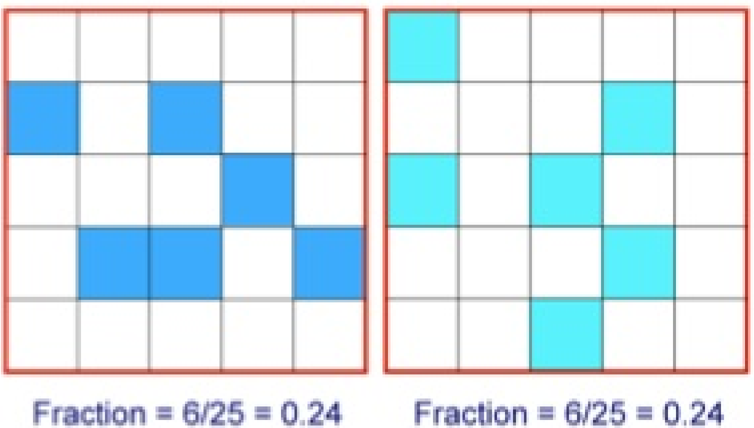
Disadvantages: Limited diagnostic information. Do not inform about specific error types, but may be sensitive to them. Do not inform about spatial structure errors.

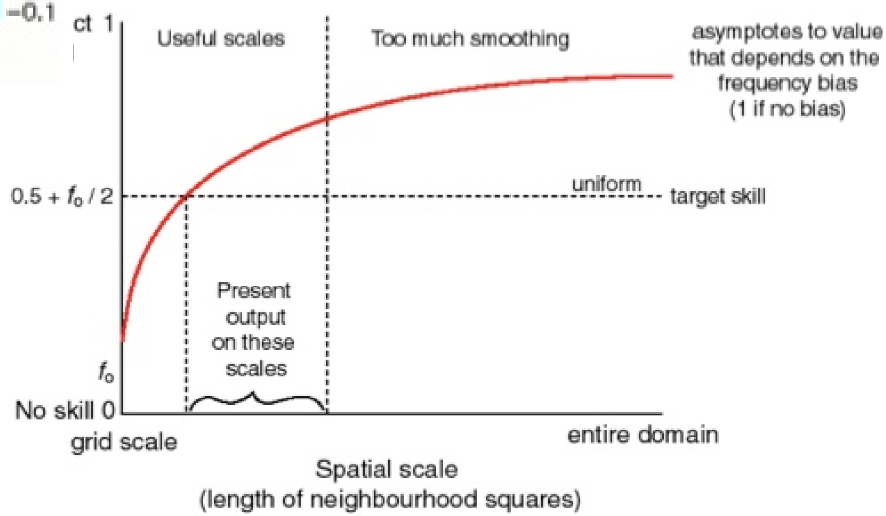
Examples: Simplest example is *upscaling*. Many such methods have been proposed (Ebert, 2008 gives a nice review).

Filter Methods: Neighborhood



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

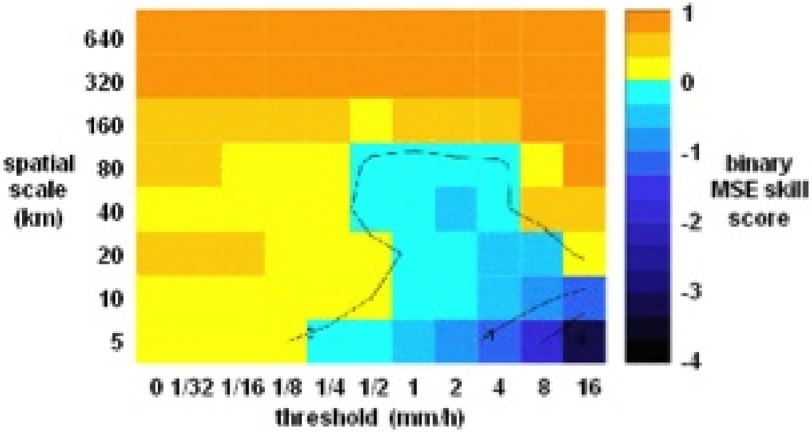




Fractions Skill Score (Robert and Lean, 2008)

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)

Displacement Methods: Features based

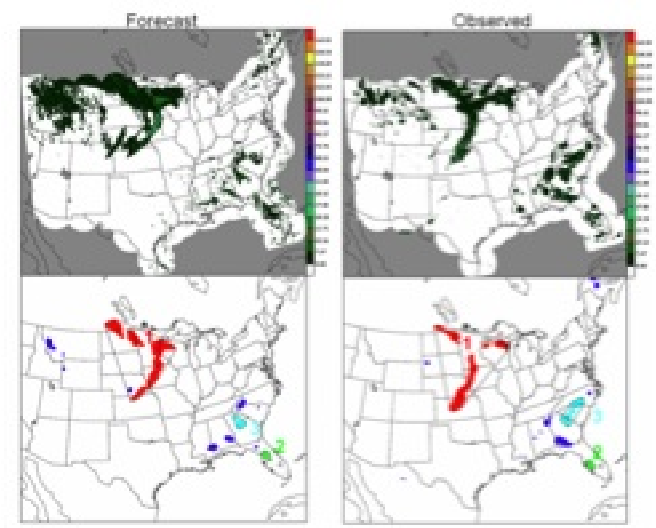
Model: $O_A(x, y) = F_B(x, y) + \varepsilon$

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

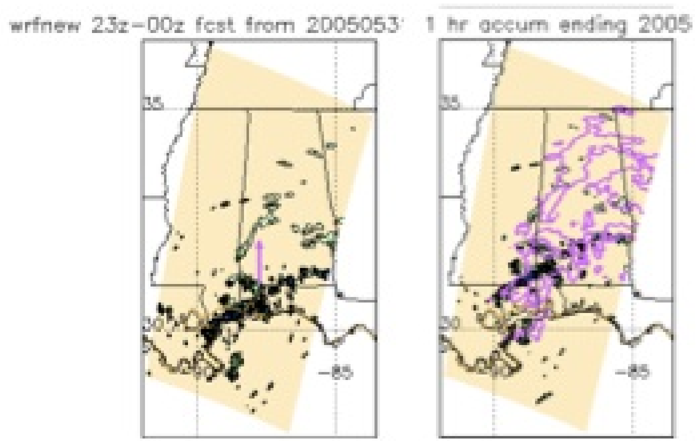
Examples:

- 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.

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 (Venugopal *et al.*, 2005; G. 2011;
Schwedler and Baldwin, 2011;
Zhu *et al.*, Submitted)

Optical Flow (e.g., Keil and Craig, 2008, 2009)

Image Warping (e.g., Alexander *et al.*, 1998;
G., Lindström and Lindgren, 2010)

Distortion representation (e.g., Hoffman *et al.*, 1995)

Gaussian mixtures (Lakshmanan and Kain, 2009)

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., G., 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: Field Deformation

Field Deformation Methods: Image Warping

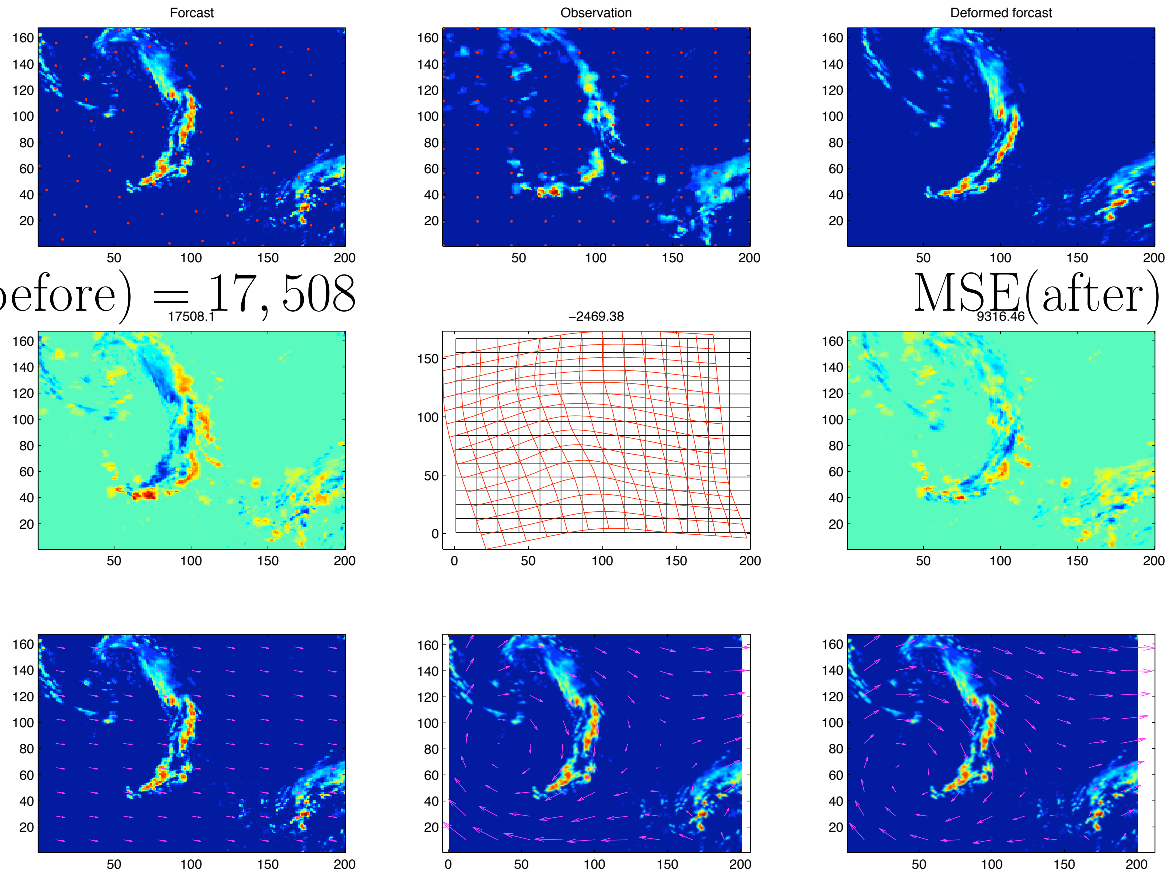
For the TPS Warp, the likelihood leads to the following loss function. The optimization of which yields the optimal warp.

$$Q(\boldsymbol{p}^F) = \frac{1}{2\sigma_\varepsilon^2} \sum (O(\boldsymbol{s}) - F(\boldsymbol{W}(\boldsymbol{s})))^2 +$$

$$\frac{1}{2\sigma_\Delta}\left[(\boldsymbol{p}_x^F - \boldsymbol{p}_x^O)^T(\boldsymbol{I} - \boldsymbol{C})(\boldsymbol{p}_x^F - \boldsymbol{p}_x^O) + (\boldsymbol{p}_y^F - \boldsymbol{p}_y^O)^T(\boldsymbol{I} - \boldsymbol{C})(\boldsymbol{p}_y^F - \boldsymbol{p}_y^O)\right]$$

See, e.g., Åberg *et al.*, *Environmetrics*, **16**(8):833–848, 2005.

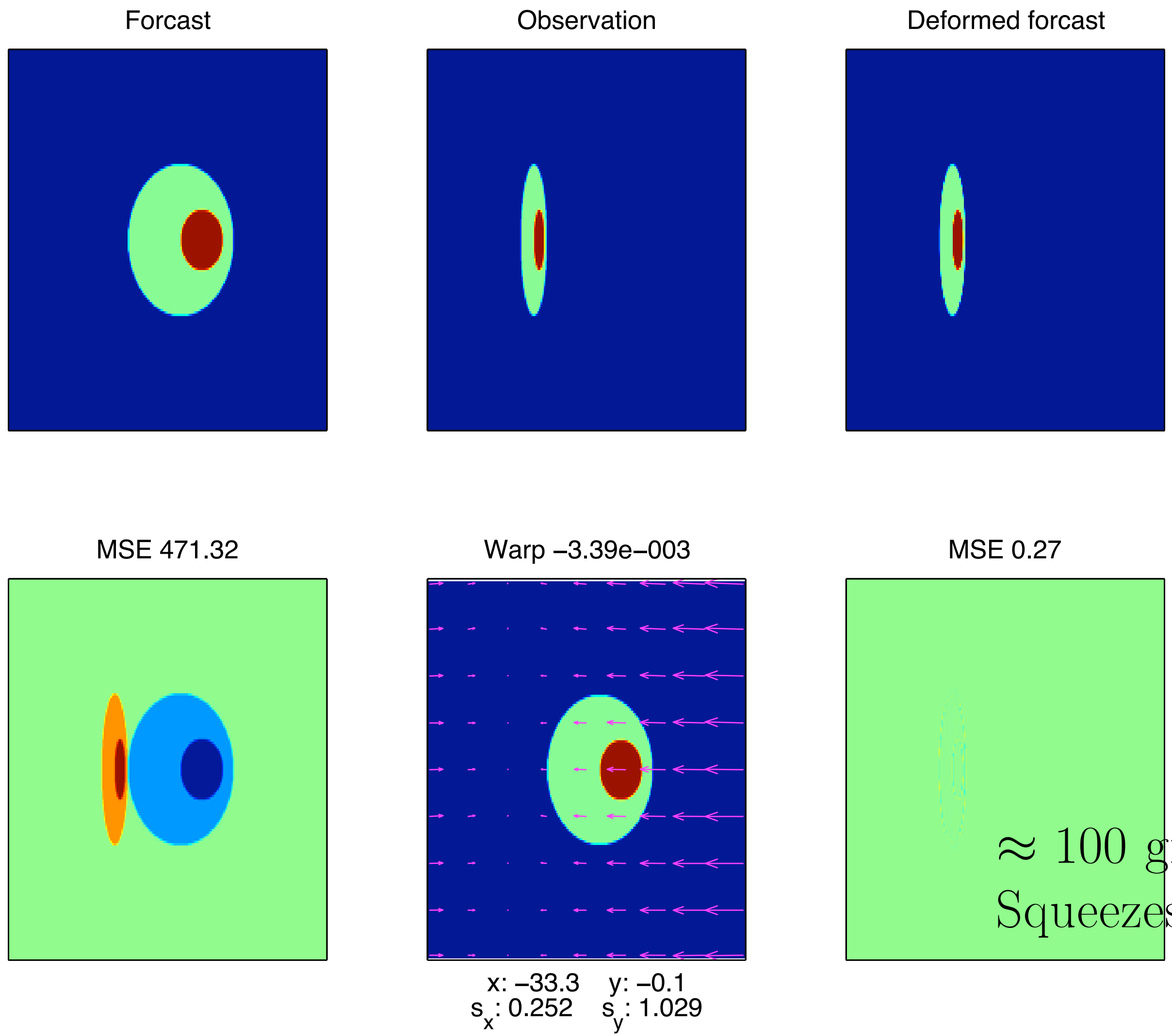
ICP Test Cases



$\mathrm{MSE(before)} = 17,508$

$\mathrm{MSE(after)} = 9,316$

$$\frac{17{,}508 - 9316}{17{,}508} \approx 47\%$$



≈ 100 grid points west
Squeezes horizontally.

Geometric 3; 125 grid points too far east and larger spatial coverage

Space-Time Image Warp

Industrial Mathematical and Statistical Modeling Workshop for Graduate Students

July 19-27, 2010
North Carolina State University, Raleigh, NC

Sponsored by
Statistical and Applied Mathematical Sciences Institute (SAMSI), RTP, NC

Center for Research in Scientific Computation, Raleigh, NC

G. *et al.*, 2011. Spatial Forecast Verification: Image Warping. NCAR Technical Note, TN-482+STR, 23pp.

Space-Time Image Warp

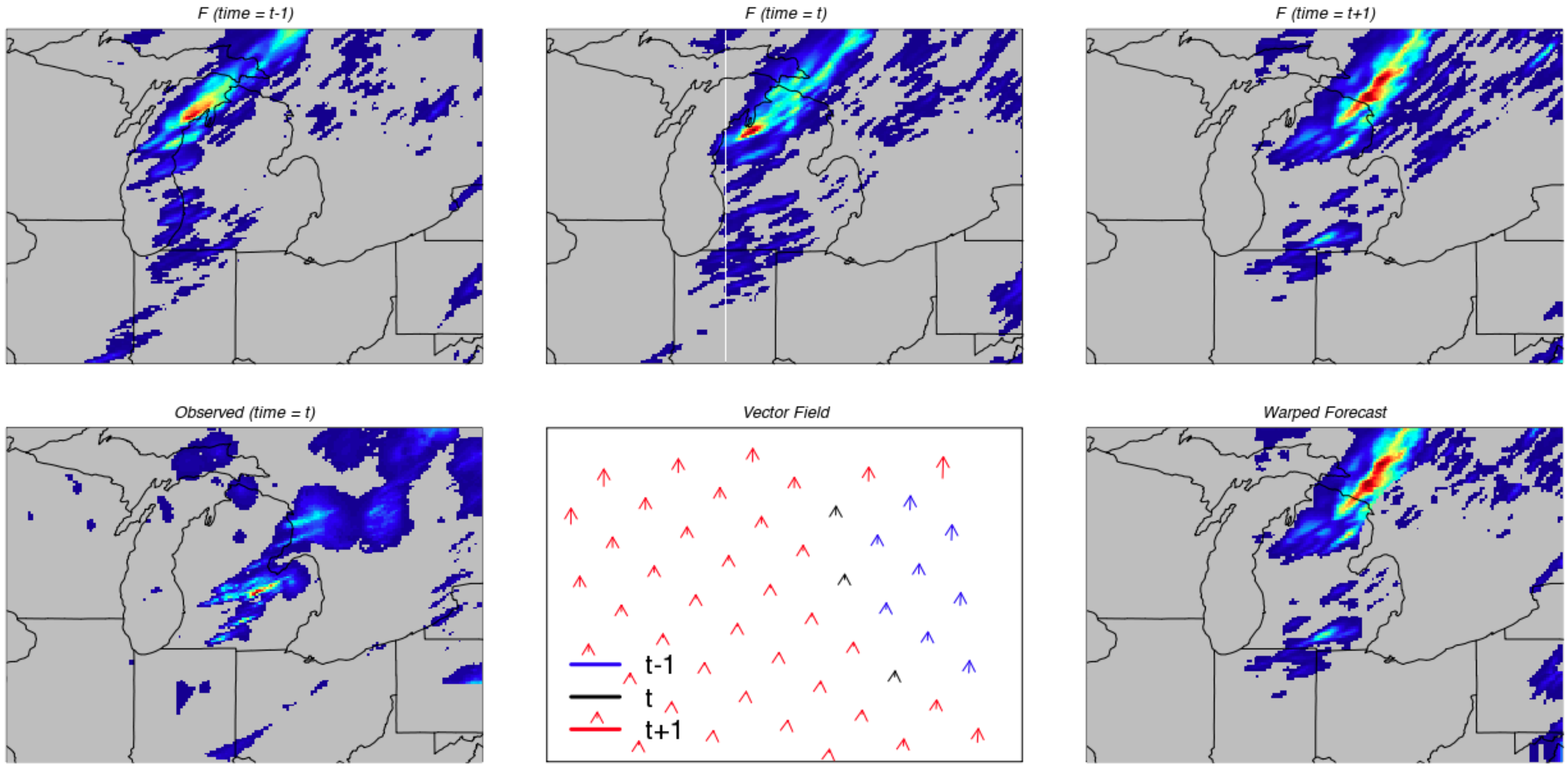
For a given *initialization time*, a forecast model will give predictions, say every hour, up to some distance in time. Timing error implies that the forecast at time $t-1$ (or $t+1$) is better than that at time t. But, maybe only for one storm feature or area. Can these errors be distinguished from true spatial displacement errors?

Extension of 2-d spatial warping to space-time

Equations about the same, but with the added dimension. Tri-harmonic basis functions instead of 2-d TPS radial basis functions.

Space-Time Image Warp

Example



Reduction in RMSE is over 50% after applying space-time warp. Most errors were spatial only.

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)?