



NCAR



What we learned from the Spatial Forecast Verification Inter-Comparison Project (ICP)

Eric Gilleland

Research Applications Laboratory

National Center for Atmospheric Research

MesoVICT Kick-off Meeting, 2 October
2014, Vienna, Austria

National Center for Atmospheric Research

Recalling the ICP



Motivation for ICP:

- Great number of new verification methods were being developed rapidly by many different researchers
- Confusion about how the methods related to one another, how each method worked, what information could be gleaned by each method, and whether any given method actually conveyed any useful information or not

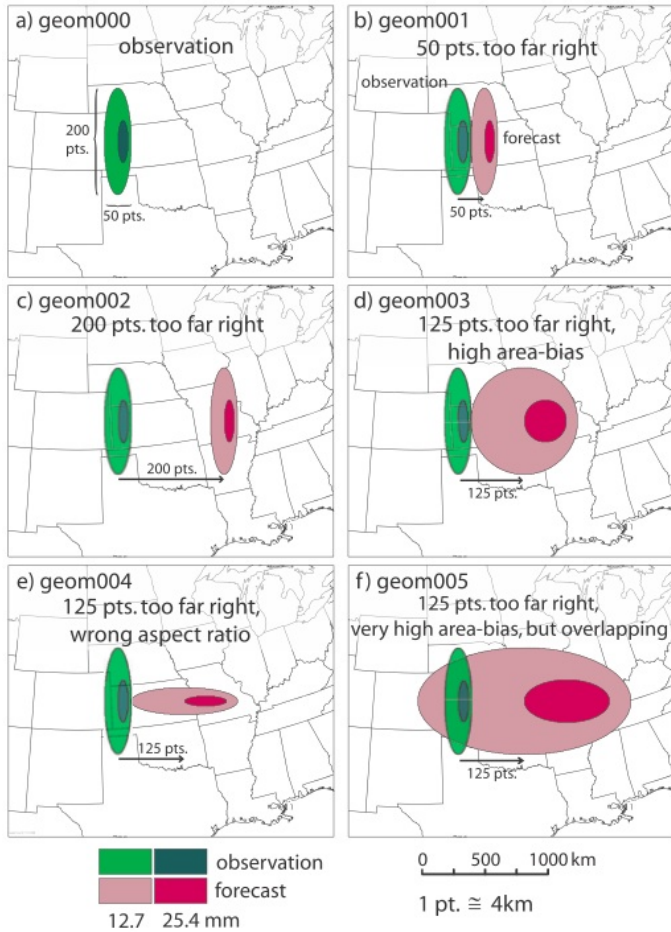
Recalling the ICP



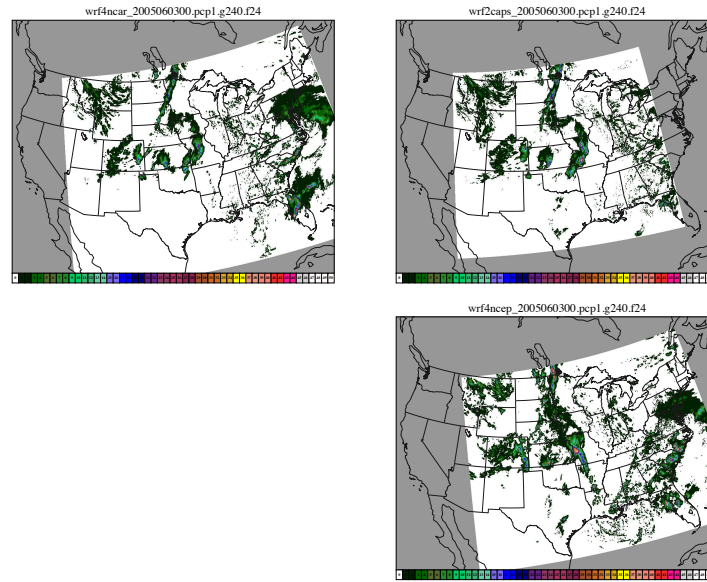
Primary Goals were:

- Try to understand how each method related to other methods
- Determine if some methods could give more information if used in concert
- Learn some of the pros and cons of each method
- Determine if the methods jive with a subjective assessment made by a human

Recalling the ICP



Geometric Cases



9 real cases (additional real cases also available) with three models from the NSSL/SPC Spring 2005 Experiment

7 perturbed cases based on one of the real cases

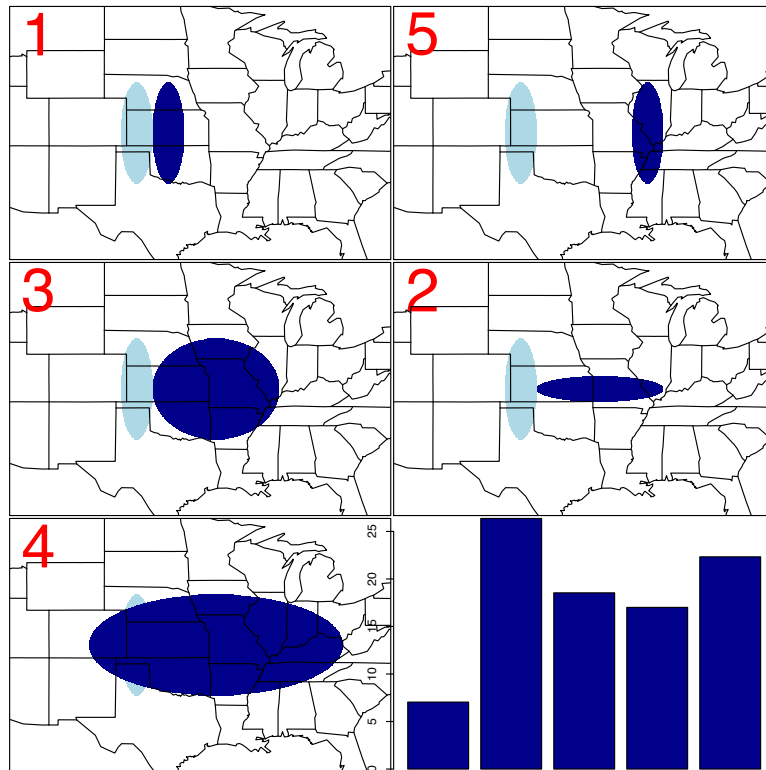
Recalling the ICP



Results papers:

- Special collection of papers (16 in total) in *Weather and Forecasting*
- Overview of data cases paper with traditional verification applied to them, as well as subjective evaluation results: Ahijevych *et al.*, 2009, WAF, 24, 1485 – 1497
- Review papers covering methods employed in the ICP: G. *et al.*, 2009, WAF, 24, 1416 – 1430, G. *et al.*, 2010, BAMS, 91, 1365 – 1373, and Brown *et al.*, 2012, chapter 6 of second edition of “Forecast Verification: A Practitioner’s Guide in Atmospheric Science” edited by I.T. Jolliffe and D. B. Stephenson.

Recalling the ICP



	Baddeley's Δ	DAS	MODE
geom001	1	1	1
geom002	5	2	5
geom003	3	5	3
geom004	4	3	2
geom005	4	3	2

Need for a core case that everyone uses!!!

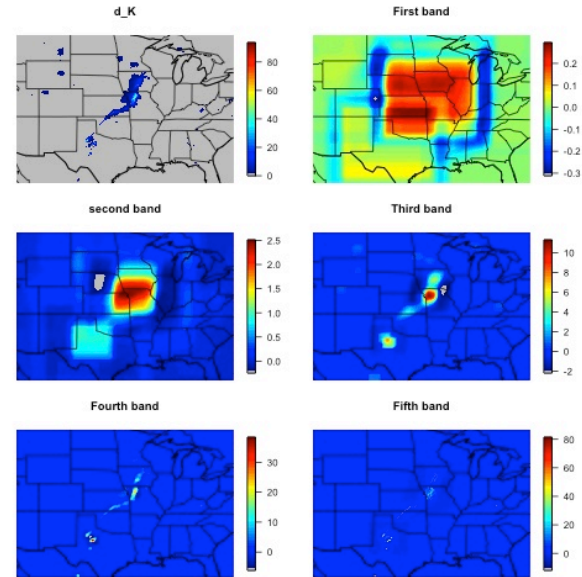
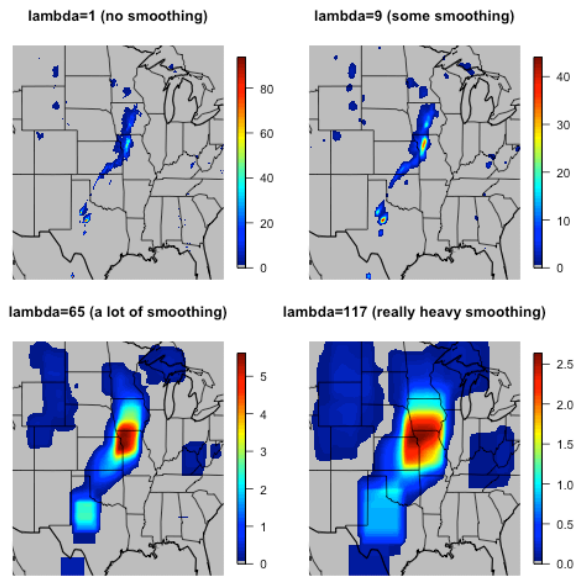
G. (2011, WAF, 26, 409 – 415)

smoothing / neighborhood



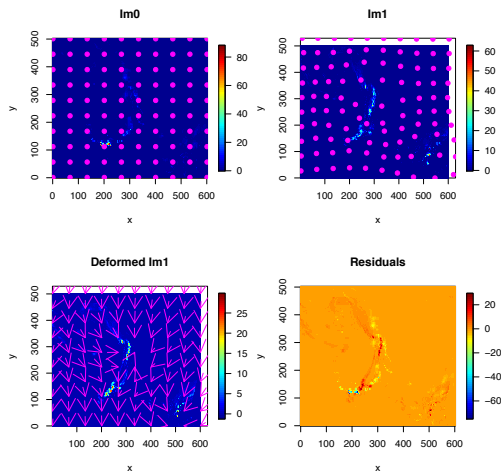
NCAR

Filter

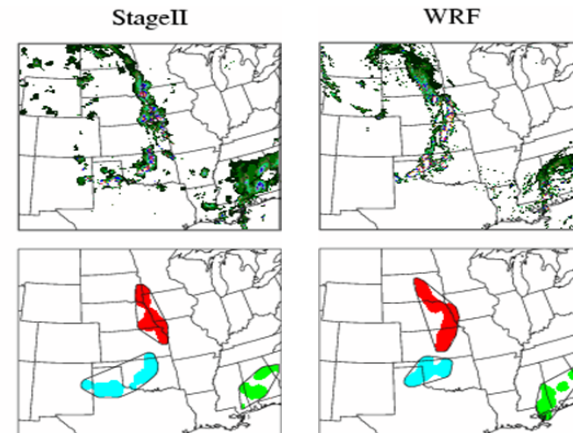


Scale Separation

Entire field



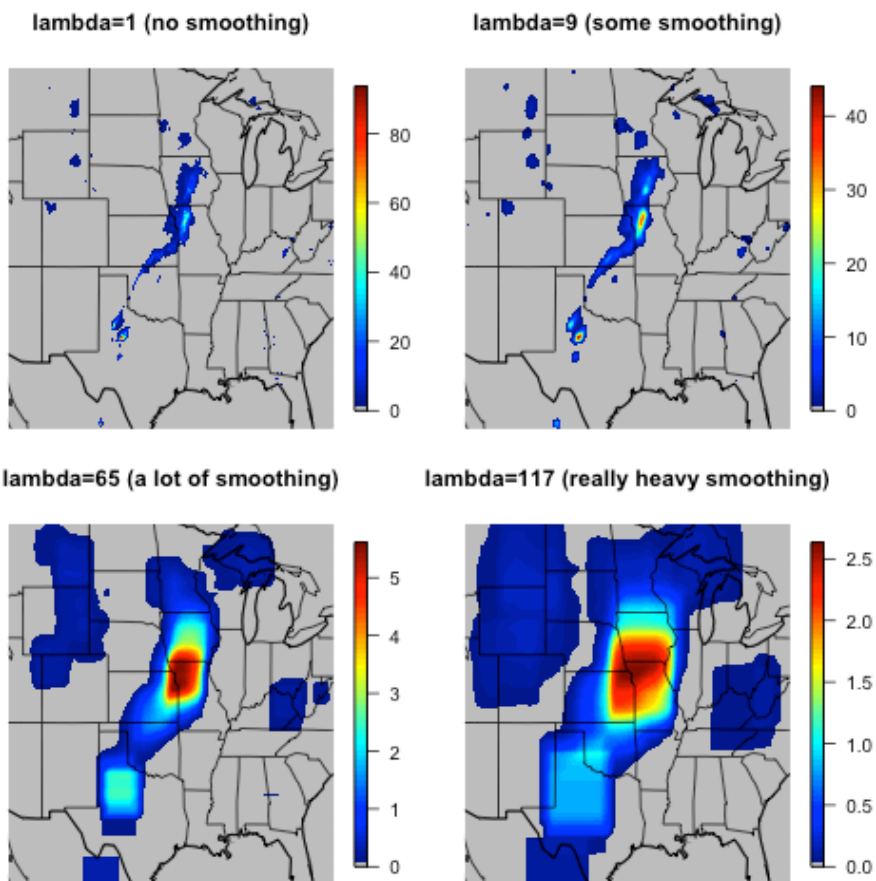
specific features



Spatial Displacement

Recalling the ICP

smoothing filter / Neighborhood Methods



Apply filter to:

- raw field
- binary “thresholded” field (event field)

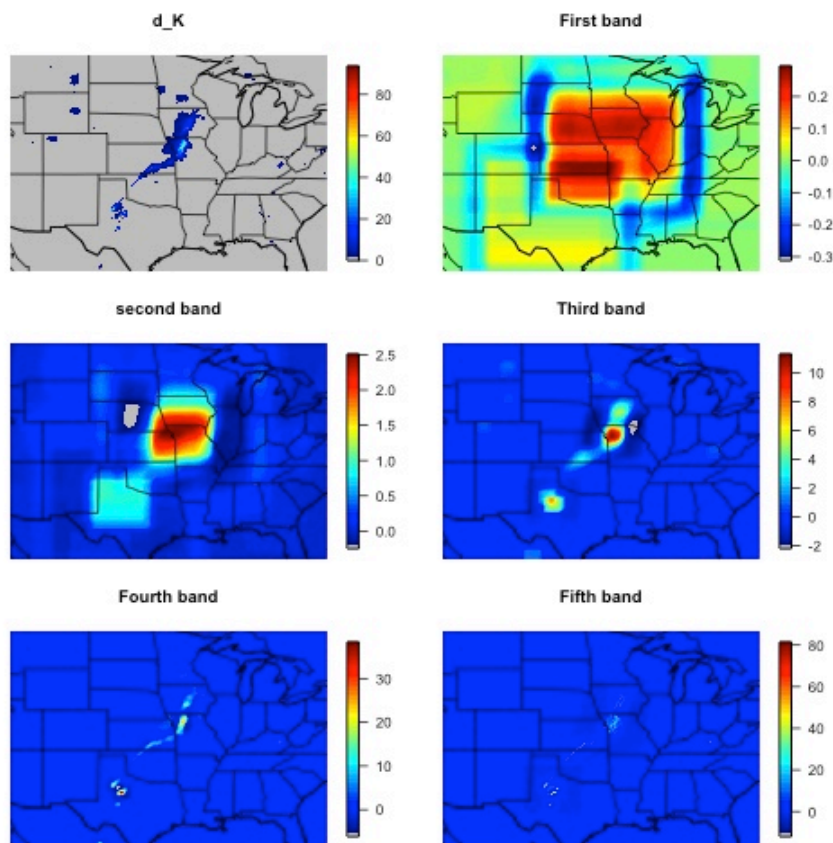
Apply filter to:

- both fields
- forecast only
- observations only

Ebert, 2008, Meteorol. Appl., 15, 51 – 64

Recalling the ICP

scale separation (make use of a band-pass filter)

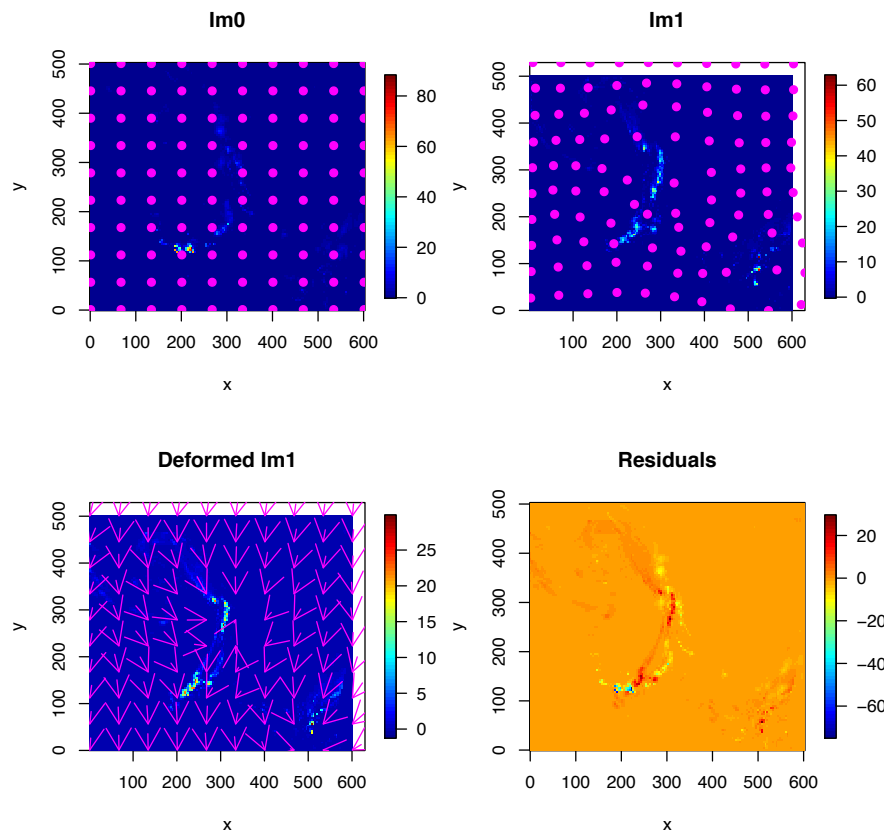


Analogous to a spectral decomposition strategy from univariate time series verification, but perhaps more challenging in space.

Investigate forecast performance at physical scales of interest.

Recalling the ICP

Entire field spatial displacement



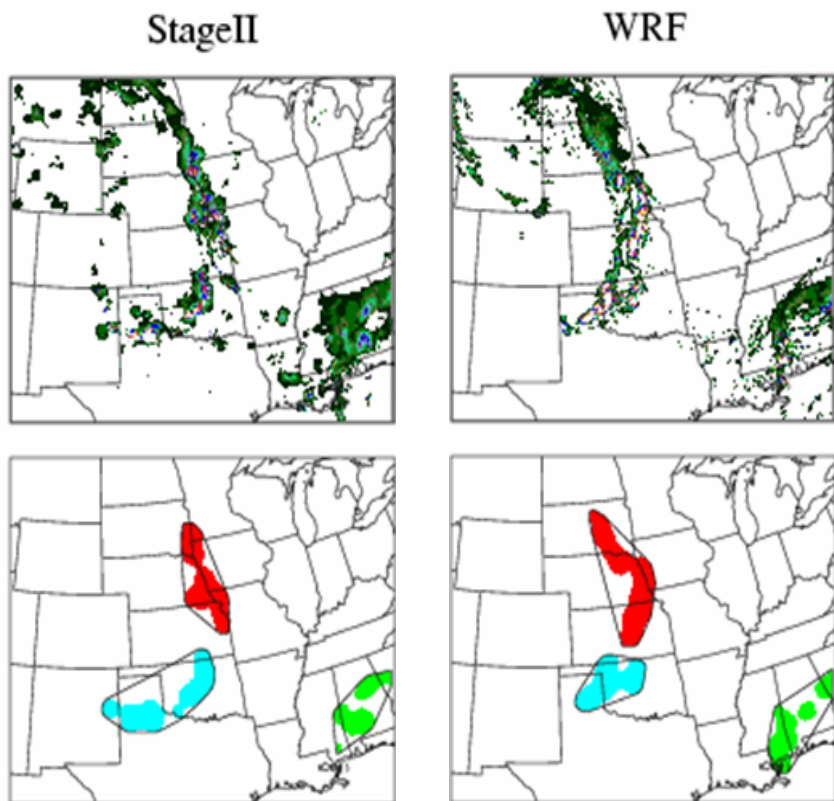
Assumption that small location errors lead to larger than desired penalties (or desire to diagnose presence/impact of location errors)

Two main categories:

- Summary metrics
- re-gridding / field deformation methods to re-calibrate the forecast to align better with the observed field.

Recalling the ICP

Specific features (feature-based) spatial displacement



Investigate individual storms within a field.

Application of entire field displacement methods to these individual features.

Allows for a new “spatial” contingency table.

Excellent for diagnostic evaluation, but also useful for quantitative summaries, etc.

Summary / Conclusions



- Much was learned in the ICP about the fundamentals of the methods
 - Most methods fell into one of two categories, each of which can be further delineated into two separate categories
 - Some methods fit less well, but generally still have some mechanics or facets in common (e.g., variogram vs. scale separation)
- Hybridizing the methods became a clear extension of the methods
- Many methods aimed to inform about the same type of errors, but possibly in very different ways (e.g., feature-based concept of centroid distance and angle difference vs. shape analysis)
- Geometric cases were most informative for understanding most of the methods (some methods do not handle these cases at all!)
- Subjective evaluation is fraught with difficulties, but nevertheless provided a talking point for how the methods behave
- Perturbed cases were difficult to assess true rankings because of small but important features in the corner of the cases
- Cases covered only the relatively flat terrain of the central United States
- No meteorological scenarios included
- Assumption that reanalysis fields are good surrogates for the truth (no point observations used as most methods require gridded fields; or do they?)
- some methods attempted to incorporate sampling uncertainty (usually by way of bootstrapping)