Current State of Verification

Barbara Brown

Joint Numerical Testbed
Research Applications Laboratory, NCAR

2 November 2010

DTC Verification Workshop

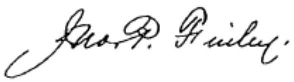
Historical perspective: The Finley example

- U.S. Army (Signal Service/Corps) 1877-~1920
- Tornado predictions 1884-1885
  - Two 8-h outlooks per day
  - Spotter reports (~1,000 reporters)
  - 18 districts, eastern U.S.; 4 parts in each

John Park Finley
1854-1943

(from Galway, 1985; BAMS)

LIEUTENANT JOHN P. FINLEY, SIGNAL CORPS, UNITED STATES ARMY.



The Finley example

- Finley forecasts

	Obs. Yes	Obs. No	Sum
Fcst.Yes	28	72	100
Fcst. No	23	2680	2703
Sum	51	2752	2803

- 96.6% accurate
- "Accuracy" if no tornado forecasts issued: 98.2%

The Finley example: Outcomes

- The first (?) scientific discussion of verification
- Numerous verification measures developed (e.g., Equitable Threat Score, Heidke skill score)
- Many issues raised
  - Definition of forecast “event”
  - Quality of observations
  - Baselines of no skill
  - “Dimensionality” of the verification problem
  - Specifying purpose of verification
  - Use and value of forecasts
  - Asymmetric costs of misclassification

Verification evolution

Measures-based approach (traditional)

- Summarize forecast quality using one or more standard measures
  - ***Continuous***: Mean Absolute Error (MAE), Mean Error (ME), Root Mean Squared Error (RMSE), S1, Anomaly Correlation
  - ***Dichotomous or categorical***: POD, FAR, CSI, ETS
  - ***Probability***: Reliability, Brier skill score
  - ***Skill scores***: Comparative verification
- Heidke score (1930s)
- Brier score (1950s)
- Focus on upper air (e.g., 500 mb ht) and a few other variables (e.g., T, PoP)

Purposes of verification

Administrative

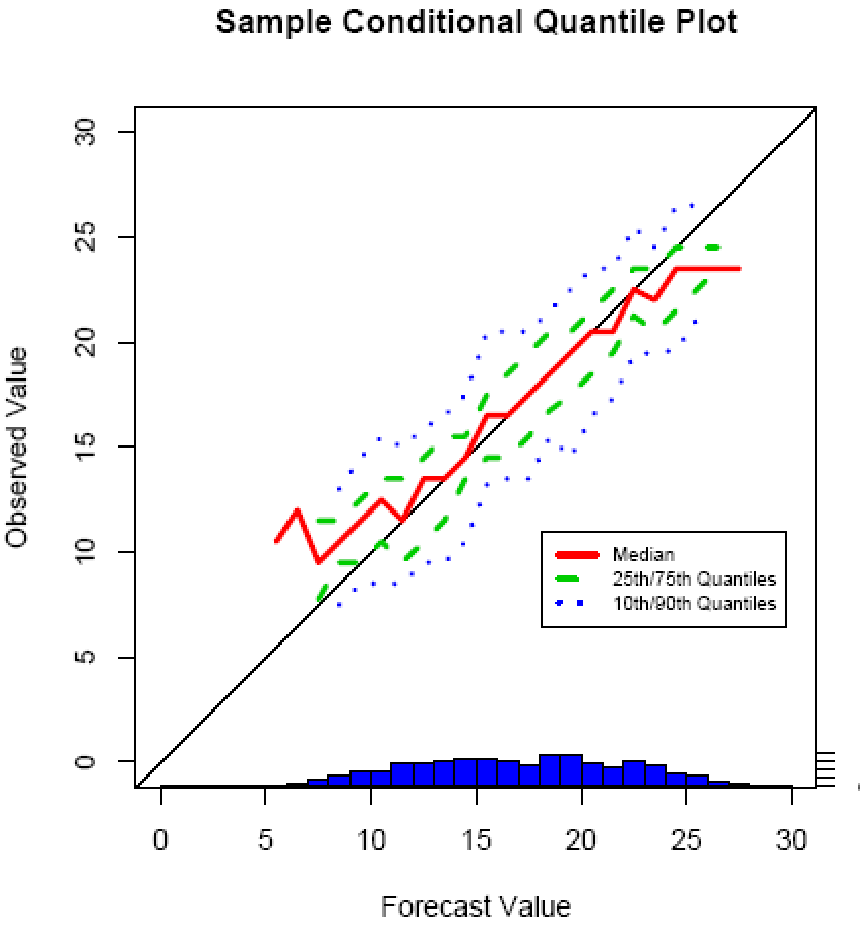
Scientific

Economic

From Brier and Allen, 1951

Verification evolution

- "Meta-verification": Propriety, Equitability
- Distributions-oriented approach (Murphy and Winkler; 1980s)
  - Joint distribution of forecasts and observations
  - Underlying conditional and marginal distributions
- Advent of "diagnostic" verification



Note from Murphy (1993):

“p(f,x) contains all of the non-time-dependent information relevant to evaluating forecast quality”

One could imagine extending this joint distribution to *p(f,x,t)*…

…and that Murphy considered this possibility.

Issue: Dimensionality

Current situation

- Most of the scores used today are the same as the ones developed in the Finley period
- Operational verification typically focuses on “management” needs and model-centric applications of verification
  - i.e., not on diagnostic or user-focused approaches
- Typical focus is on
  - A few traditional <u>measures</u>
  - Aggregated <u>statistics</u>
  - A few <u>parameters</u> (e.g., 500 mb ht, T, PoP)

Current situation cont.

- Model verification “drives” choices in model parameterizations, development, etc.

  Ex: verification of models using RMSE applied to 500 mb heights

- Uncertainty in verification measures is not commonly estimated in operational verification
- Forecast use/value is rarely considered

Current situation cont.

But… recent advances and efforts represent progress and growth

New focus on…

- Spatial methods
- Confidence intervals and hypothesis tests
- Ensemble and other methods
- "High impact" weather

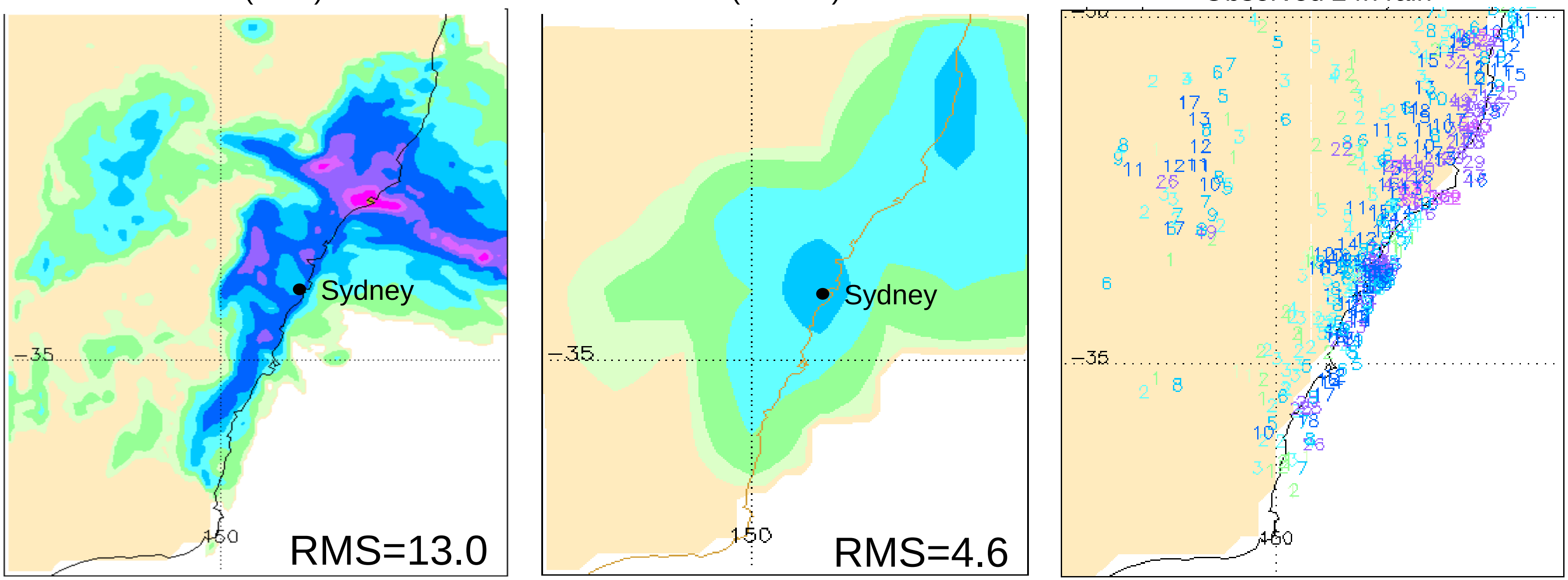
Driven by…

- High resolution forecasts
- New remotely sensed observations
- New types of forecasts (e.g., ensembles, hurricanes)
- Community efforts

Challenges of high resolution forecasts

Which rain forecast would you rather use?
Which is a "better" forecast?

Mesoscale model (5 km) 21 Mar 2004 | Global model (100 km) 21 Mar 2004 | Observed 24h rain

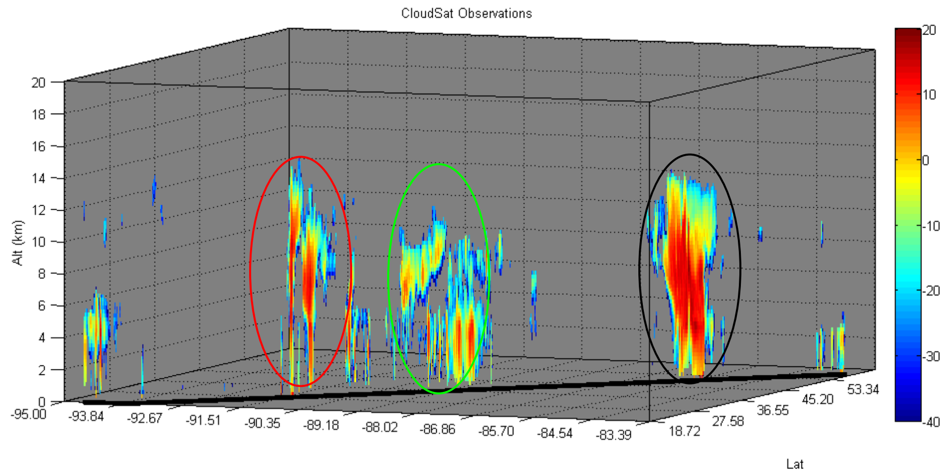


From E. Ebert

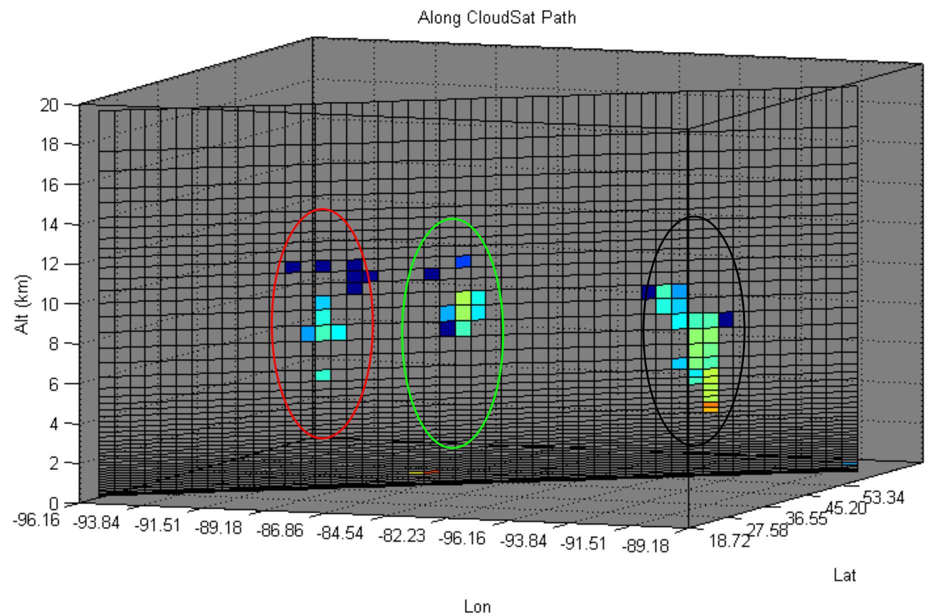
"Smooth" forecasts generally "Win" according to traditional verification approaches.

Cloud-Sat radar vs. RUC model output

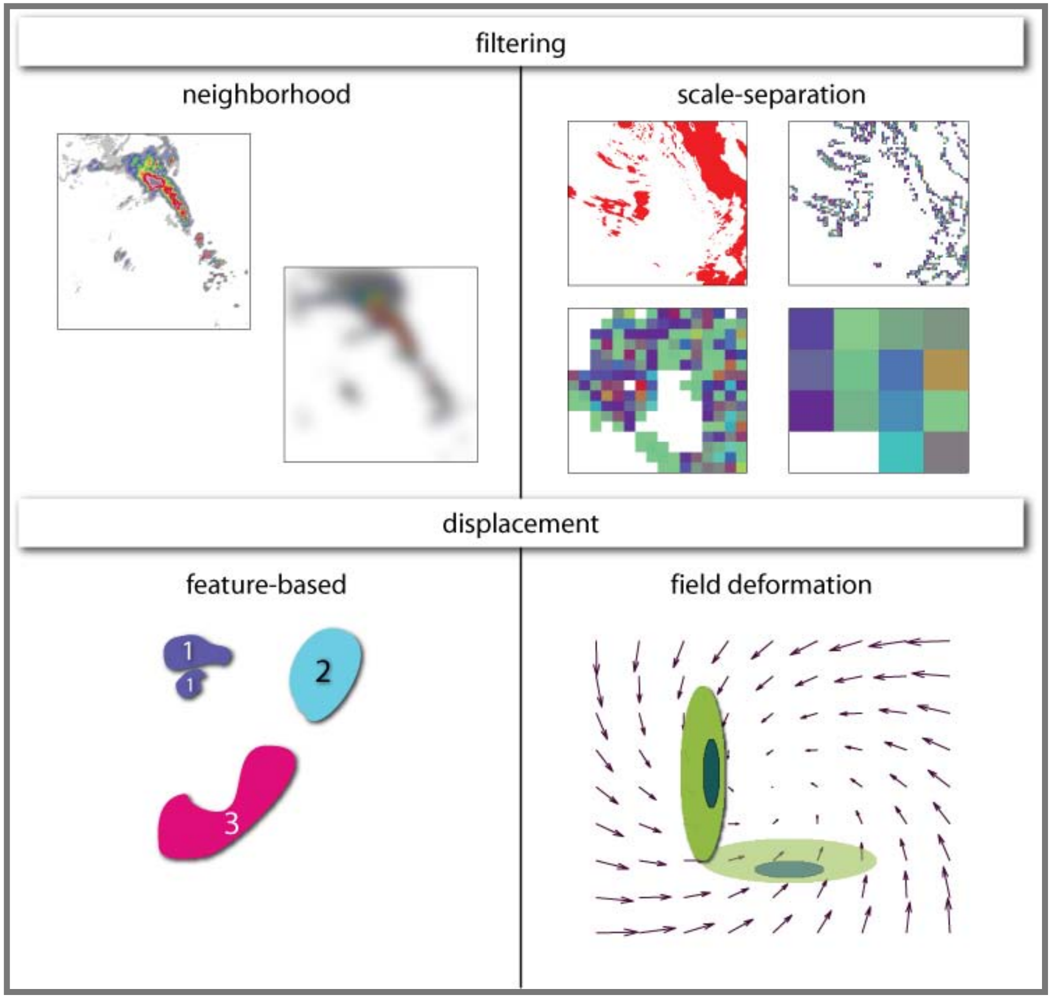
CPR
reflectivity

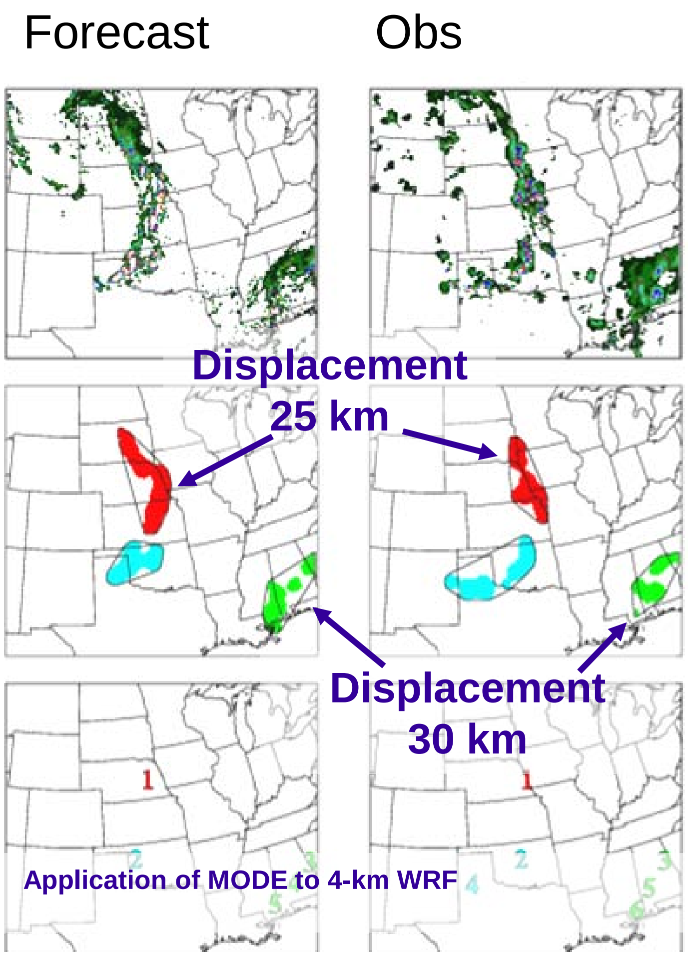


RUC
reflectivity

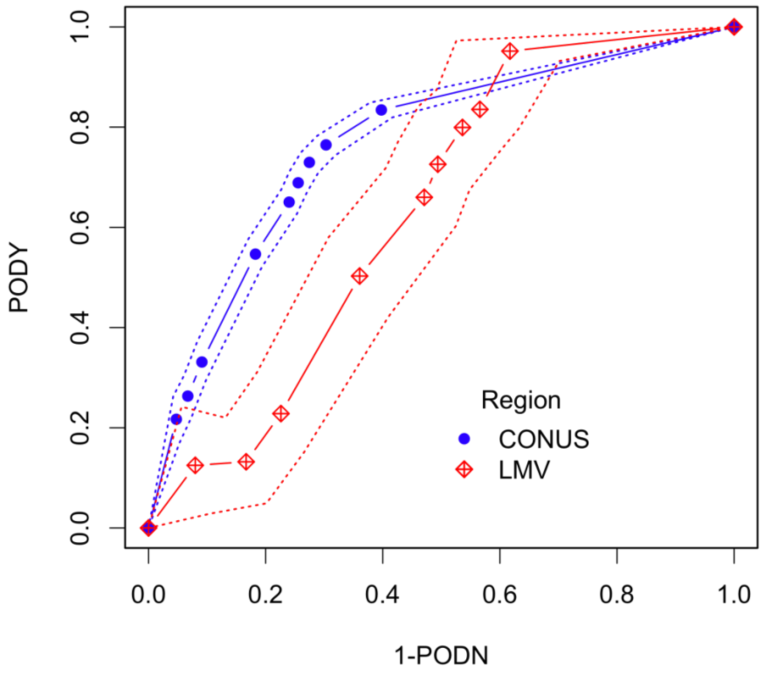


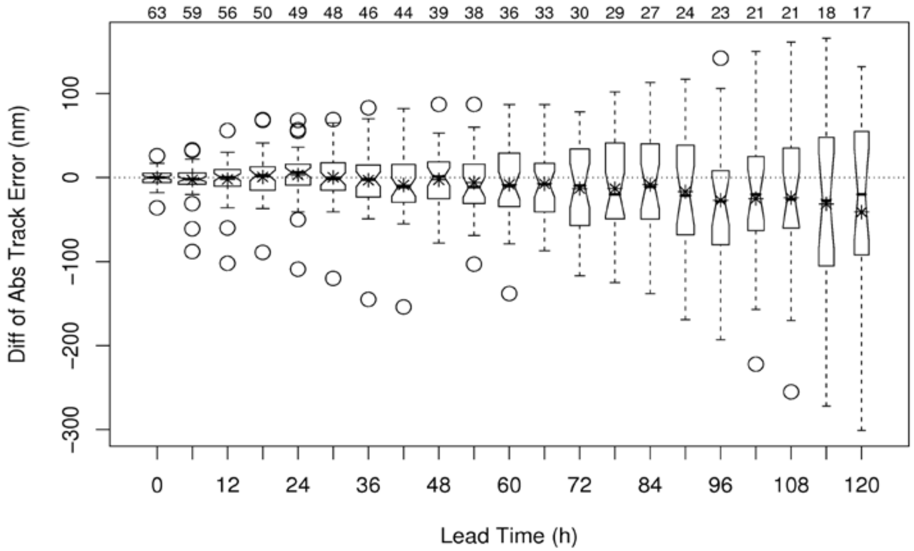
Spatial verification methods





Uncertainty in verification measures

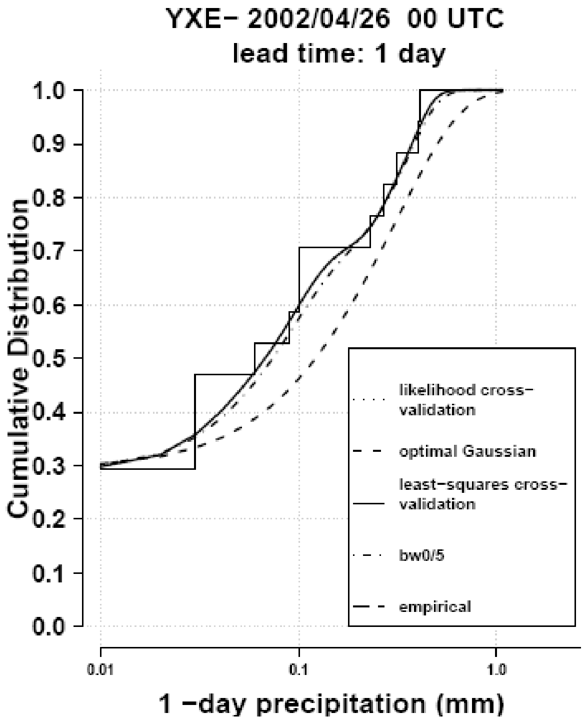




Confidence intervals and hypothesis tests are becoming more commonly used, particularly in research

New forecast types

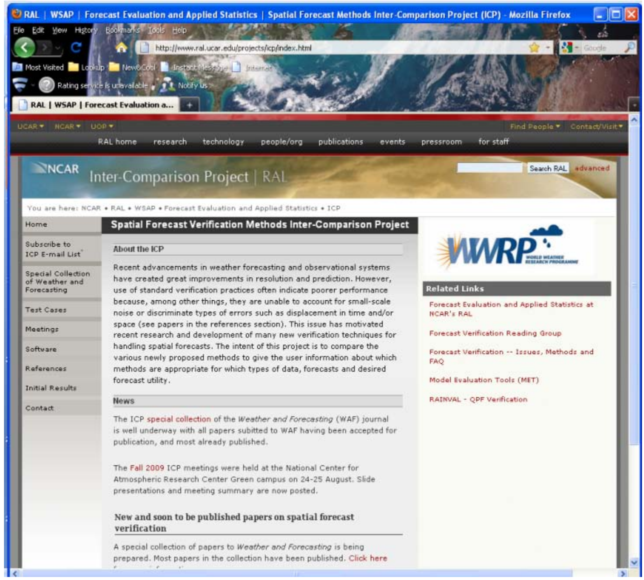
- Many kinds of forecasts require specialized verification approaches
- *Examples*:
  - Ensembles – representing a distribution
  - Hurricanes – multiple related variables
  - Hurricane ensembles



From Wilson 2009

Verification Community

- Rapidly expanding international community developed over last 10 years
- Workshops
- Tutorials
- Intercomparison project(s)
- Books, Chapters, Journal articles



ASP Colloquium, 2010

NCAR-ASP Summer Colloquium on Forecast Verification in the Atmospheric Sciences, June 6-18, 2010, Boulder, Colorado, USA

Purposes of verification

Administrative

Scientific

Economic

From Brier and Allen, 1951

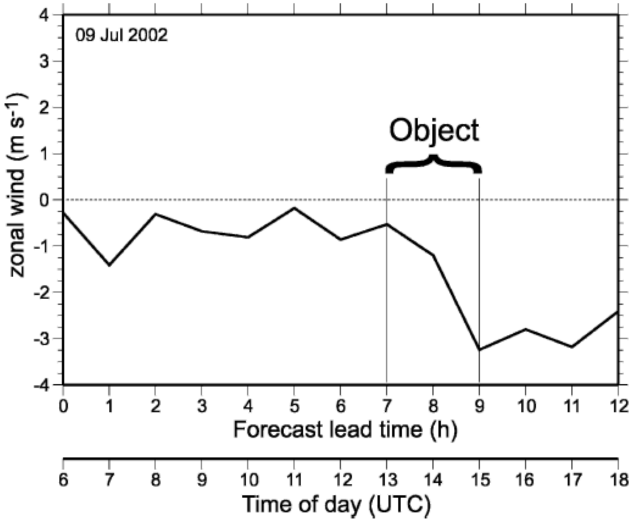
Verification challenges

Note: Some of our recent accomplishments remain our greatest challenges…

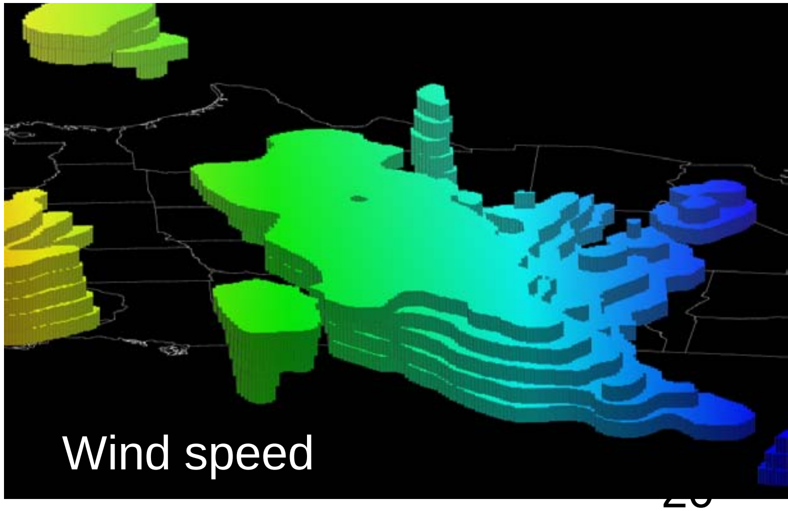
- New forecast types
- Verification of extremes and rare events (or "high impact weather")
- Quantification of uncertainty in verification measures *and* in verification observations
- Evaluation of high-resolution forecasts
- Use of remotely sensed observations
- Methods for spatial forecasts
- User-focused verification
- Verification across time

Examples of "timing" questions

- Does the model / forecast correctly capture the timing of a maximum or minimum value?
- Are forecasts issued at different times *consistent* with each other?
- Does the forecast capture changes in temperature, pressure, wind, etc. with time?
- How can we characterize and evaluate the combination of temporal and spatial features?



Rife et al., *WAF* 2005



Using temporal and spatial uncertainty information

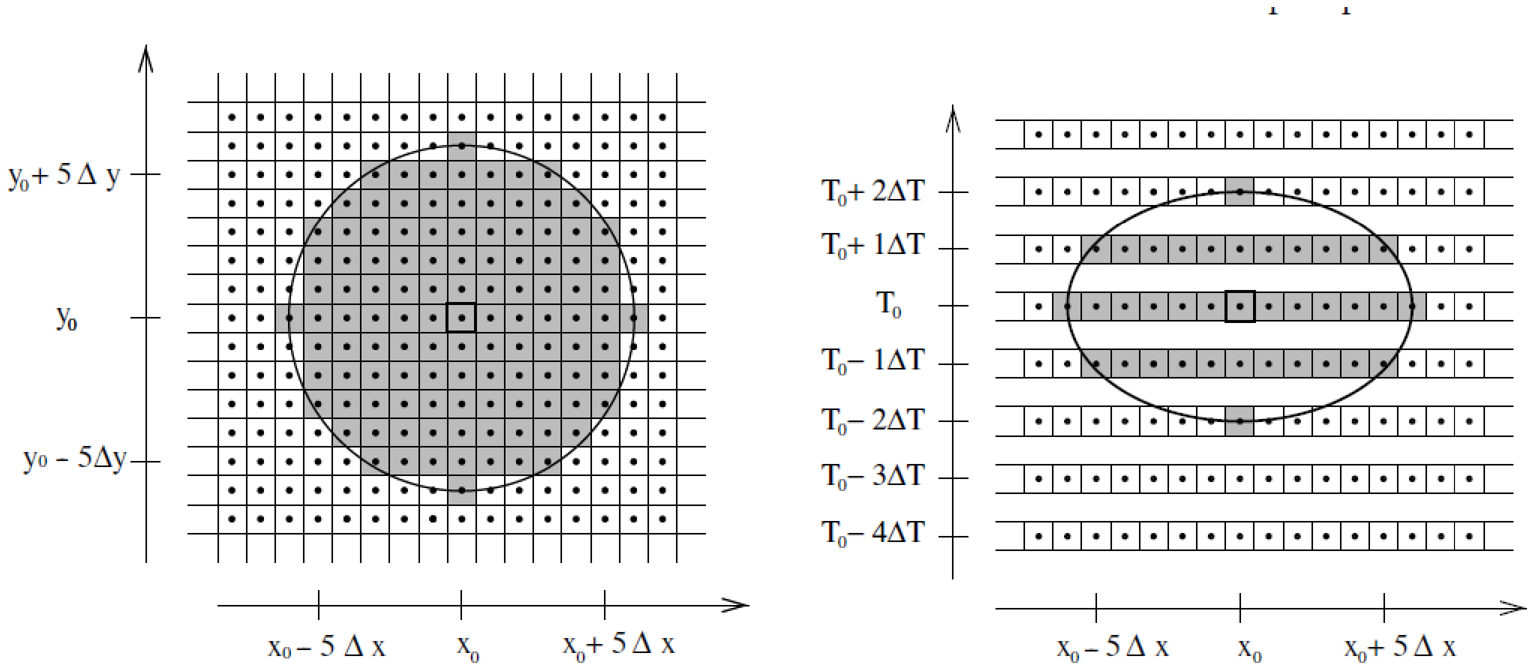


Figure 1. *Example of a spatio-temporal neighbourhood of a given grid point at location (x_0, y_0) and forecast lead time T_0. Left: the spatial neighbourhood in the (x, y)-plane. Right: the spatio-temporal neighbourhood in the (x, T)-plane. Δx and Δy denote the size of a grid box and ΔT denotes the time step between successive model output times. Shaded grid boxes belong to the neighbourhood.*

Theis et al. 2005

What is the current state of verification?

ARKANSAS

ℸ} {∫(((λ{}(}↕

(Just kidding…)

The current state of verification is…

Exciting…

Dynamic…

Challenging

Full of opportunities for

- Collaborations
- Research
- Applications
- Maybe a verification testbed?