Spatial and Object-Oriented Verification

Barbara Brown (NCAR)
Beth Ebert (BMRC)

International Verification Methods Workshop
Montreal, Quebec, Canada
16 September 2004

Outline

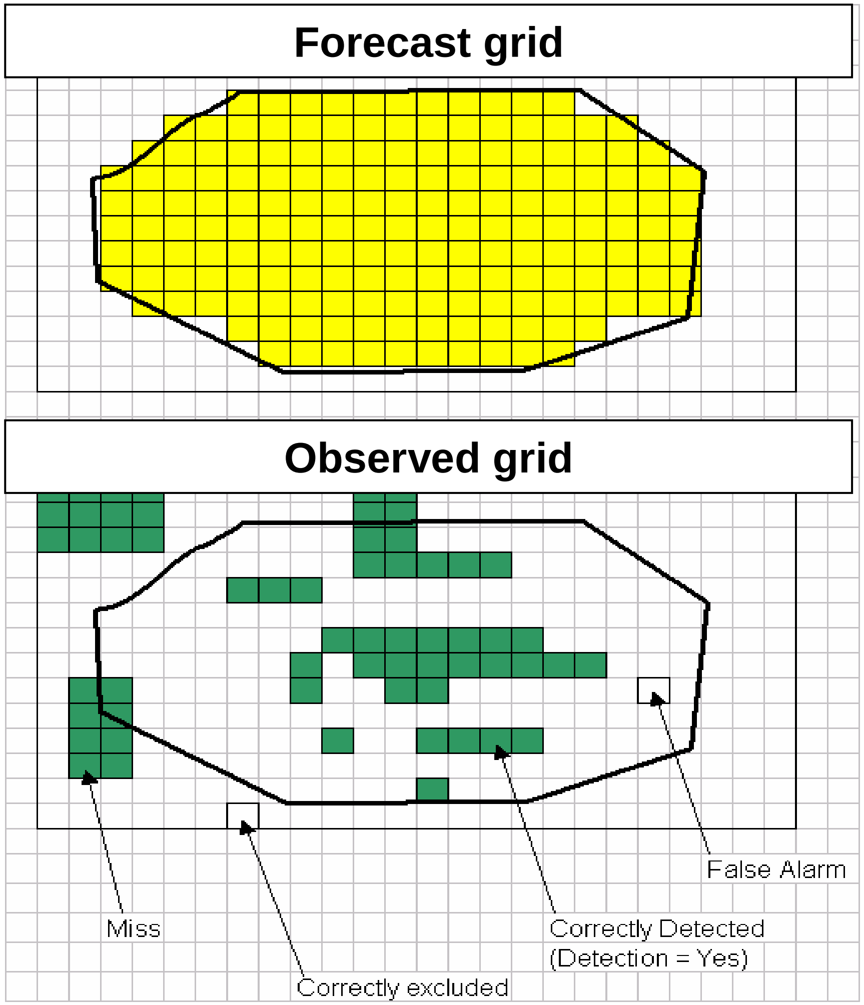
- Motivation
- Traditional approach
- Challenges and issues – high resolution forecasts (Beth)
- Goals of alternative approaches (Beth)
- Emerging approaches (Beth)
- Object-oriented and composite approaches (Barb)
- Prospects for the future

Aggregation/Stratification

- Aggregation
  - Across time?
    - Results for each grid point/location
  - Space?
    - Results for each time
  - Space and time?
    - Results summarized across spatial region and across time
- Stratification
  - By grid point/forecast? Region? Time period (e.g., according to diurnal variation)?

Matching Forecasts and Observations

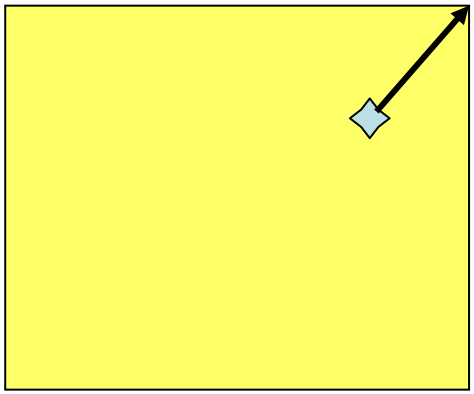
- Matching approach depends on
  - Nature of forecasts and observations
    - Scale
    - Consistency
    - Sparseness
  - Other matching criteria
    - Verification goals
    - Use of forecasts
- Grid to Grid approach
  - Overlay forecast and observed grids
  - Match each forecast and observation

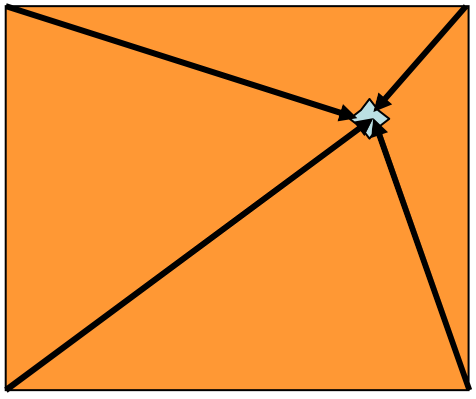


Matching forecasts and observations

- Point-to-Grid and
  Grid-to-Point

- Matching approach can impact the results of the verification





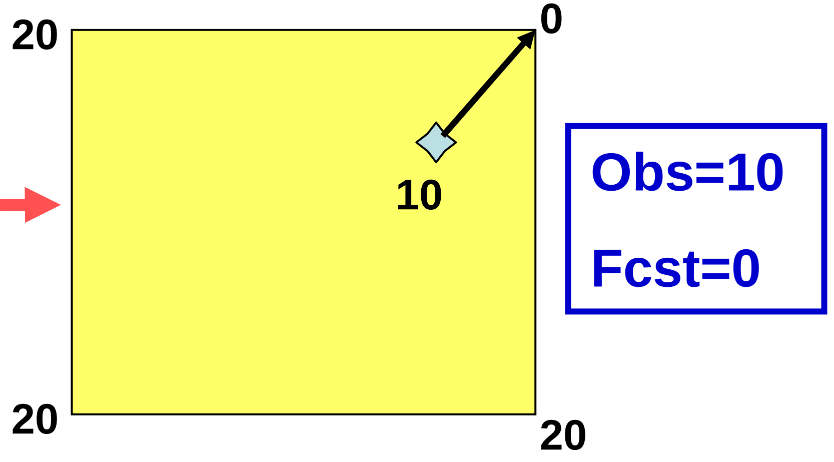
Matching forecasts and observations

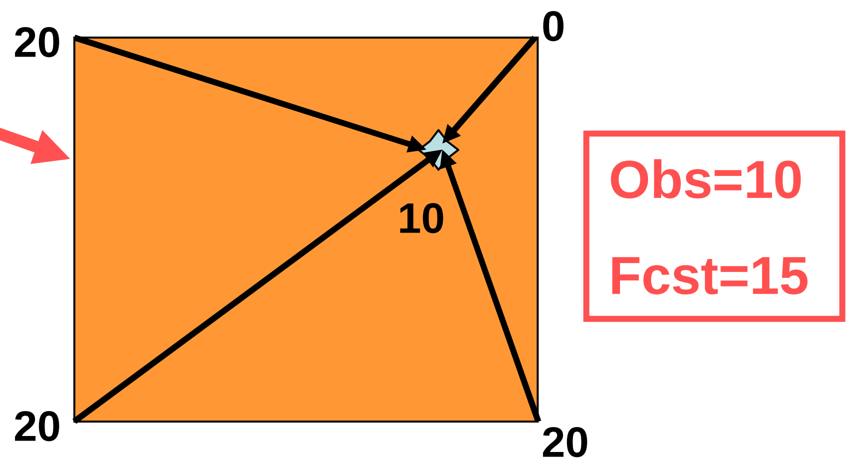
Example:

- Two approaches:
  - Match rain gauge to nearest gridpoint

  or

  - Interpolate grid values to rain gauge location
    - Crude assumption: equal weight to each gridpoint





- Differences in results associated with matching:

  "representativeness" error

The "good" news: Representativeness error generally is smaller than the forecast errors

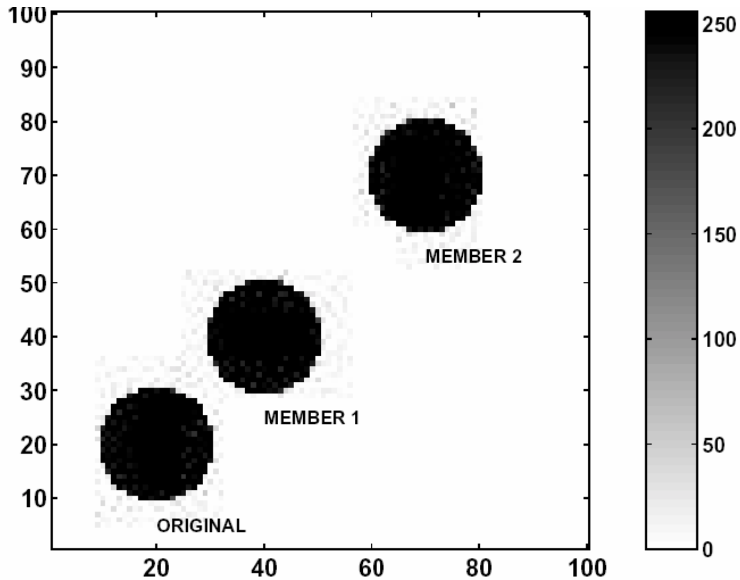
Traditional verification approaches

Compute statistics on forecast-observation pairs

- Continuous values (e.g., precip amount, temperature, NWP variables):
  - MSE, ME, Correlation
  - S1 score, Anomaly correlation
  - Others (e.g., new score under development by Venugopal et al. 2004)
- Categorical values (e.g., precip occurrence):
  - Contingency table statistics (POD, FAR, Heidke skill score, Equitable threat score, Hanssen-Kuipers statistic)

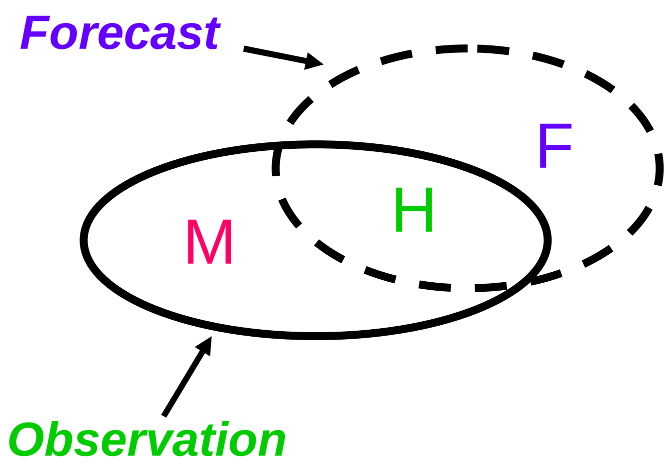
Venugopal et al. approach for grid comparisons

- Ratio of an image distance measure to a magnitude difference metric
  - *Distance measure*: Modified Hausdorff distance
  - *Magnitude metric*: Based on means, standard deviations, covariances of precipitation field
- May be more sensitive to true differences between images than other measures (e.g., RMSE, ETS)



	RMSE	EqTh	FQI
Original and Member 1	68.41	-0.02	0.3924
Original and Member 2	68.41	-0.02	1.15

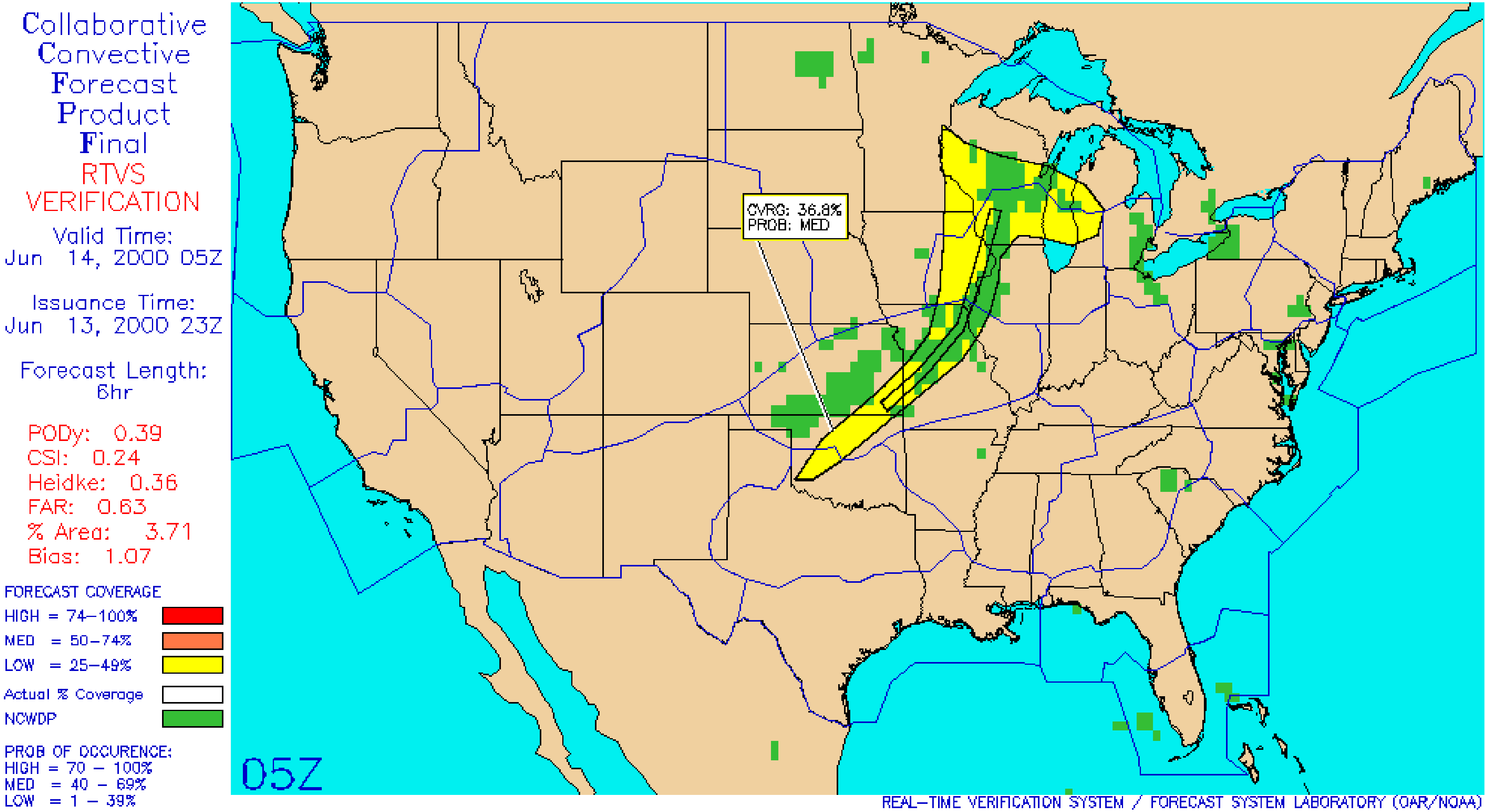
“Standard” Verification Measures (Yes/No forecasts)



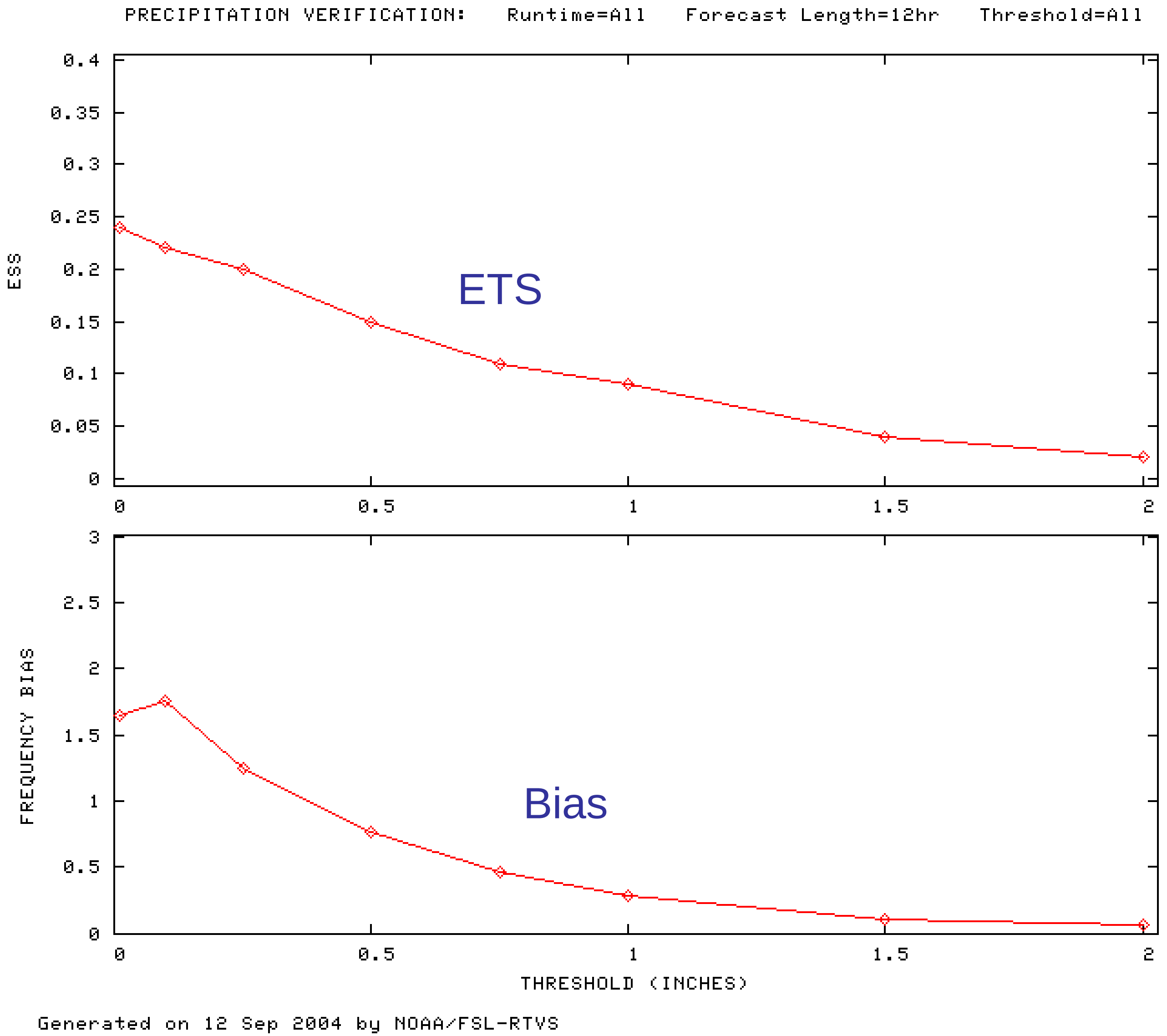
- H = Hits
- M = Misses
- F = False Alarms
- POD = H / (H + M)
- POFD = proportion of “No” area that was correctly forecast to be “No”
- FAR = F / (H + F)
- Bias = (F + H) / (M + H)

- **CSI = H / (M + H + F)**
  Measures “relative accuracy”
- **H-K = POD + POFD -1**
  Measures “discrimination” between Yes and No observations
- **POD (PODy)**
  Measures proportion of observed area that is correctly forecast to be “Yes”
- **POFD (PODn)**
  Measures proportion of area that is correctly forecast to **be “No”**
- **FAR**
  Measures proportion of forecast convective area that is incorrect
- **Bias**
  Measures the extent of over- or under-forecasting
- **Skill scores (Heidke, Gilbert/ETS)**
  Measure the improvement in percent correct and CSI, respectively over what’s expected by chance

PODy=0.39, FAR=0.63, CSI=0.24



Example: Model precipitation



Eta 12-h forecasts

Continental U.S.

(From NOAA/FSL Real-time verification System, RTVS)

What’s missing?

- Traditional approaches provide overall measures of skill… BUT
- They provide minimal *diagnostic* information about the forecast:
  - What went wrong?
  - What went right?
  - How can I improve this forecast?
  - How can I use it to make a decision?
- Best performance for *smooth* forecasts
- Insensitive to the ***size*** of the errors…

What makes a good spatial forecast, anyway?

For spatial forecasts we tend to focus on *features* on maps. In a good spatial forecast, the feature* will:

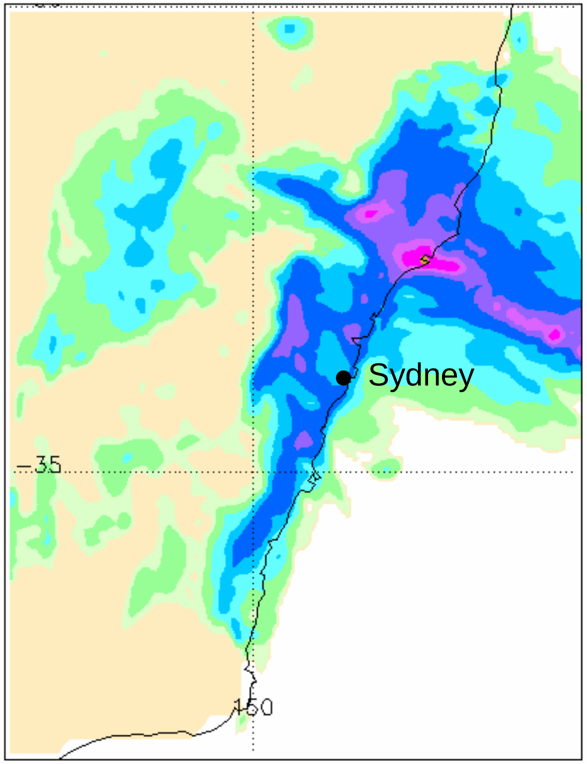
- be predicted to occur
- be located in the right place
- have the correct amplitude
- be about the right size
- have the right shape
- have appropriate spatial variability and structure(s)

*Depending on the spatial and temporal scales of interest, the feature may actually be a group of weather elements (thunderstorms, for example)

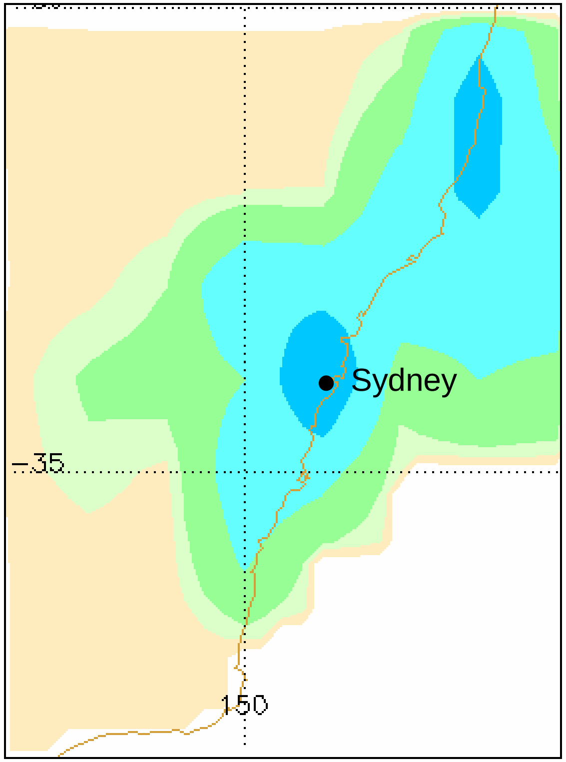
Challenges and issues – high resolution forecasts

Which rain forecast would you rather use?

Mesoscale model (5 km) 21 Mar 2004



Global model (100 km) 21 Mar 2004



Challenges and issues – high resolution forecasts

High resolution forecasts give detailed spatial and temporal structures

Positive impact on usefulness of forecast

- More realistic representation of convection
- Better representation of topographic effects and diurnal circulations
- Better definition of frontal zones
- Local maxima and minima better resolved

Negative impact on standard verification scores

- Amplitude errors magnified compared to low-res forecasts
- Small spatial or temporal offsets likely to count as misses and false alarms

Position / timing errors

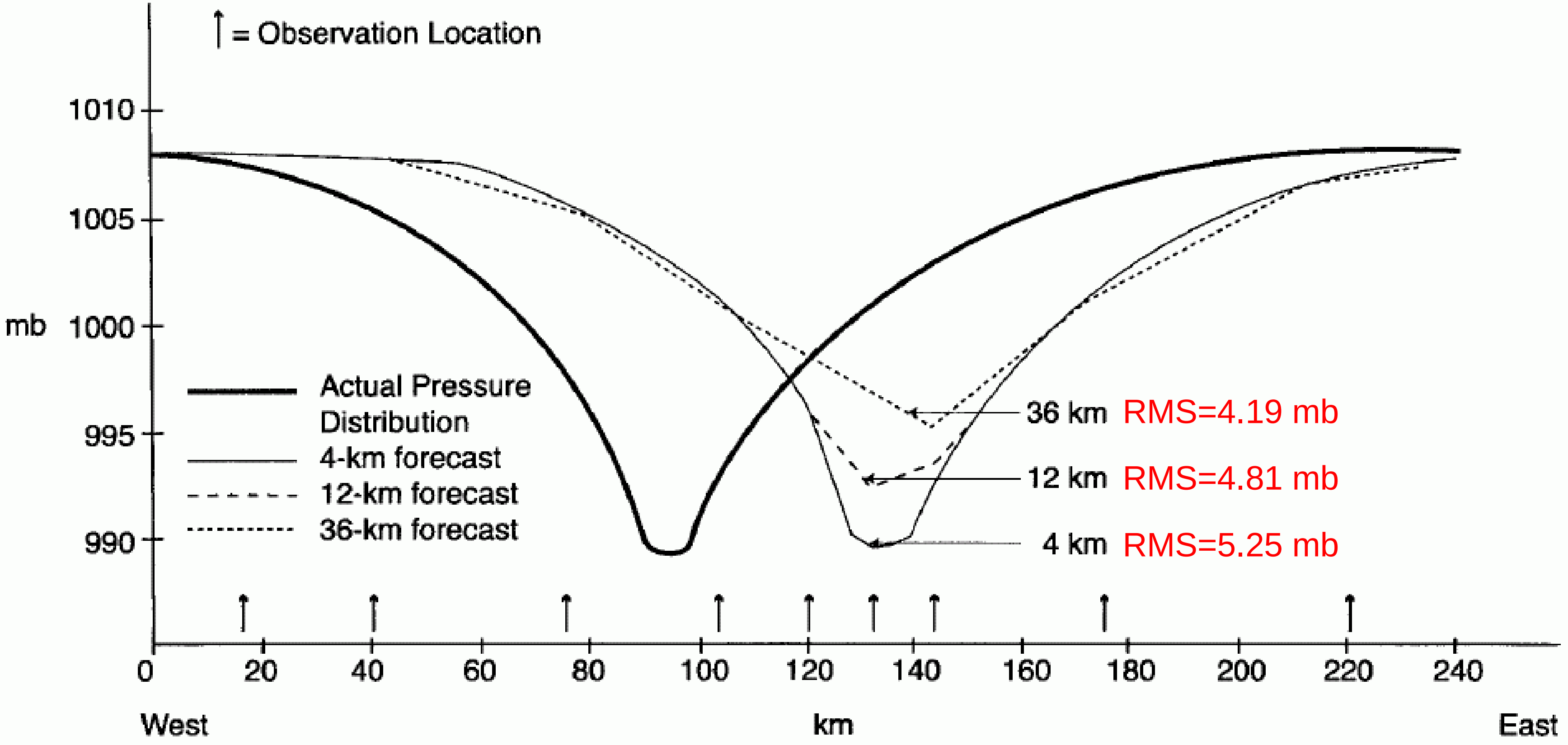


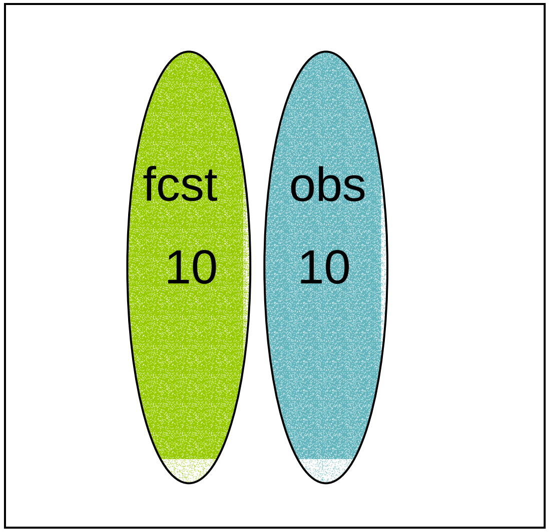
Fig. 16. Sea level pressure variation in the east–west direction of an observed trough (solid line), and forecast troughs from 36-, 12-, and 4-km domains. The vertical arrows indicate hypothetical observation locations. (Mass et al., *BAMS*, 2002)

More realistic structures in high resolution forecasts get penalized more if there are positional or timing errors

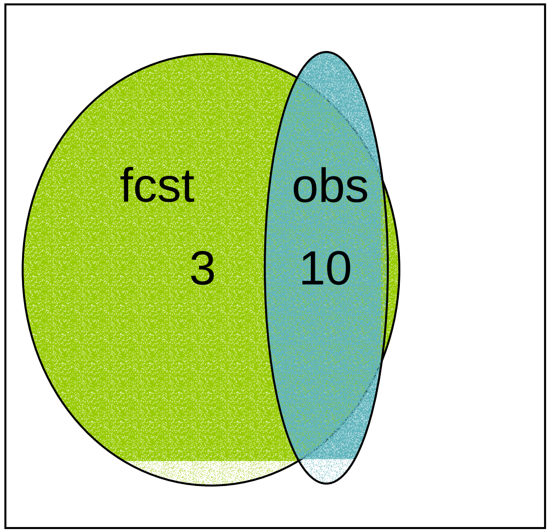
"Double penalty"

Event predicted where it did not occur,
no event predicted where it did occur

Ex: Two rain forecasts giving the same volume



High resolution forecast
RMS ~ 4.7
POD=0, FAR=1, TS=0



Low resolution forecast
RMS ~ 2.7
POD~1, FAR~0.7, TS~0.3

Verification philosophies for high resolution forecasts

Need we get the forecast *exactly* right? Often close enough is good enough...

YES

- Topographically influenced weather (valley winds, orographic rain, etc.)
- Hydrological applications (e.g. flash floods)
- High stakes situations (e.g. space shuttle launch, hurricane landfall)

↓

Standard verification methods

NO

- Guidance for weather forecasters
- Model *validation* (does it predict what we expect it to predict?)
- Observations may not allow standard verification of high resolution forecasts

↓

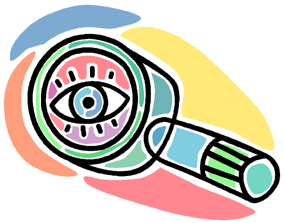
Diagnostic methods

verify attributes of forecast

Goal of *diagnostic* verification – understanding the sources of the errors

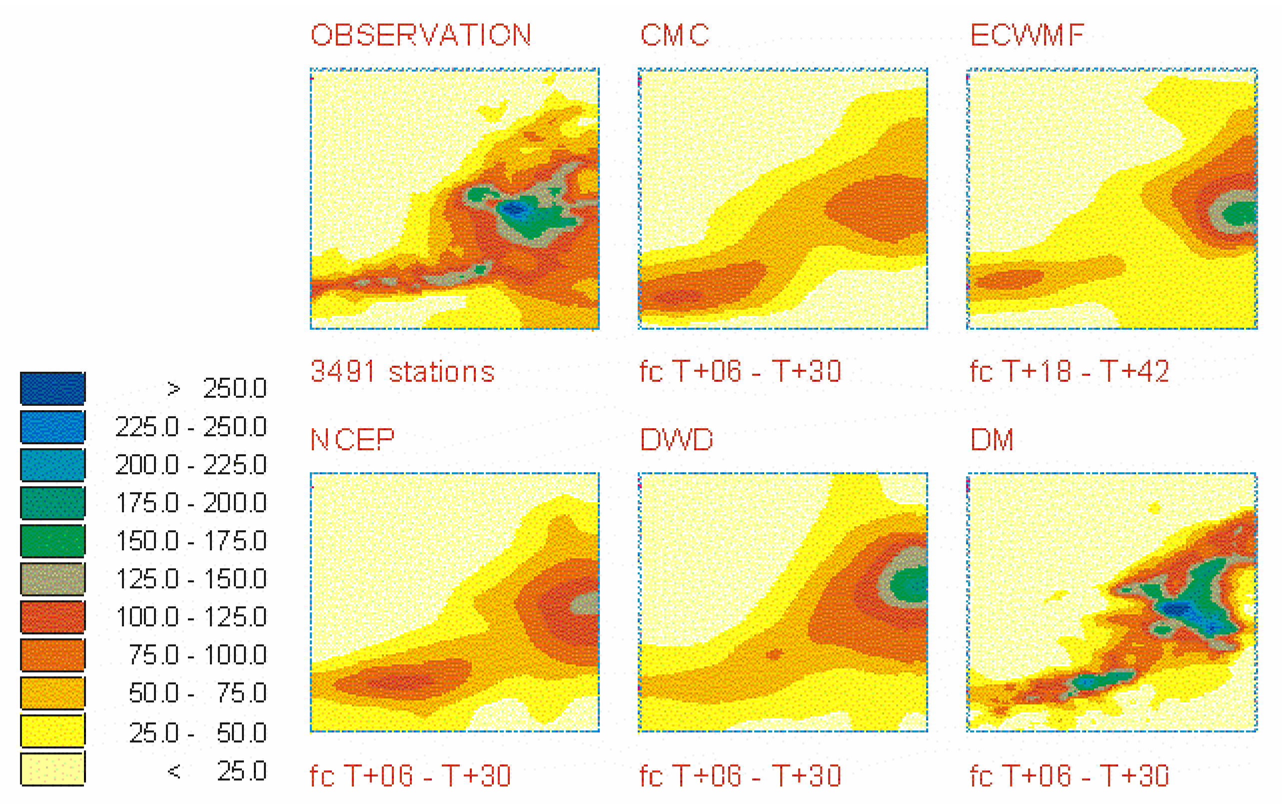
- Advanced diagnostic methods help quantify errors in:
  - Occurrence
  - Location
  - Amplitude
  - Size
  - Shape
  - Variability
- Can we determine the relative importance of these sources of error?
  - Error decomposition methods
  - Scale separation methods

Visual ("eyeball") verification



Oldest, perhaps most informative, method

but not quantitative...



Practically perfect hindcasts
(Brooks & Kay)

Rare events are, by nature, difficult to predict accurately **and** difficult to observe accurately. Ex: tornadoes

Forecasts are often given as watch or warning *regions*

Standard verification doesn't work well

- Most of warning region is false alarm, even when the event occurs in the region
- Observations of non-events generally not reported
- Persistence or climatology aren't really appropriate reference forecasts for skill scores

Practically perfect hindcasts
(Brooks & Kay)

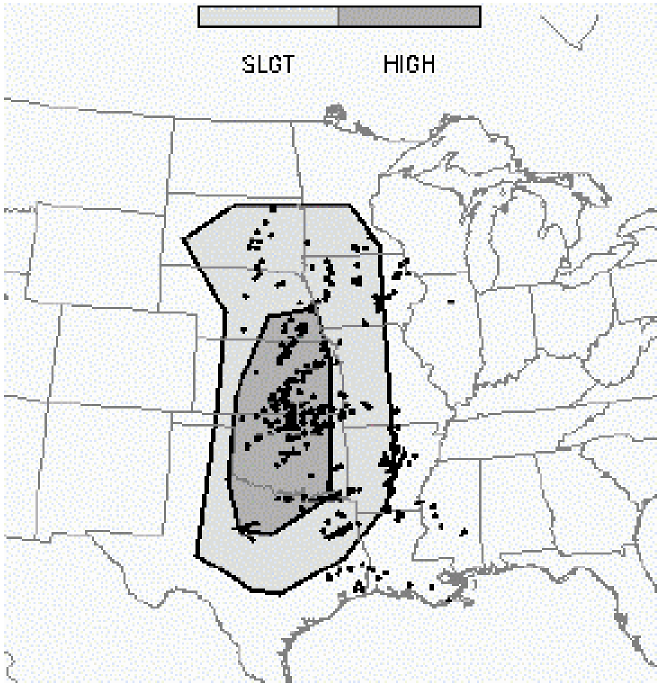
PP approach: If the forecaster had all of the observations in advance, what would the "practically perfect" forecast look like? Apply a smoothing function to the observations to get probability contours, choose an appropriate yes/no threshold

→ Did the actual forecast look like the practically perfect forecast?

How would the practically perfect forecast score when verified against the observations?

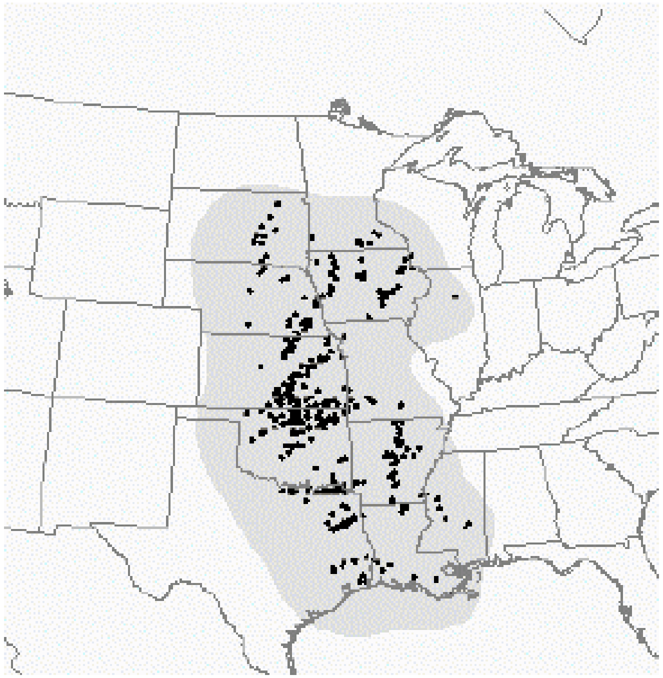
→ How did the verification score for the actual forecast compare to the score for the practically perfect forecast?

Practically perfect hindcasts
(Brooks & Kay)



SPC convective outlook

Threat score=0.34



Practically Perfect forecast

Threat score=0.48

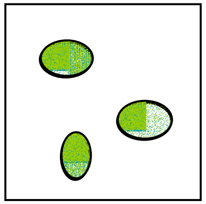
Convective outlook is 75% of the way to being "practically perfect"

Fuzzy verification (Damrath)

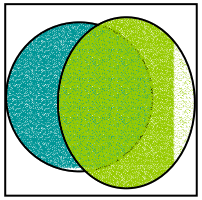
For a given region of interest, when is a forecast useful?

Some possible decision criteria:

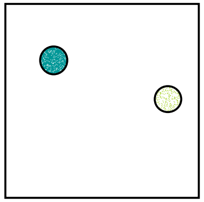
- If the event of interest is forecast and observed at precisely the same points



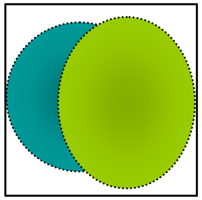
- If the event of interest is forecast and observed over a minimum fraction of the region of interest (50%?)



- If the event of interest is forecast and observed at any point in the region of interest

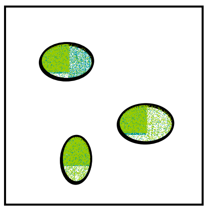


- If the event of interest is forecast and observed with certain frequencies P_f and P_o – how to use this information?

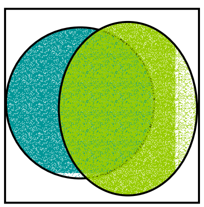


Fuzzy verification (Damrath)

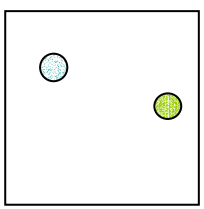
Contingency table elements depend on decision criterion:



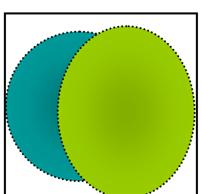
Agreement at every point in region



Minimum coverage X% of forecast and observations



Event forecast and observed at any point in region



Forecast and observed events with certain frequencies

traditional

Predicted \ Observed	yes	no
yes	hits	false alarms
no	misses	correct negatives

proportion

Predicted \ Observed	$P_o \geq X\%$	$P_o < X\%$
$P_f \geq X\%$	hits	false alarms
$P_f \geq X\%$	misses	correct negatives

fuzzy

Predicted \ Observed	P_o	$1\text{-}P_o$
P_f	$\min(P_o, P_f)$	$\min(1\text{-}P_o, P_f)$
$1\text{-}P_f$	$\min(P_o, 1\text{-}P_f)$	$\min(1\text{-}P_o, 1\text{-}P_f)$

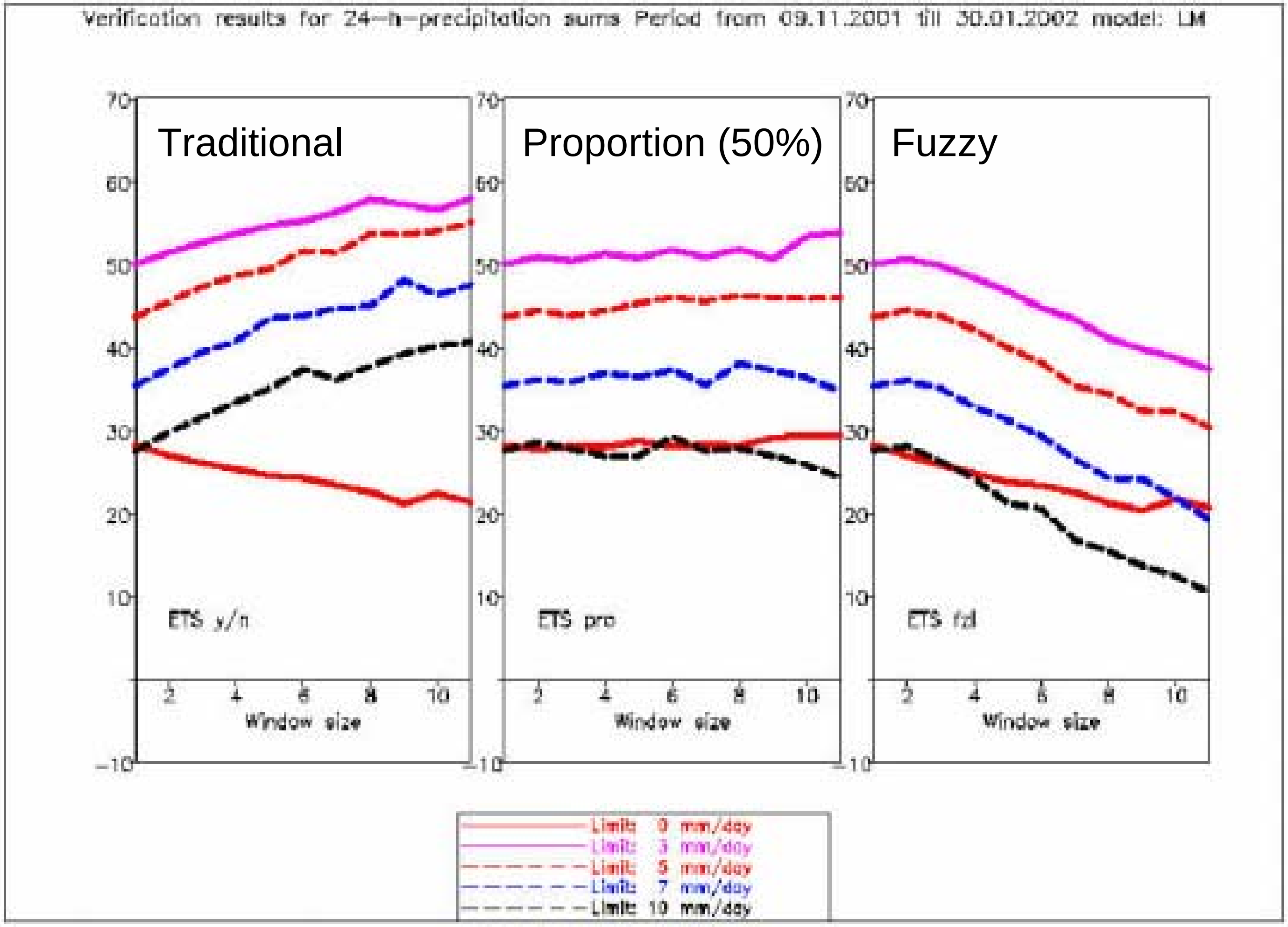
Forecasts partially correct, partially incorrect

Fuzzy verification (Damrath)

Results:

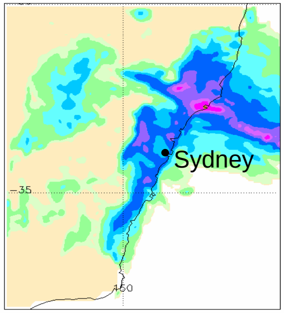
For *traditional* verification the scores improve as the window (grid) size increases (location errors less important) .

For *fuzzy* verification the scores tend to get poorer as the window (grid) size increases (forecasts and obs less certain).



Spatial multi-event contingency table (Atger, 2001)

Experienced forecasters interpret output from a high resolution deterministic forecast in a *probabilistic* way



← "high probability of some heavy rain near Sydney", ***not*** "62 mm of rain will fall in Sydney"

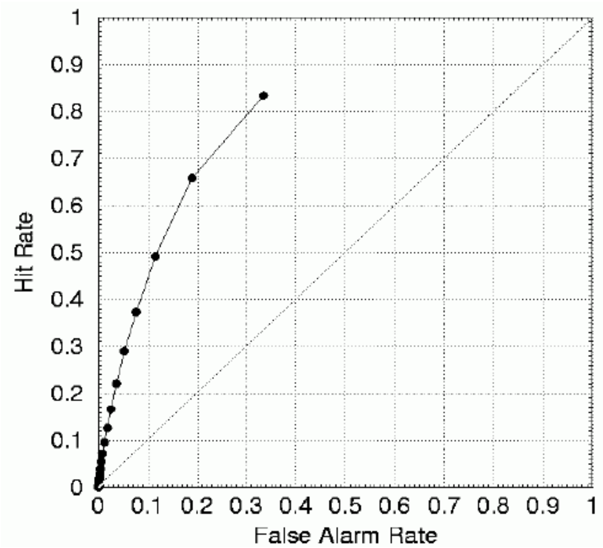
The deterministic forecast is mentally "calibrated" according to how "close" the forecast is to the place / time / magnitude of interest.

Very close → high probability
Not very close → low probability

Spatial multi-event contingency table (Atger, 2001)

Verify using the Relative Operating Characteristic (ROC)

Measures how well the forecast can separate events from non-events based on some decision threshold



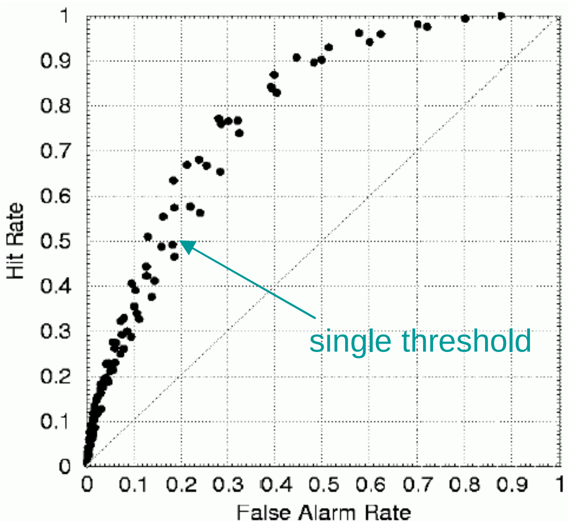
ROC curve for varying rain threshold

Decision thresholds to vary:

- magnitude (ex: 1 mm h^{-1} to 20 mm h^{-1})
- distance from point of interest (ex: within 10 km, ... , within 100 km)
- timing (ex: within 1 h, ... , within 12 h)
- anything else that may be important in interpreting the forecast

Spatial multi-event contingency table (Atger, 2001)

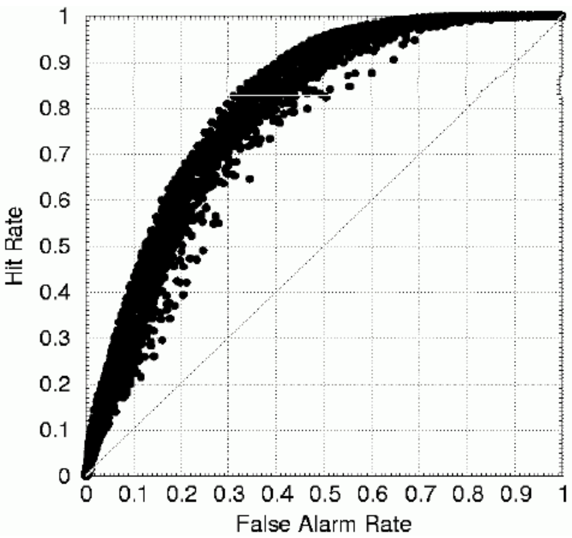
Vary more than one decision threshold



ROC curve for varying rain threshold and varying distance from point of interest

N = (# rain thresholds) x (# distance thresholds)

Can apply to ensembles, and to compare deterministic forecasts to ensemble forecasts

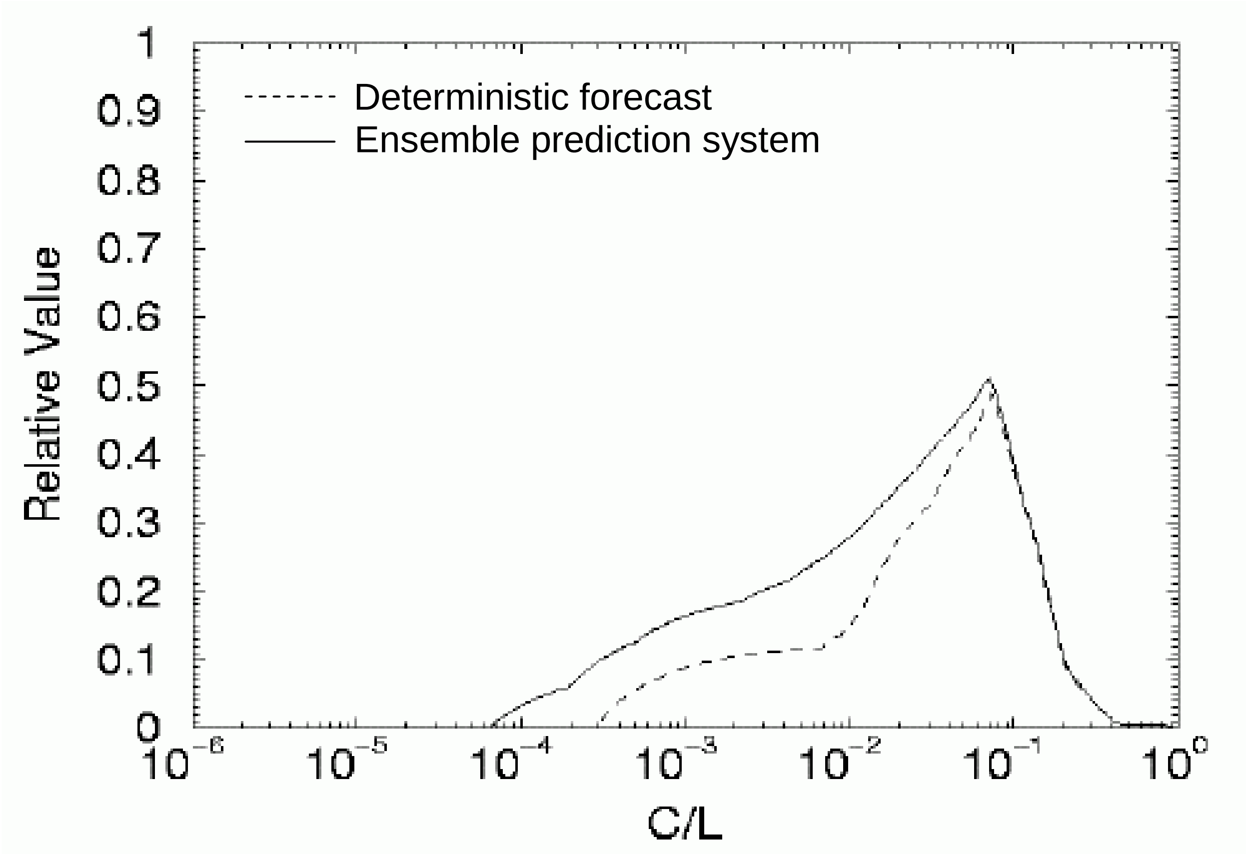


ROC curve for varying rain threshold and varying distance from point of interest, for an EPS

N = (# rain thresholds) x (# distance thresholds) x (# ensemble members)

Spatial multi-event contingency table (Atger, 2001)

Can also evaluate and compare forecasts in terms of Relative Value



Scale separation methods

Measure error as a function of spatial scale

Examples:

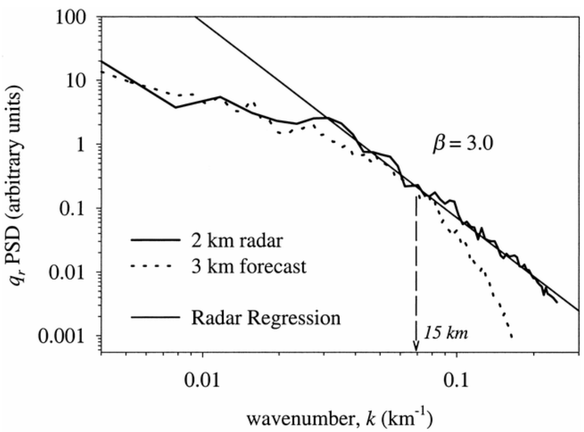
- Wavelet decomposition (Briggs and Levine, 1997; Alvera-Azcárate, 2004)
- Multiscale statistical properties (Zepeda-Arce et al., 2000; Harris et al., 2001)
- Discrete cosine transformation (Denis et al., 2002)
- Scale recursive estimation (Tustison et al., 2003)
- Intensity-scale verification approach (Casati et al., 2004)

→ Barbara Casati's talk this afternoon!

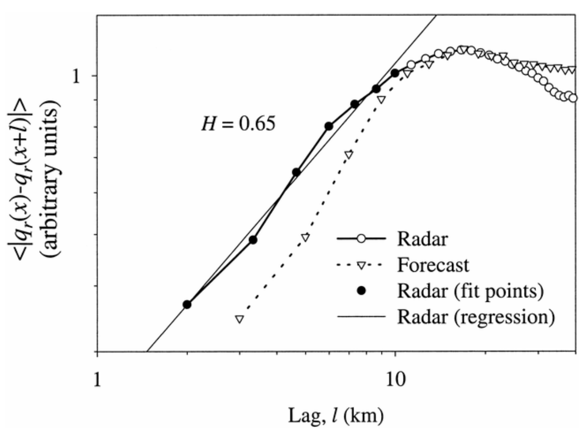
Multiscale statistical properties
(Harris et al., 2001)

Does a model produce the observed precipitation scale-dependent variability, i.e. does is look like real rain?

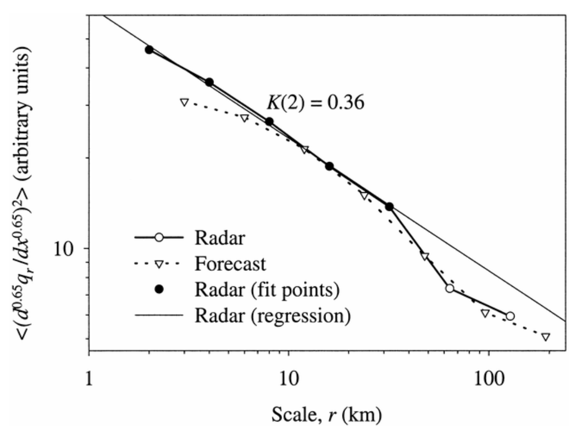
Compare multiscale statistics for model and radar data



Power spectrum



Structure function



Moment scaling

Feature calibration and alignment (Hoffman et al., 1995; Nehrkorn et al., 2003)

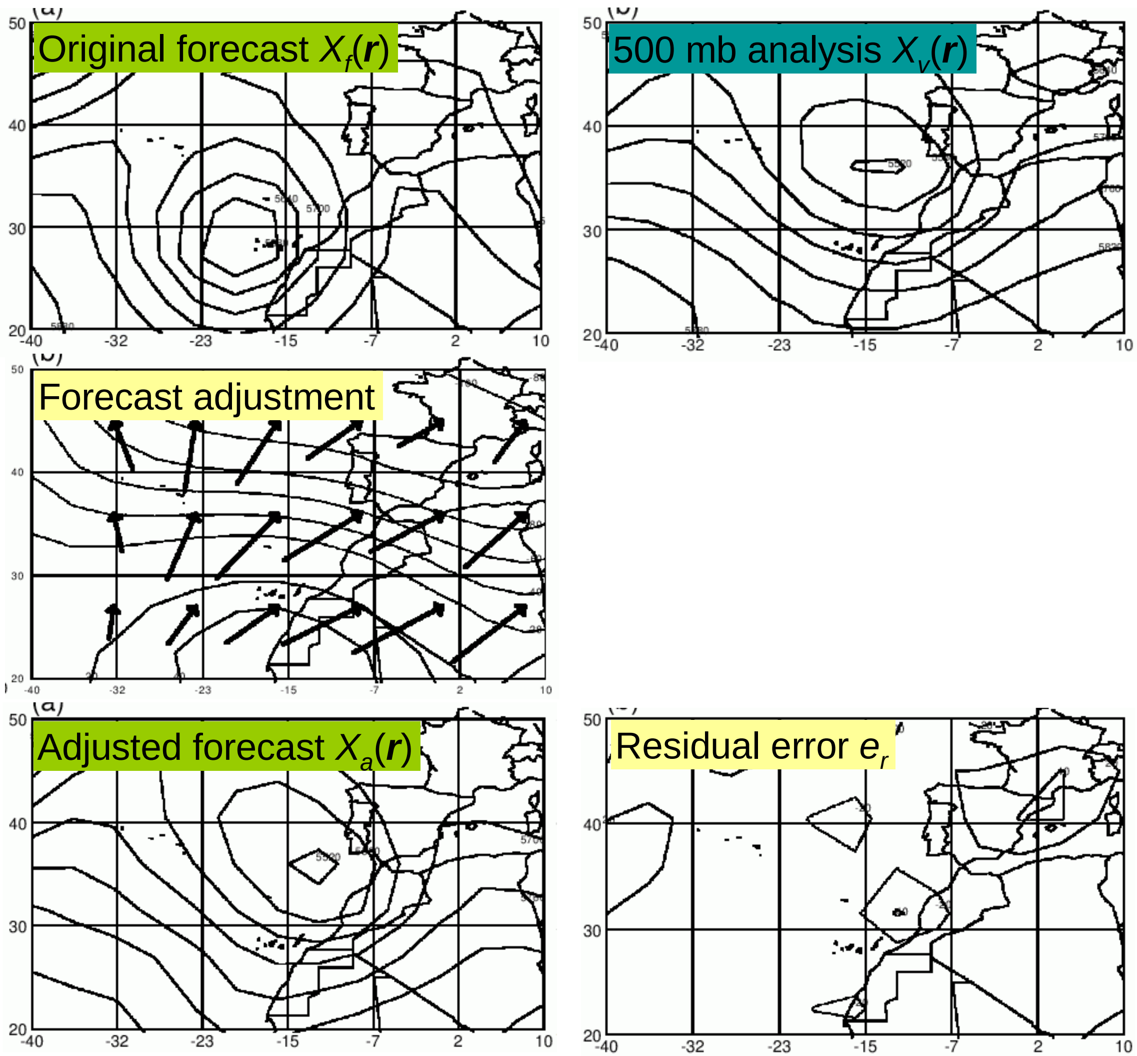
Subjective evaluation of a spatial forecast may suggest "front moved too fast in the model" or "forecast low too far to the south", etc.

Need for an objective evaluation method that accounts for phase and amplitude errors

Feature calibration and alignment (FCA) strategy uses pattern matching of forecast field to analysis field to determine fields of phase, amplitude, residual errors

- Hoffman et al. (1995) – maximize spatial correlation
- Nehrkorn et al. (2003) – spectral variational techniques

Feature calibration and alignment (Hoffman et al., 1995; Nehrkorn et al., 2003)



Error decomposition

$$e = X_f(\boldsymbol{r}) - X_v(\boldsymbol{r})$$

where $X_f(\boldsymbol{r})$ is the forecast, $X_v(\boldsymbol{r})$ is the verifying analysis, and $\boldsymbol{r}$ is the position.

$$e = e_p + e_b + e_r$$

where

$e_p = X_f(\boldsymbol{r}) - X_d(\boldsymbol{r})$ phase error

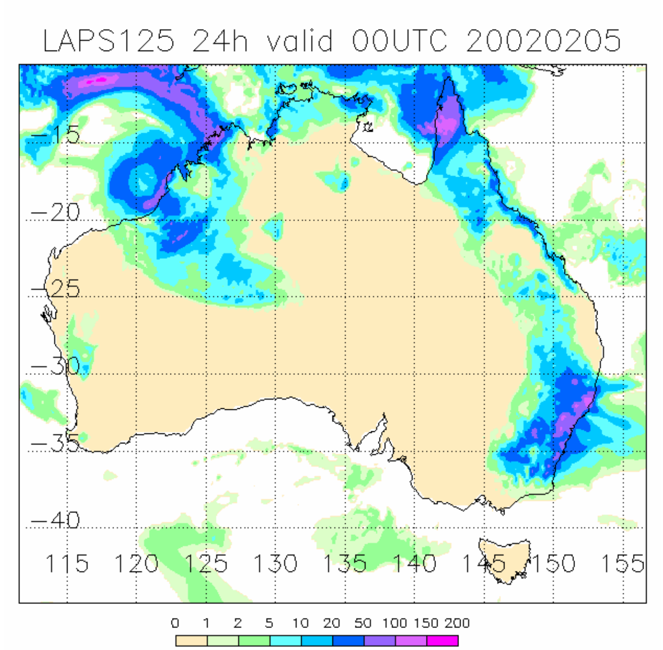
$e_b = X_d(\boldsymbol{r}) - X_a(\boldsymbol{r})$ local bias error

$e_r = X_a(\boldsymbol{r}) - X_v(\boldsymbol{r})$ residual error

Entity-based approach
(Ebert and McBride, 2000)

When one or more features of interest are present in a forecast the spatial verification does not separate the performance for one feature from the performance for another.

Example: National scale QPF verification

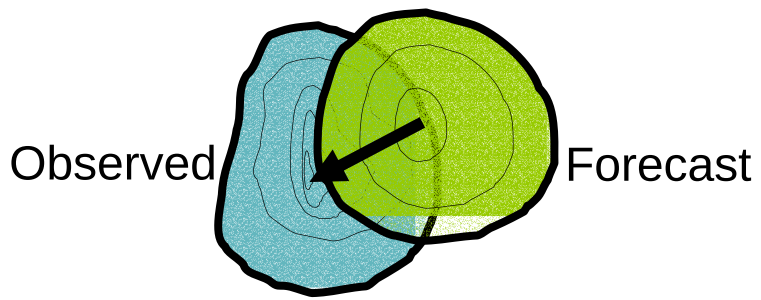


Entity-based verification focuses only on features of interest (entities)

- Determines location errors
- Verifies properties of location-corrected entities
- Decomposes error into location, amplitude, and pattern components

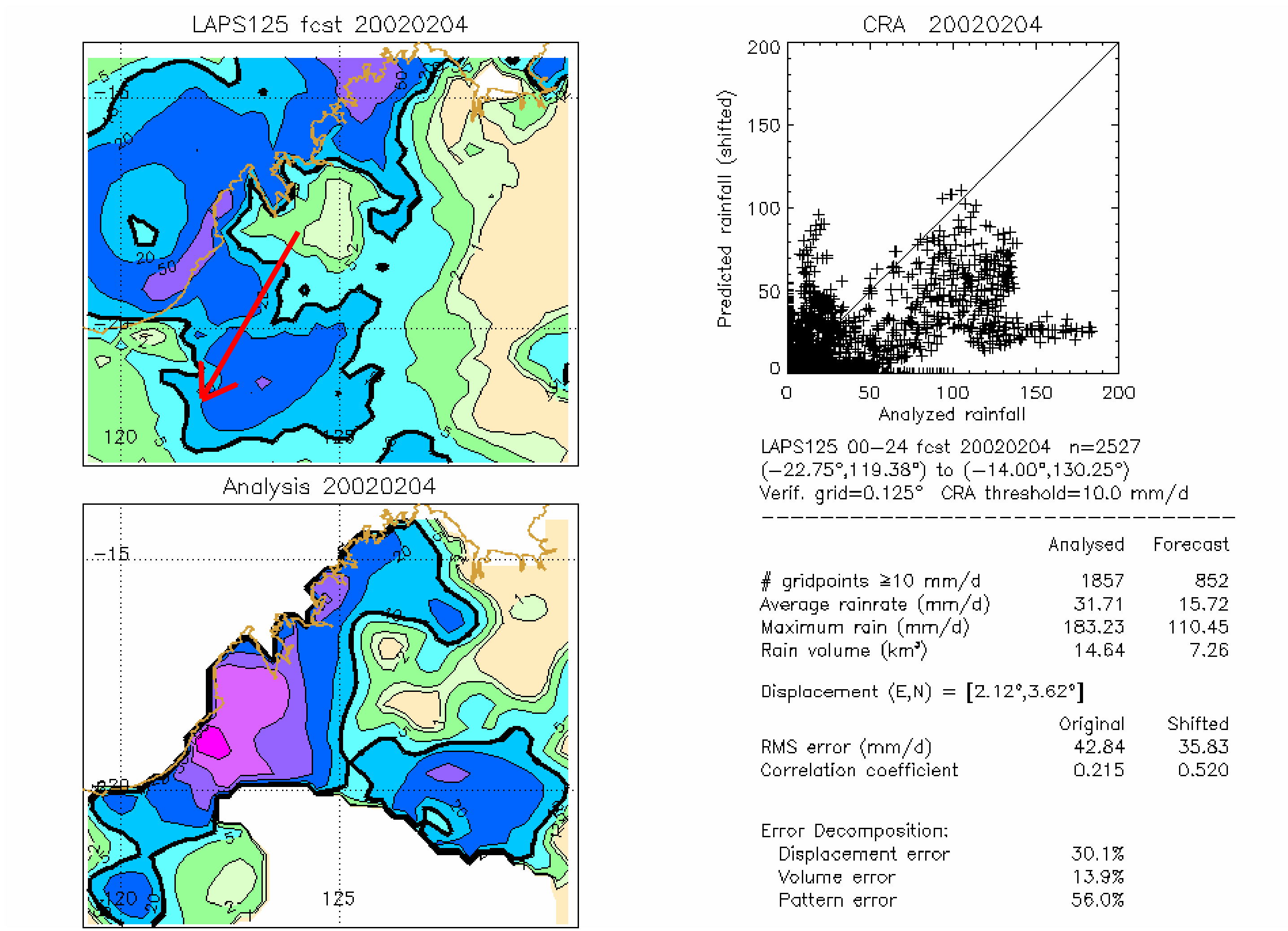
Entity-based approach
(Ebert and McBride, 2000)

- Define entities using threshold (Contiguous Rain Areas)
- Horizontally translate the forecast until a *pattern matching* criterion is met:
  - minimum total squared error between forecast and observations
  - maximum correlation
  - maximum overlap
- The displacement is the vector difference between the original and final locations of the forecast.



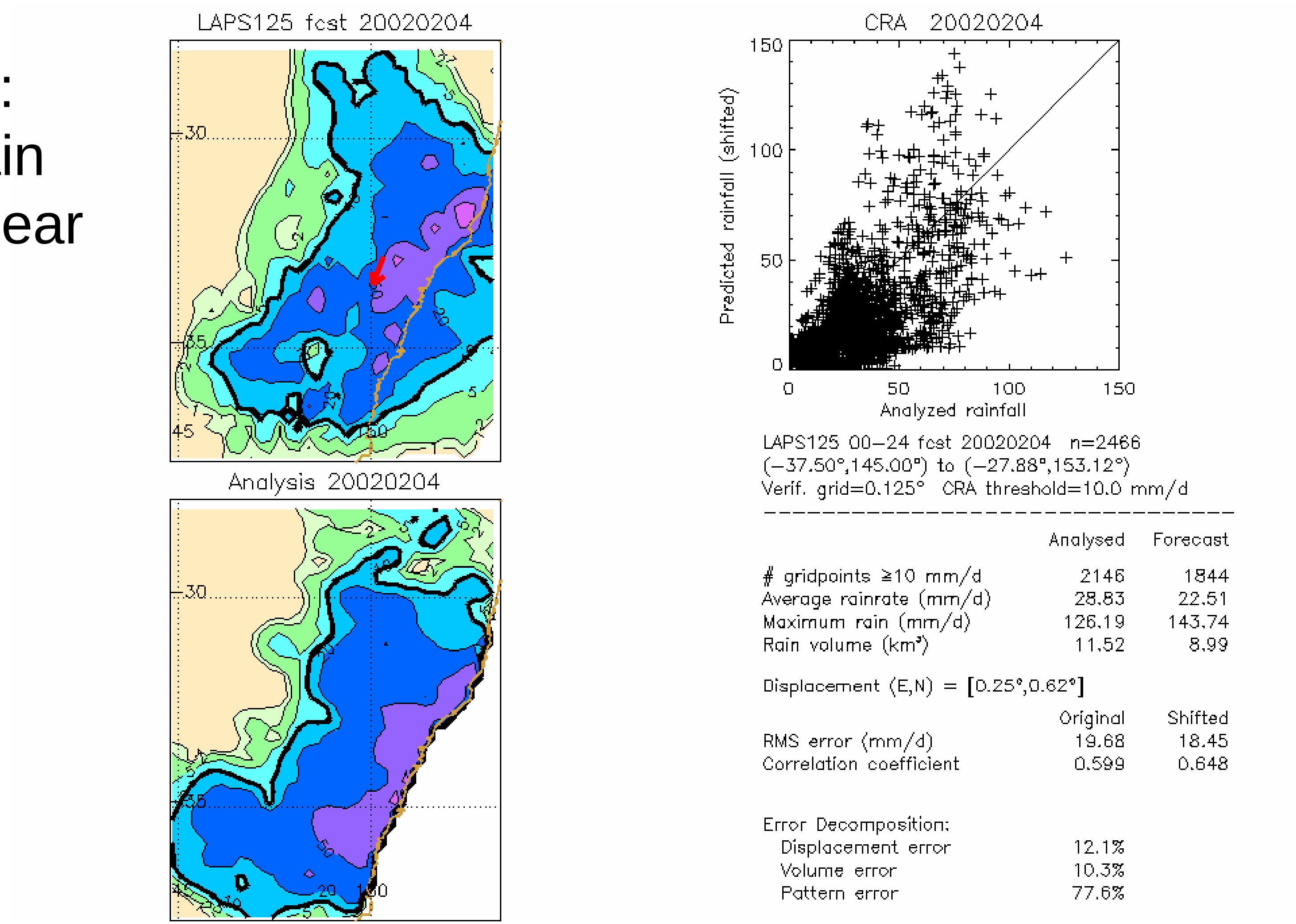
Entity-based approach
(Ebert and McBride, 2000)

1st CRA: Tropical Cyclone Chris



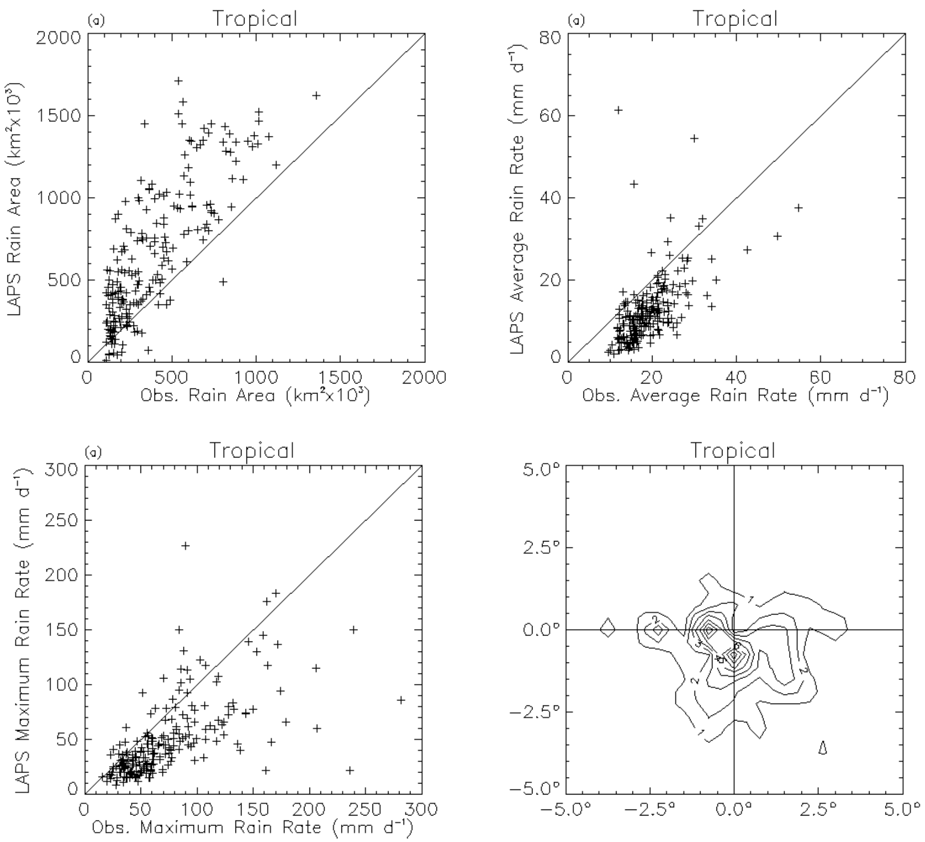
Entity-based approach
(Ebert and McBride, 2000)

2nd CRA: Heavy rain system near Sydney



Entity-based approach
(Ebert and McBride, 2000)

Distribution-oriented verification for entity properties



Entity-based approach
(Ebert and McBride, 2000)

Error decomposition based on optimizing total squared error:

The total mean squared error (MSE) can be written as:

$$MSE_{total} = MSE_{displacement} + MSE_{volume} + MSE_{pattern}$$

The difference between the mean square error before and after translation is the contribution to total error due to *displacement*,

$$MSE_{displacement} = MSE_{total} - MSE_{shifted}$$

The error component due to *volume* represents the bias in mean intensity,

$$MSE_{volume} = (\overline{F} - \overline{X})^2$$

where $\overline{F}$ and $\overline{X}$ are the mean forecast and observed values after the shift.

The *pattern error*, computed as a residual, accounts for differences in the fine structure of the forecast and observed fields,

$$MSE_{pattern} = MSE_{shifted} - MSE_{volume}$$

Entity-based approach
(Ebert and McBride, 2000)

Alternate error decomposition based on optimizing spatial correlation:

The total mean squared error (MSE) can be written as:

$$MSE_{total} = MSE_{displacement} + MSE_{volume} + MSE_{pattern}$$

Murphy's (1995) decomposition of the MSE:

$$MSE = (\overline{F} - \overline{X})^2 + (s_X - r s_F)^2 + (1 - r^2) s_F^2$$

where $\overline{F}$ and $\overline{X}$ are the mean forecast and observed values before the shift. Correcting the forecast location improves its correlation with the observations, r_{opt}. Adding and subtracting r_{opt} and rearranging,

$$MSE_{displacement} = 2 s_F s_X (r_{opt} - r)$$

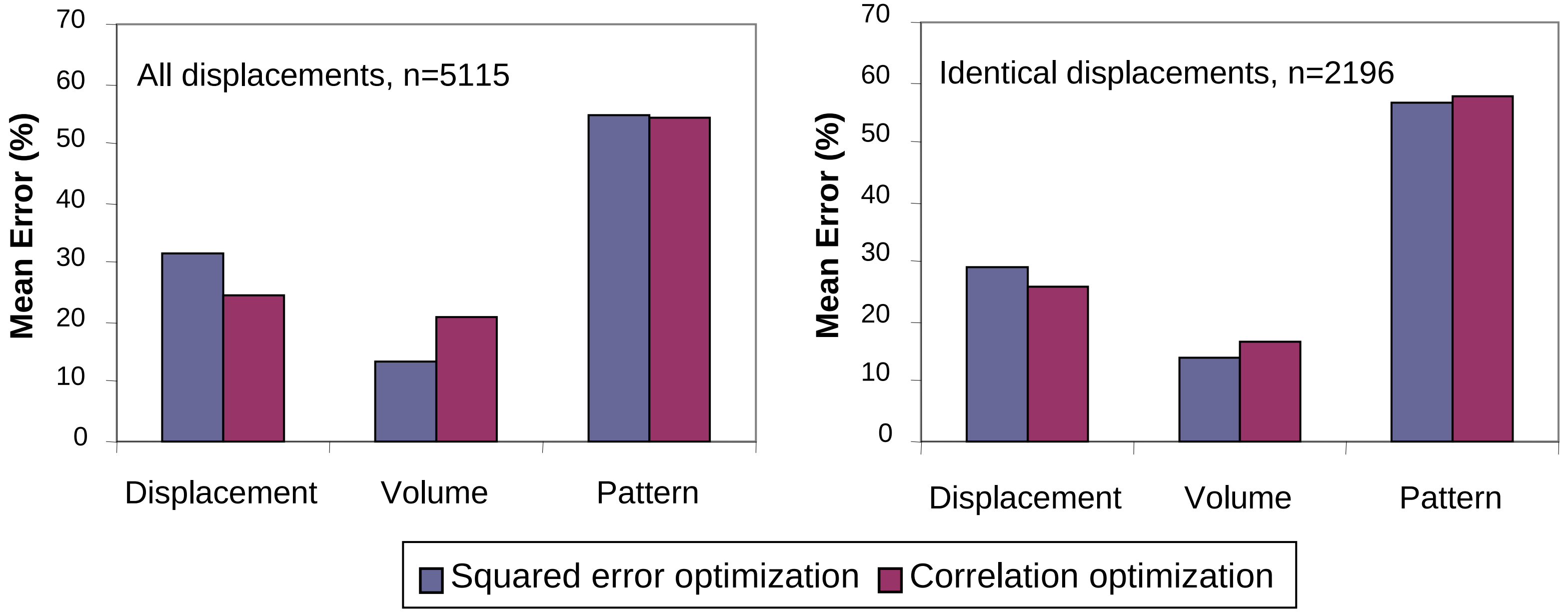
$$MSE_{volume} = (\overline{F} - \overline{X})^2$$

$$MSE_{pattern} = 2 s_F s_X (1 - r_{opt}) + (s_F - s_X)^2$$

Entity-based approach
(Ebert and McBride, 2000)

How do these two error decompositions compare to each other?

Comparison of CRA error decompositions for 24 h QPFs from 7 global and regional NWP models over Australia, 2000-2003.



Entity-based approach
(Ebert and McBride, 2000)

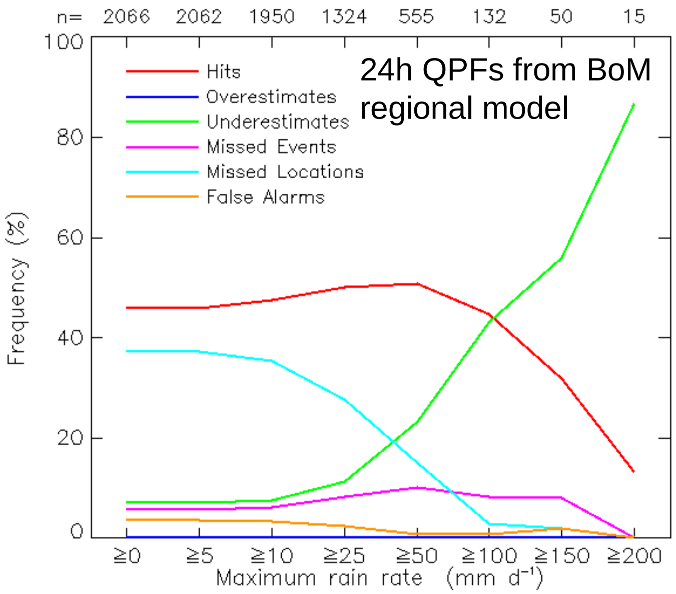
Event forecast classification

Select the two most important aspects of a useful forecast for an event, for example:

- Forecast location must be close to the observed location
- Predicted magnitude must be close to the observed

Forecast Magnitude

Displacement of forecast event		Too small	Approx. correct*	Too great
	Close	*Under-estimate*	***Hit***	*Over-estimate*
	Far	*Missed Event*	*Missed Location*	*False Alarm*



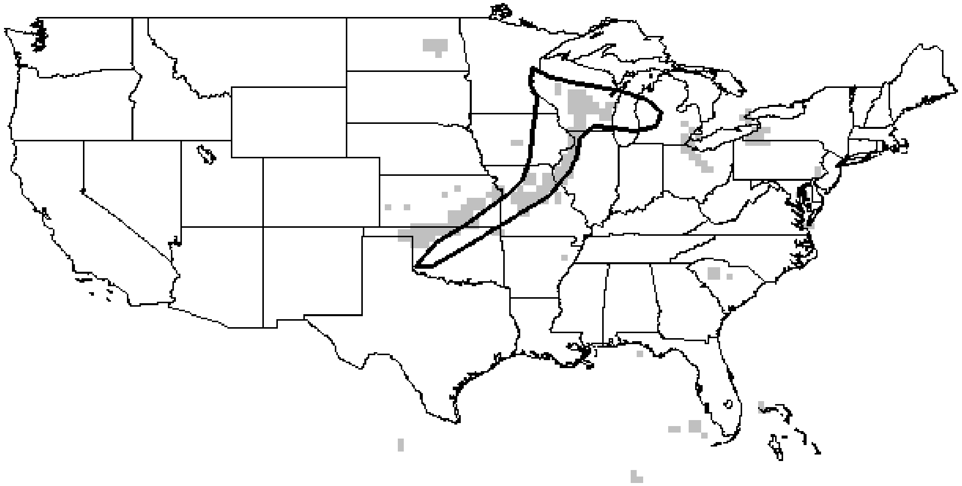
Object-oriented approaches

- GOAL: A verification approach that is consistent with "eyeball" verification
- Diagnostic information that
  - Can be used to improve the forecasts
  - Provides useful information to decision makers
- Types of information
  - Location
  - Amplitude
  - Size
  - Shape
- Examples: Marzban and Sandgathe (U. Washington); Baldwin (NSSL); Bullock, Brown et al. (NCAR)

Example: *13 June 2000, 23Z, 6-h*

Collaborative Convective Forecast Product (CCFP)

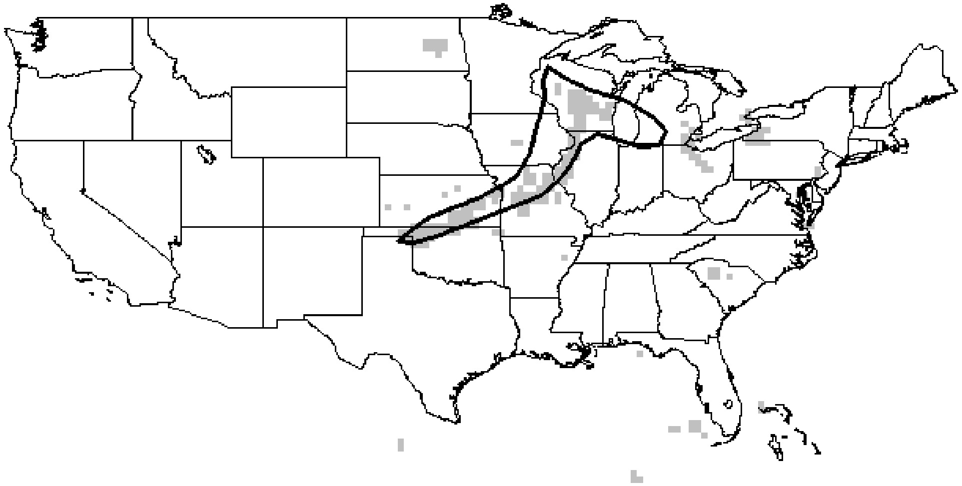
(Human generated forecast)



Original shape:

POD = 0.37 Bias = 1.3

FAR = 0.71 CSI = 0.19



Translated and rotated shape:

POD = 0.52 Bias = 1.3

FAR = 0.60 CSI = 0.29

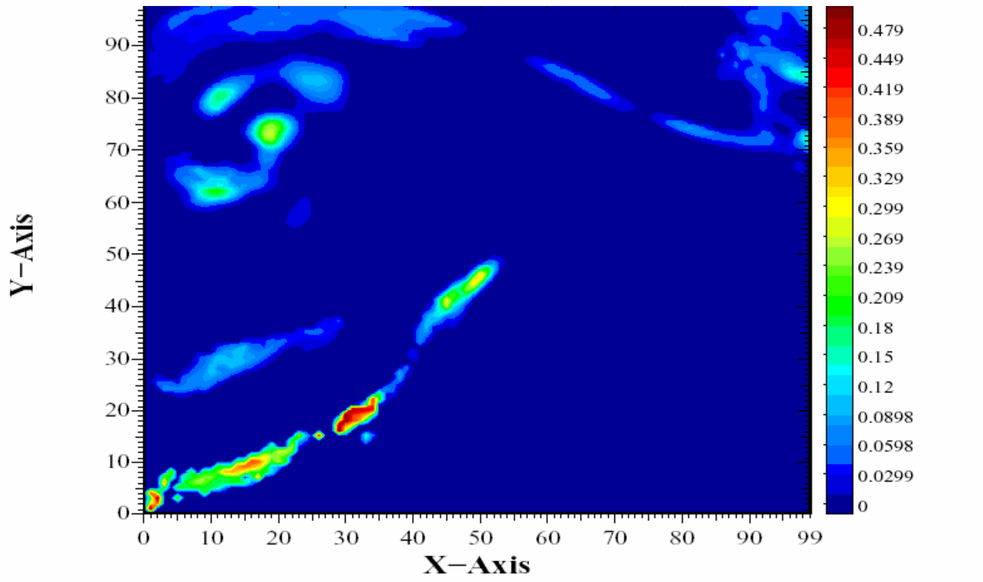
***Translation*: 0 km**

***Rotation*: -10°**

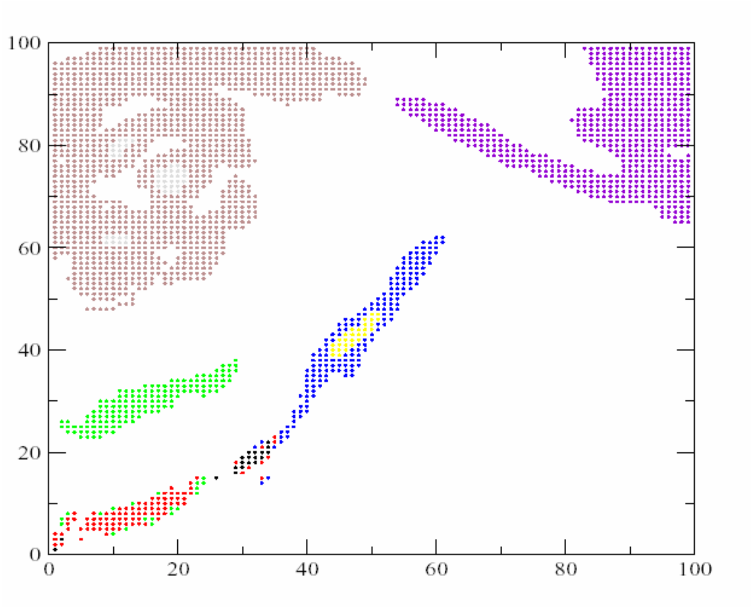
Cluster analysis approach (Marzban and Sandgathe)

- See talk later today (Marzban)
- **Goal**: Assess the agreement between fields using clusters identified using agglomerative hierarchical cluster analysis (CA)
- CA identifies clusters that generally agree with human visual interpretation
- Optimize clusters (and numbers of clusters) based on
  - Binary images (x-y optimization)
  - Magnitude images (x-y-p optimization)
- Compute Euclidean distance between clusters in forecast and observed fields (in x-y and x-y-p space)
  - Evaluate significance using t-test

Cluster analysis approach (Marzban and Sandgathe)



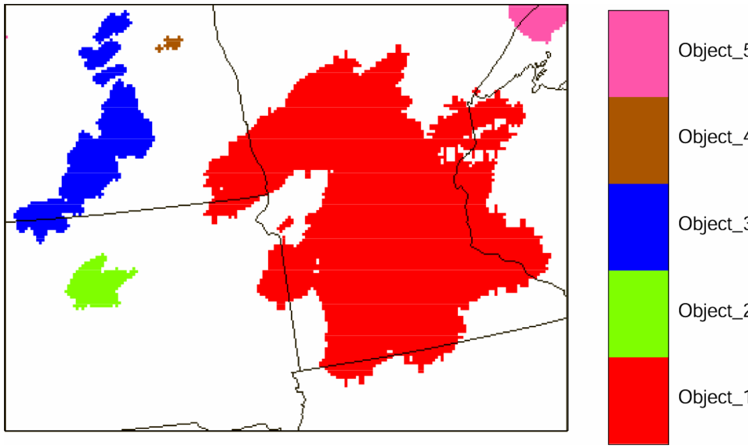
Cluster example: MM5 precipitation forecasts



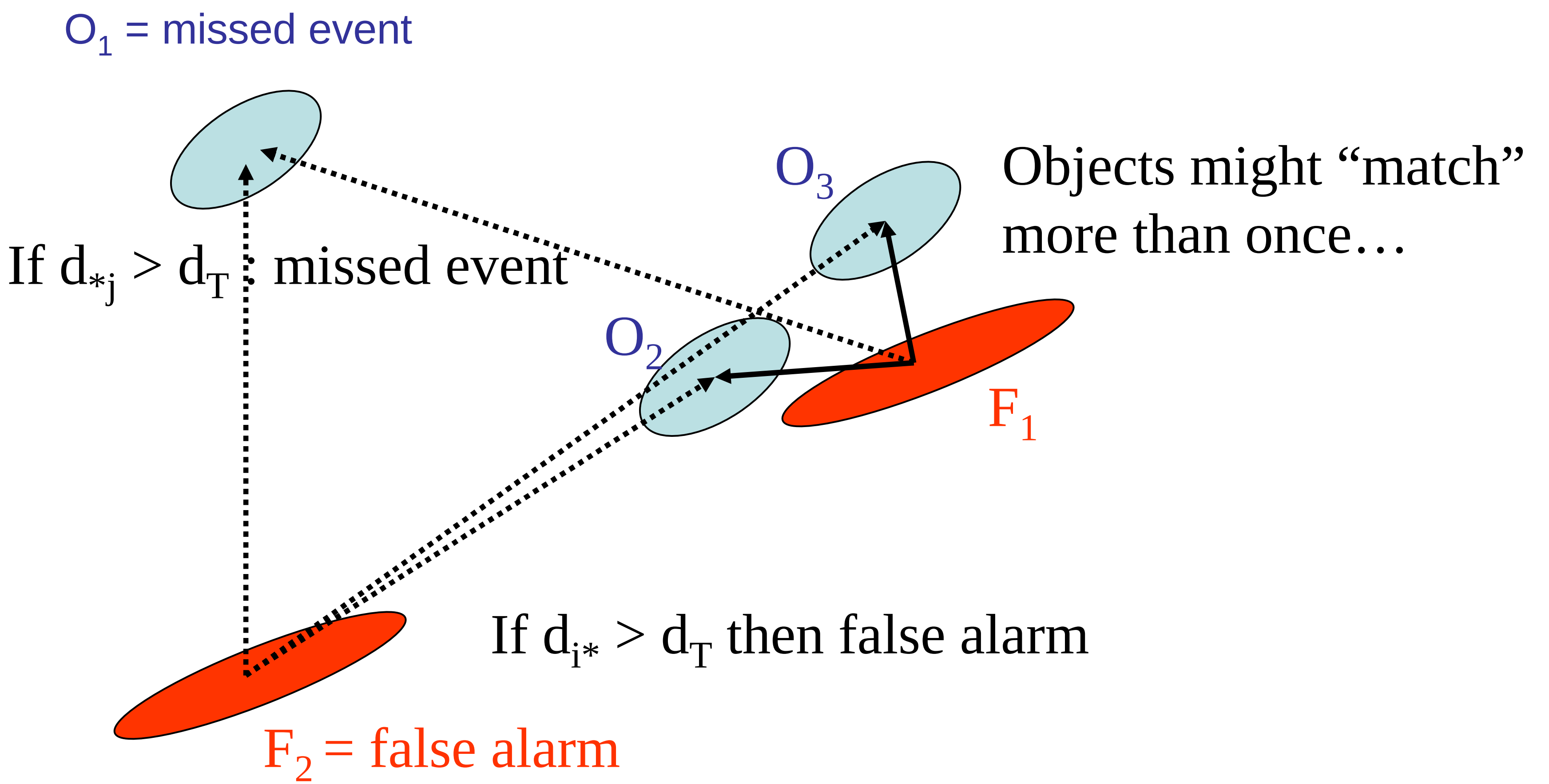
8 clusters identified in x-y-p space

Object-oriented approach (Baldwin)

- See talk later this morning (Brooks/Baldwin)
- Goal: Measure and compare attributes of forecast and observed rain areas
- Use CRA-type approach to identify rainfall areas
- Measure various (11) attributes (area, mean and variance of precip intensity, orientation angle, etc.)
- Measure multi-attribute "distance" between forecast and observed areas
  - Standardize attributes using LEPS
  - Select distance threshold to identify matches, misses and false alarms
- Summarize attributes and distances



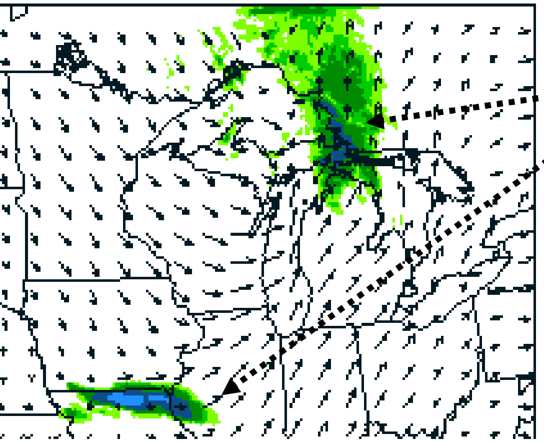
Object-oriented approach (Baldwin)



Baldwin: Object Attributes

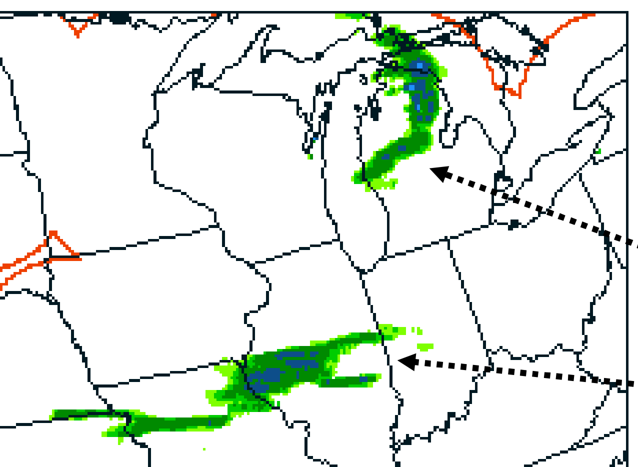
NCAR WRF 4km

WRF NCAR
19-H FCST
4.0 KM LMB CON GRD



Stage II radar ppt

4.8 KM POL STR GRD



Fcst_1	Fcst_2	Obs_1	Obs_2
Area=70000 km^2	Area=285000	Area=135000	Area=70000 km^2
mean(ppt)=0.97	mean(ppt)=0.32	mean(ppt)=0.45	mean(ppt)=0.60
σ (ppt)= 1.26	σ (ppt)= 0.44	σ (ppt)= 0.57	σ (ppt)= 0.67
Δcorr(50)=1.17	Δcorr(50)=0.27	Δcorr(50)=0.37	Δcorr(50)=0.36
Δcorr(100)=0.99	Δcorr(100)=0.42	Δcorr(100)=0.54	Δcorr(100)=0.52
Δcorr(150)=0.84	Δcorr(150)=0.48	Δcorr(150)=0.58	Δcorr(150)=0.49
θ(50)=173°	θ(50)=95°	θ(50)=171°	θ(50)=85°
θ(100)=173°	θ(100)=85°	θ(100)=11°	θ(100)=75°
θ(150)=173°	θ(150)=85°	θ(150)=11°	θ(150)=65°
lat = 40.2°N	lat = 47.3°N	lat = 39.9°N	lat = 44.9°N
lon = 92.5°W	lon = 84.7°W	lon = 91.2°W	lon =84.5°W

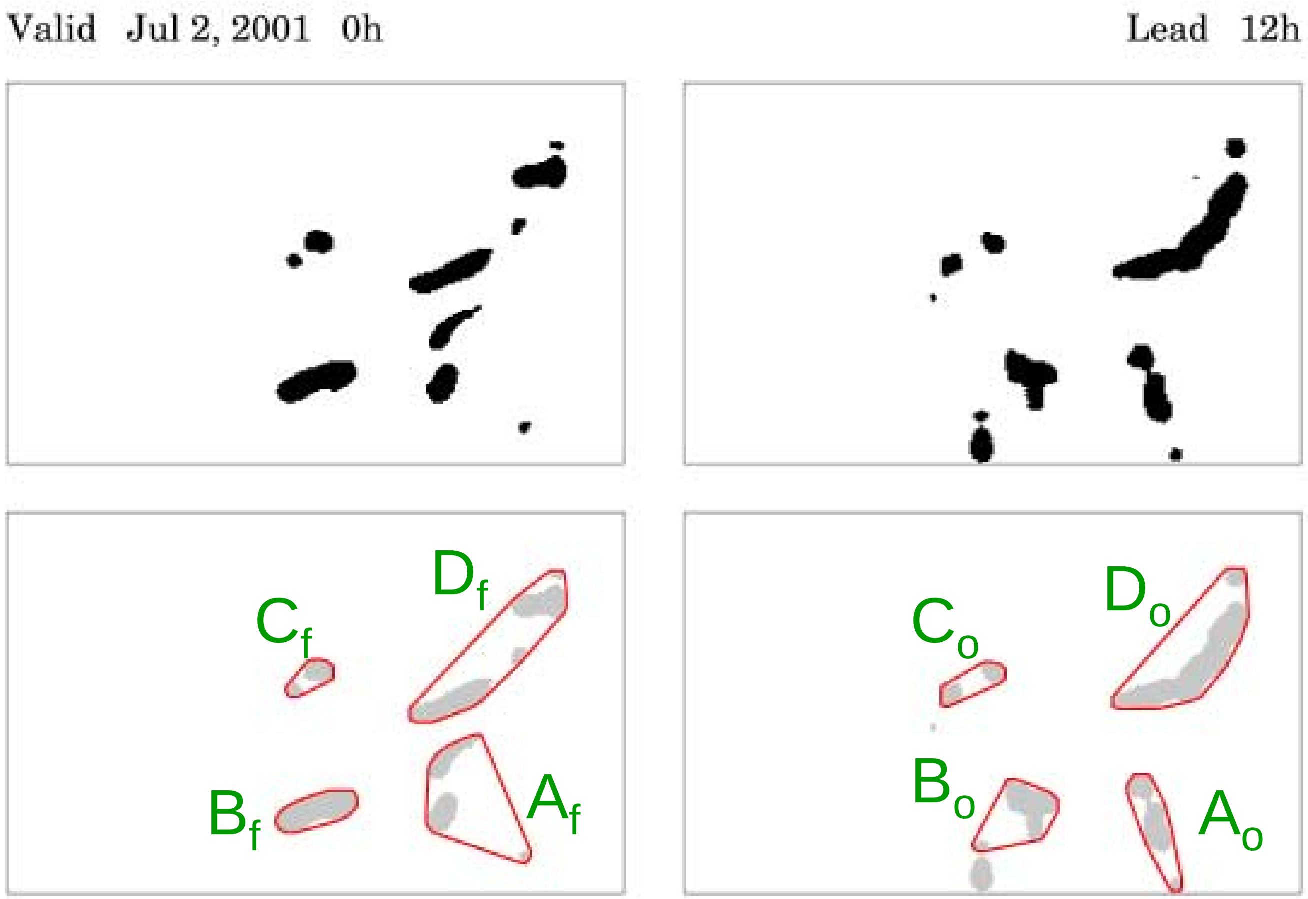
Obs_2

Obs_1

NCAR object-oriented approach

- See talks later today (Bullock, Gilleland, Brown)
- Define objects in forecast and observation field
  - Convolution/threshold approach
- Define characteristics of objects
  - Shape, Location, Orientation angle, Precipitation intensity, “Ugliness”, etc.
- Merge objects in individual fields
- Match objects between fields
  - Two approaches: Fuzzy Logic; Baddeley’s delta
- Characterize performance: examine object differences
- Summarize performance across a set of forecasts

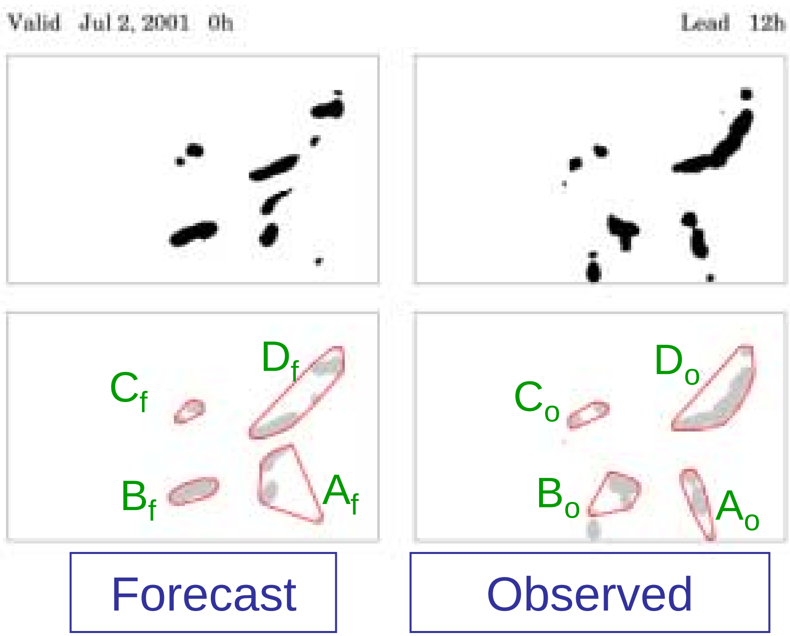
NCAR object-oriented approach



Forecasts: 12-h WRF precipitation

Obs: Stage IV precipitation

NCAR object-oriented approach



Composite object	Intersection Area (IA)	Symmetric Difference (SD) Area
A	0	638
B	100	488
C	66	134
D	259	905

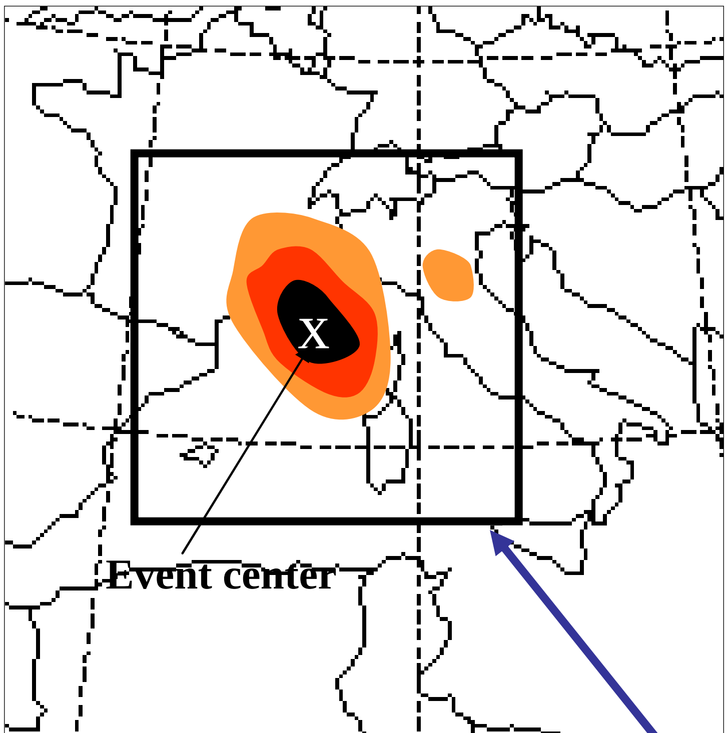
Attribute	WRF	Stage IV	Difference
Composite Objects "A"			
Centroid X	187	197	-10
Centroid Y	44	31	13
Intensity (0.50)	4.7	2.5	2.2
Intensity (0.90)	8.5	13.9	-5.4
Area	319	319	0
Composite Objects "B"			
Centroid X	130	144	-14
Centroid Y	36	36	0
Intensity (0.50)	4.7	2.0	2.7
Intensity (0.90)	8.7	9.7	-1.0
Area	355	333	22
Composite Objects "C"			
Centroid X	128	121	7
Centroid Y	93	90	3
Intensity (0.50)	4.0	2.4	1.6
Intensity (0.90)	8.5	11.3	-2.8
Area	126	140	-14
Composite Objects "D"			
Centroid X	205	215	-10
Centroid Y	102	100	2
Intensity (0.50)	3.8	3.8	0
Intensity (0.90)	7.4	13.8	-6.4
Area	585	838	-253

Composite approach (Nachamkin)

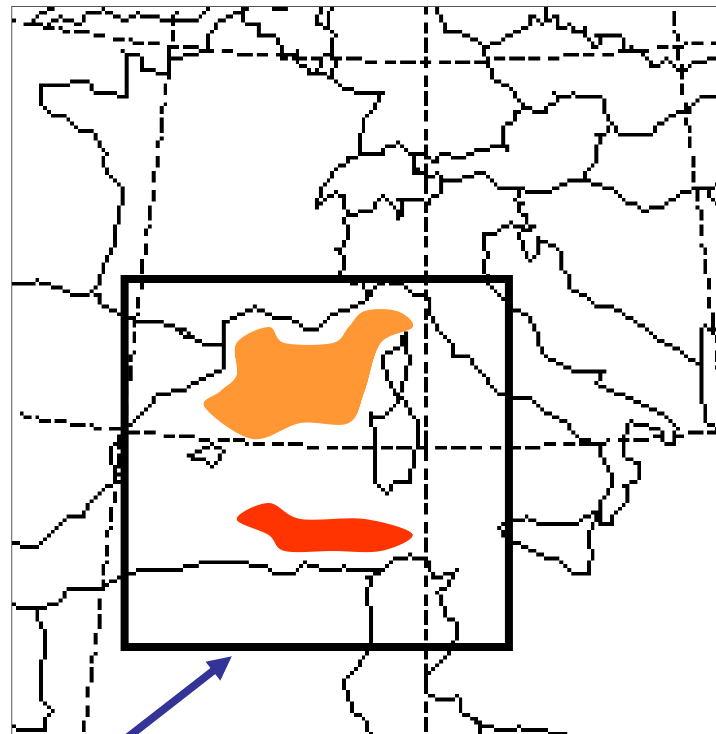
- Talk yesterday afternoon
- **Goal**: Characterize distributions of errors from both a forecast and observation perspective
- Procedure:
  - Identify events of interest in the forecasts
    - Rainfall greater than 25 mm
    - Event contains between 50 and 500 grid points
  - Define a kernel and collect coordinated samples
    - Square box
    - 31x31 grid points (837x837 km for 27 km grid)
  - Compare forecast PDF to observed PDF
  - Repeat process for observed events

Composite approach: Collecting the Samples

Forecast event



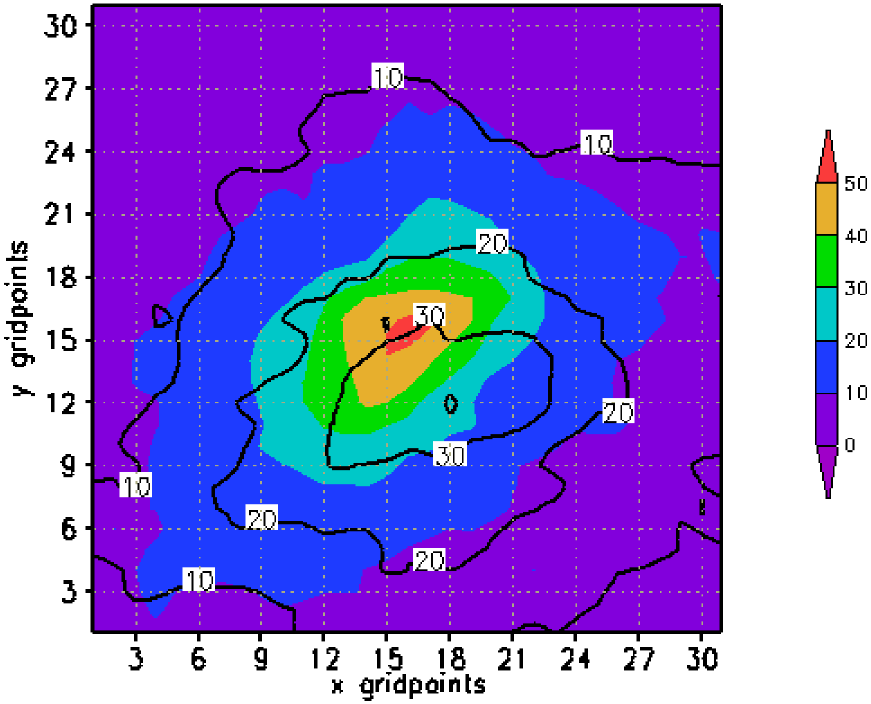
Observations



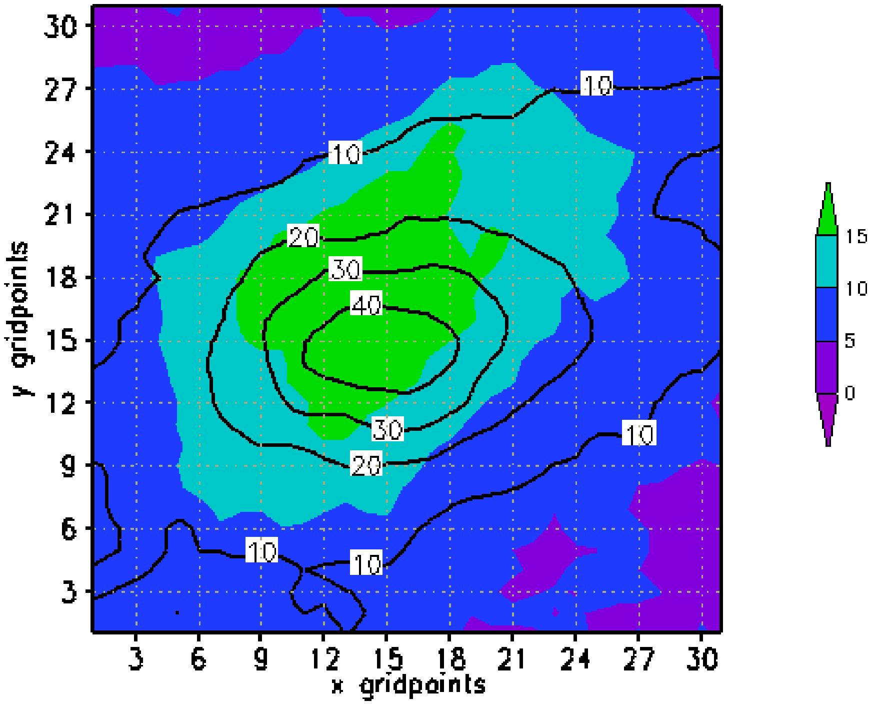
Collection kernel

Composite approach: Kernel grid-average precipitation

Average rain (mm) given an event was predicted



Average rain (mm) given an event was observed



Summary

- Standard methods provide basic information about performance of spatial forecasts, but are not diagnostic and may not provide information needed to improve the forecasts or to optimally use the forecasts
- Issues of scaling are still of concern but are generally ignored or only considered minimally
- Much progress! in the last 2 years
  - Event- or object-based approaches
  - Fuzzy verification approaches
  - Other alternative approaches (e.g., composite methods; new skill measures)

The future?

- Evolving ideas about "What makes a 'good' forecast"
  - Look at bigger picture
  - Reliance on probability distributions
- Further development of diagnostic approaches
  - Measure attributes of interest to users
  - Application to other forecast elements
- Improved approaches for coping with
  - Scaling issues
  - Spatial correlations
- Application of approaches that provide information that is "operationally relevant"
- Incorporation of observation uncertainty

References (1)

Atger, F., 2001: Verification of intense precipitation forecasts from single models and ensemble prediction systems. *Nonlin. Proc. Geophys.*, **8**, 401-417.

Baldwin, M., 2004: Object identification techniques for object-oriented verification. *Int'l Verification Methods Workshop, 15-17 Sept. 2004, Montreal.*

Briggs, W.M. and R.A. Levine, 1997: Wavelets and field forecast verification. *Mon. Wea. Rev.*, **125**, 1329-1341.

Brooks, H.E., M. Kay and J.A. Hart, 1998: Objective limits on forecasting skill of rare events. *19th Conf. Severe Local Storms, AMS*, 552-555.

Brown, B.G., R.R. Bullock, C.A. David, J.H. Gotway, M.B. Chapman, A. Takacs, E. Gilleland, K. Manning, J. Mahoney, 2004: New verification approaches for convective weather forecasts. *11th Conf. Aviation, Range, and Aerospace Meteorology, 4-8 Oct 2004, Hyannis, MA.*

Casati, B., Ross, D.B. Stephenson, 2004: A new intensity-scale approach for the verification of spatial precipitation forecasts, *Met. App*., **11**, 141-154.

Damrath, U., 2004: Verification against precipitation observations of a high density network – what did we learn? *Int'l Verification Methods Workshop, 15-17 Sept. 2004, Montreal.*

Denis, B., J. Côté and R. Laprise, 2002a: Spectral decomposition of two-dimensional atmospheric fields on limited-area domains using the discrete cosine transform (DCT). *Mon. Wea. Rev.*, **130**, 1812-1829.

Ebert, E.E. and J.L. McBride, 2000: Verification of precipitation in weather systems: Determination of systematic errors. *J. Hydrology*, **239**, 179-202.

References (2)

Harris, D., E. Foufoula-Georgiou, K.K. Droegemeier and J.J. Levit, 2001: Multiscale statistical properties of a high-resolution precipitation forecast. *J. Hydromet.*, **2**, 406-418.

Hoffman, R.N., Z. Liu, J.-F. Louis, and C. Grassotti, 1995: Distortion representation of forecast errors. Mon. Wea. Rev., **123**, 2758-2770.

Marzban, C. and S. Sandgathe, 2004: Cluster analysis for verification of precipitation fields. *Wea. Forecasting*, submitted.

Mass, C.F., D.Ovens, K. Westrick and B.A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? *Bull. Amer. Met. Soc.*, **83**, 407-430.

Nachamkin, J.E., 2004: Mesoscale verification using meteorological composites. *Mon. Wea. Rev.*, **132**, 941-955.

Nehrkorn, T., R.N. Hoffman, C.Grassotti and J.-F. Louis, 2003: Feature calibration and alignment to represent model forecast errors: Empirical regularization. *Q. J. R. Meteorol. Soc.*, **129**, 195-218.

Tustison, B., E. Foufoula-Georgiou, and D. Harris, 2003: Scale-recursive estimation for multisensor Quantitative Precipitation Forecast verification: A preliminary assessment. *J. Geophys. Res.*, **108**, D8, 8377.

Venugopal, V., S. Basu and E. Foufoula-Georgiou, 2004: A new metric for comparing precipitation patterns with an application to ensemble forecasts. *J. Geophys. Res*., submitted.

Zepeda-Arce, J., E. Foufoula-Georgiou, and K.K. Droegemeier, 2000: Space-time rainfall organization and its role in validating quantitative precipitation forecasts. *J. Geophys. Res.*, **105** (D8), 10,129-10,146.