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)

<table>
<thead>
<tr>
<th></th>
<th>RMSE</th>
<th>EqTh</th>
<th>FQI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original and Member 1</td>
<td>68.41</td>
<td>-0.02</td>
<td>0.3924</td>
</tr>
<tr>
<td>Original and Member 2</td>
<td>68.41</td>
<td>-0.02</td>
<td>1.15</td>
</tr>
</tbody>
</table>
“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)
- H-K = POD + POFD -1
- Skill scores (Heidke, Gilbert/ETS)

Measures “relative accuracy”

Measures “discrimination” between Yes and No observations

Measures proportion of observed area that is correctly forecast to be “Yes”

Measures proportion of area that is correctly forecast to be “No”

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)

Generated on 12 Sep 2004 by NOAA/FSL-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
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)

**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

Standard verification methods

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...

```
<table>
<thead>
<tr>
<th>Observation</th>
<th>CMC</th>
<th>ECWMF</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCEP</td>
<td>DWD</td>
<td>DM</td>
</tr>
<tr>
<td>fc T+06 - T+30</td>
<td>fc T+06 - T+30</td>
<td>fc T+06 - T+30</td>
</tr>
<tr>
<td>fc T+18 - T+42</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

Legend:
- > 250.0
- 225.0 - 250.0
- 200.0 - 225.0
- 175.0 - 200.0
- 150.0 - 175.0
- 125.0 - 150.0
- 100.0 - 125.0
- 75.0 - 100.0
- 50.0 - 75.0
- 25.0 - 50.0
- < 25.0

3481 stations
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
Convective outlook is 75\% of the way to being "practically perfect"

Practically Perfect forecast
Threat score=0.48
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

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>yes</td>
<td>no</td>
</tr>
</tbody>
</table>

Forecasted proportion

\[
P_o \geq X\% \quad P_o < X\%
\]

<table>
<thead>
<tr>
<th>Predicted</th>
<th>Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>yes</td>
<td>no</td>
</tr>
</tbody>
</table>

Fuzzy forecasts

\[
P_o \geq X\%
\]

\[
P_f \geq X\%
\]

\[
P_o \quad 1-P_o
\]

\[
P_f \quad 1-P_f
\]

\[
\min(P_o, P_f) \quad \min(1-P_o, P_f)
\]

\[
\min(P_o, 1-P_f) \quad \min(1-P_o, 1-P_f)
\]

Forecasts partially correct, partially incorrect
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  $\rightarrow$ high probability  
Not very close  $\rightarrow$ 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

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

\[ N = (\# \text{ rain thresholds}) \times (\# \text{ distance thresholds}) \]

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

\[ N = (\# \text{ rain thresholds}) \times (\# \text{ distance thresholds}) \times (\# \text{ ensemble members}) \]
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 it 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

\[ e = X_f(r) - X_v(r) \]

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

\[ e = e_p + e_b + e_r \]

where

- \( e_p = X_f(r) - X_d(r) \) phase error
- \( e_b = X_d(r) - X_a(r) \) local bias error
- \( e_r = X_a(r) - X_v(r) \) 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,

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

where $\bar{F}$ and $\bar{X}$ 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:

$$MSE = (F - \bar{X})^2 + (S_X - rs_F)^2 + (1 - r^2)s_F^2$$

where $\bar{F}$ and $\bar{X}$ 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,

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

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

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

**Displacement of forecast event**

<table>
<thead>
<tr>
<th>Forecast Magnitude</th>
<th>Too small</th>
<th>Approx. correct*</th>
<th>Too great</th>
</tr>
</thead>
<tbody>
<tr>
<td>Under-estimate</td>
<td>Missed Event</td>
<td>Missed Location</td>
<td>False Alarm</td>
</tr>
<tr>
<td>Hit</td>
<td>Hit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Over-estimate</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**24h QPFs from BoM regional model**
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)

Objects might “match” more than once…

$O_1 = \text{missed event}$

If $d_{*j} > d_T$: missed event

$F_2 = \text{false alarm}$

If $d_{i*} > d_T$ then false alarm

$F_1$

$O_2$

$O_3$
NCAR WRF 4km

**Stage II radar ppt**

<table>
<thead>
<tr>
<th>Obj</th>
<th>Area (km²)</th>
<th>mean (ppt)</th>
<th>σ (ppt)</th>
<th>Δcorr(50)</th>
<th>Δcorr(100)</th>
<th>Δcorr(150)</th>
<th>θ(50)</th>
<th>θ(100)</th>
<th>θ(150)</th>
<th>lat</th>
<th>lon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fcst_1</td>
<td>70000</td>
<td>0.97</td>
<td>1.26</td>
<td>1.17</td>
<td>0.99</td>
<td>0.84</td>
<td>173°</td>
<td>173°</td>
<td>173°</td>
<td>40.2°</td>
<td>92.5°</td>
</tr>
<tr>
<td>Fcst_2</td>
<td>285000</td>
<td>0.32</td>
<td>0.44</td>
<td>0.27</td>
<td>0.42</td>
<td>0.48</td>
<td>95°</td>
<td>85°</td>
<td>85°</td>
<td>47.3°</td>
<td>84.7°</td>
</tr>
<tr>
<td>Obs_1</td>
<td>135000</td>
<td>0.45</td>
<td>0.57</td>
<td>0.37</td>
<td>0.54</td>
<td>0.58</td>
<td>171°</td>
<td>11°</td>
<td>11°</td>
<td>39.9°</td>
<td>91.2°</td>
</tr>
<tr>
<td>Obs_2</td>
<td>70000</td>
<td>0.60</td>
<td>0.67</td>
<td>0.36</td>
<td>0.52</td>
<td>0.49</td>
<td>85°</td>
<td>75°</td>
<td>65°</td>
<td>44.9°</td>
<td>84.5°</td>
</tr>
</tbody>
</table>

**Note:**
- **Area:** The area covered by the radar observations.
- **mean (ppt):** The average precipitation rate.
- **σ (ppt):** The standard deviation of precipitation rate.
- **Δcorr:** The correlation coefficient at different time lags.
- **θ:** The direction of the precipitation.
- **lat, lon:** The latitude and longitude coordinates.
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

![Diagram showing forecast and observed patterns with labeled components: C, D, A, B, F, O, Fo, Do, Ao, Bo, Cf, Dj, Af, Ci, Do, Ao, Co, Do.]

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

Event center

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)


