Spatial methods

Beth Ebert
Center for Australian Weather and Climate Research
Melbourne, Australia
What is a *spatial* verification method?

- Applies to mapped forecasts and observations
- Takes account of coherent spatial structure and features
- Provides information on errors in location, scale, and structure in forecast fields
- Does not rely exclusively on point-by-point matching
  - Particularly appropriate for high resolution forecasts where perfect performance at fine scale may be impossible due to predictability limits
Classes of spatial verification methods

- **Scale decomposition methods**
  - isolate scale-dependent error

- **Fuzzy (neighborhood) verification methods**
  - give credit to "close" forecasts

- **Object-oriented methods**
  - evaluate attributes of identifiable features

- **Field verification methods**
  - evaluate phase errors

Most methods were developed to compare model QPF to radar rainfall, but could be used for other types of data.
Scale decomposition methods
→ scale-dependent error
Intensity-scale method
Casati et al. (2004)

Evaluate forecast skill as a function of the precipitation intensity and the spatial scale of the error.
Intensity threshold $\rightarrow$ binary images

**Binary Analysis**

**Binary Forecast**

$E_u = I_{Y'>u} - I_{X'>u}$
Scale → wavelet decomposition of binary error

\[
E_u = \sum_{l=1}^{L} E_{u,l} \quad \text{MSE}_u = \sum_{l=1}^{L} \text{MSE}_{u,l}
\]
MSE skill score

\[ SS_{u,l} = \frac{MSE_{u,l} - MSE_{u,l,\text{random}}}{MSE_{u,l,\text{best}} - MSE_{u,l,\text{random}}} = 1 - \frac{MSE_{u,l}}{2\varepsilon(1 - \varepsilon)/L} \]

Status: Available in MET
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 multi-scale statistics for model and radar data

Status: Not implemented in MET
Fuzzy (neighborhood) verification methods
→ give credit to "close" forecasts
"Fuzzy" (neighborhood) verification methods
Ebert (2008)

- Don't require an exact match between forecasts and observations
  - Unpredictable scales
  - Uncertainty in observations

- Look in a space / time neighborhood around the point of interest
  - Evaluate using categorical, continuous, probabilistic scores / methods
Fuzzy verification framework

Treatment of forecast data within a window:

- Mean value (upscaling)
- Occurrence of event* somewhere in window
- Frequency of event in window $\rightarrow$ probability
- Distribution of values within window

* Event = value exceeding a given threshold, for example, rain exceeding 1 mm/h
Upscaling

Average the forecast and observations to successively larger grid resolutions, then verify as usual

% change in ETS

 Improvement

 RUC20

 Weygandt et al. (2004)

 Degradation

 RUC10

 LMM12

 ETA12

 DTC Verification Workshop, Boulder, 16-18 April 2008

 Weygandt et al. (2004)
Fractions skill score
Roberts and Lean (2008)

Compares forecast fractions with observed fractions (radar) in a *probabilistic* way over different sized neighbourhoods

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

\[
= \frac{1}{N} \sum_{i=1}^{N} P_{\text{fcst}}^2 + \frac{1}{N} \sum_{i=1}^{N} P_{\text{obs}}^2
\]

Fraction = 6/25 = 0.24

observed

forecast
Spatial multi-event contingency table
Atger (2001)

Measures how close a forecast event was to an observed event in terms of location / timing / magnitude

Vary decision thresholds:

• 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

Fuzzy methodology – compute Hanssen and Kuipers score   HK = POD – POFD
Practically perfect hindcasts
(Kay and Brooks 2000)

Q: 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 yes/no threshold that maximizes CSI when verified against obs
- Did the actual forecast look like the practically perfect forecast?
- How did the performance of the actual forecast compare to the performance of the practically perfect forecast?

Fuzzy methodology – compute

\[
\frac{\text{ETS}_{\text{forecast}}}{\text{ETS}_{\text{PracPerf}}}
\]

\[
\text{CSI}_{\text{forecast}} = 0.34 \quad \text{CSI}_{\text{PracPerf}} = 0.48
\]
## Decision models

<table>
<thead>
<tr>
<th>Fuzzy method</th>
<th>Matching strategy*</th>
<th>Decision model for useful forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Upscaling</strong> (Zepeda-Arce et al. 2000; Weygandt et al. 2004)</td>
<td>NO-NF</td>
<td>Resembles obs when averaged to coarser scales</td>
</tr>
<tr>
<td><strong>Minimum coverage</strong> (Damrath 2004)</td>
<td>NO-NF</td>
<td>Predicts event over minimum fraction of region</td>
</tr>
<tr>
<td><strong>Fuzzy logic</strong> (Damrath 2004), joint probability (Ebert 2002)</td>
<td>NO-NF</td>
<td>More correct than incorrect</td>
</tr>
<tr>
<td><strong>Fractions skill score</strong> (Roberts and Lean 2008)</td>
<td>NO-NF</td>
<td>Similar frequency of forecast and observed events</td>
</tr>
<tr>
<td><strong>Area-related RMSE</strong> (Rezacova et al. 2006)</td>
<td>NO-NF</td>
<td>Similar intensity distribution as observed</td>
</tr>
<tr>
<td><strong>Pragmatic</strong> (Theis et al. 2005)</td>
<td>SO-NF</td>
<td>Can distinguish events and non-events</td>
</tr>
<tr>
<td><strong>CSRR</strong> (Germann and Zawadzki 2004)</td>
<td>SO-NF</td>
<td>High probability of matching observed value</td>
</tr>
<tr>
<td><strong>Multi-event contingency table</strong> (Atger 2001)</td>
<td>SO-NF</td>
<td>Predicts at least one event close to observed event</td>
</tr>
<tr>
<td><strong>Practically perfect hindcast</strong> (Brooks et al. 1998)</td>
<td>SO-NF</td>
<td>Resembles forecast based on perfect knowledge of observations</td>
</tr>
</tbody>
</table>

*NO-NF = neighborhood observation-neighborhood forecast,
SO-NF = single observation-neighborhood forecast
Fuzzy verification framework

Status: Available in MET
How are scale decomposition methods and fuzzy neighborhood methods different?

- Scale decomposition methods use wavelet or Fourier transforms to isolate the information at different scales.

- Fuzzy neighborhood methods smooth the information at smaller scales.
Object-oriented methods

evaluate attributes of features
Entity-based approach (CRA)  
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.
CRA information

Gives information on:

- Location error
- RMSE and correlation before and after shift
- Attributes of forecast and observed entities
- Error components
  - displacement
  - volume
  - pattern

**Status:** Not yet implemented in MET
Method for Object-based Diagnostic Evaluation (MODE)

Davis et al. (2006)

Two parameters:
1. Convolution radius
2. Threshold
MODE object matching/merging

Compare attributes:
- centroid location
- intensity distribution
- area
- orientation
- etc.

When objects not matched:
- false alarms
- missed events
- rain volume
- etc.

24h forecast of 1h rainfall on 1 June 2005
MODE methodology

**Identification**

- Measure Attributes

**Merging**

- Matching

**Comparison**

- Summarize

**Convolution – threshold process**

**Fuzzy Logic Approach**

- Compare forecast and observed attributes
- Merge single objects into composite objects
- Compute interest values
- Identify matched pairs

**Accumulate and examine comparisons across many cases**

**Status:** Available in MET
Composite approach
Nachamkin (2004)

Characterize distributions of errors from both a forecast and observation perspective

- Procedure:
  - Identify events of interest in the forecasts
  - Define a kernel and collect coordinated samples
  - Compare forecast PDF to observed PDF
  - Repeat process for observed events
Composite example

Compare kernel grid-averaged values

Average rain (mm) given an event was predicted

Average rain (mm) given an event was observed

Status: Not implemented in MET
Structure-Amplitude-Location (SAL)  
Wernli et al. (2008)

Consider precipitation in pre-specified area (e.g. river catchment)

SAL consists of three independent components:

- components address quality of structure (S), amplitude (A) and location (L) of QPF in that area
- according to SAL a forecast is perfect if $S = A = L = 0$
- $S$ requires the definition of precipitation objects, using a threshold value
- *but* no attribution between precipitation objects in forecast and observations!
SAL: definitions of components

\[ A = \frac{(D(R_{\text{mod}}) - D(R_{\text{obs}}))}{0.5*(D(R_{\text{mod}}) + D(R_{\text{obs}}))} \]

\[ D(...) \text{ denotes the area-mean value (e.g. catchment)} \]
\[ \text{normalized amplitude error in considered area} \]
\[ A \in [-2, \ldots, 0, \ldots, +2] \]

\[ L = \frac{|r(R_{\text{mod}}) - r(R_{\text{obs}})|}{\text{dist}_{\text{max}}} \]

\[ r(...) \text{ denotes the centre of gravity of the precipitation field in the area} \]
\[ \text{normalized location error in considered area} \]
\[ L \in [0, \ldots, 1] \]

\[ S = \frac{(V(R_{\text{mod}}) - V(R_{\text{obs}}))}{0.5*(V(R_{\text{mod}}) + V(R_{\text{obs}}))} \]

\[ V(...) \text{ denotes the weighted volume average of all scaled precipitation objects in considered area} \]
\[ \text{normalized structure error in considered area} \]
\[ S \in [-2, \ldots, 0, \ldots, +2] \]
S A L statistics for 24h accumulations

summer seasons 2001-2004 for catchment Rhine

Status: Not implemented in MET
Cluster analysis verification
Marzban and Sandgathe (2008)

- Assess the agreement between fields using clusters identified using combinative cluster analysis on (x, y, intensity)
- At each iteration, a given cluster is declared as “hit” if the proportion of observed pixels is between 20% and 80%. Otherwise it is a miss (>80% obs) or a false alarm (<20% obs).

**Status:** Not implemented in MET
Field verification

→ evaluate phase errors
Forecast quality measure (FQM)  
Keil and Craig (2007)

Combines distance measure and intensity difference measure
- Pyramidal image matching (optical flow) to get vector displacement field \( e_{\text{distance}} \)
- Intensity errors of morphed image are also penalized \( e_{\text{intensity}} \)

\[
FQM = \frac{1}{A} \sum_{A} (c_1 e_{\text{distance}} + c_2 e_{\text{intensity}})
\]

\( c_1 = \text{(search radius)}^{-1} \)
\( c_2 = \text{(climatological observed value)}^{-1} \)

**Status:** Not implemented in MET
## Summary of spatial verification approaches (1)

<table>
<thead>
<tr>
<th>Spatial verification method</th>
<th>Status</th>
<th>What it measures</th>
<th>Observational data</th>
<th>scale error</th>
<th>location error</th>
<th>timing error</th>
<th>intensity error</th>
<th>false alarms/misses</th>
<th>selectable parameters</th>
<th>fast</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional continuous, categorical statistics (RMSE, POD, FAR, etc.)</td>
<td>available in MET</td>
<td>point-to-point accuracy</td>
<td>point data, gridded data</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>YES</td>
<td>Easy to understand</td>
</tr>
<tr>
<td>Intensity-scale method</td>
<td>available in MET</td>
<td>occurrence error as a function of spatial scale and intensity</td>
<td>gridded data</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td>Scale separation of errors</td>
</tr>
<tr>
<td>Multi-scale statistical properties</td>
<td>not available in MET</td>
<td>structural properties of fields</td>
<td>gridded data</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>Spectral approach</td>
</tr>
<tr>
<td>Fuzzy verification methods</td>
<td>available in MET</td>
<td>scale- and intensity-dependent similarity of forecast to observations</td>
<td>point data, gridded data</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>Works for any kind of field</td>
</tr>
</tbody>
</table>
### Summary of spatial verification approaches (2)

<table>
<thead>
<tr>
<th>Spatial verification method</th>
<th>Status</th>
<th>What it measures</th>
<th>Observational data</th>
<th>scale error</th>
<th>location error</th>
<th>timing error</th>
<th>intensity error</th>
<th>false alarms/misses</th>
<th>selectable parameters</th>
<th>fast</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRA</td>
<td>not available in MET</td>
<td>location error and attributes of objects</td>
<td>gridded data</td>
<td>YES (indirect)</td>
<td>YES</td>
<td>NO</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>Works best on well-defined objects</td>
</tr>
<tr>
<td>MODE</td>
<td>available in MET</td>
<td>location error and attributes of objects</td>
<td>gridded data</td>
<td>YES (indirect)</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>Works best on well-defined objects</td>
</tr>
<tr>
<td>Composite approach</td>
<td>not available in MET</td>
<td>conditional error distributions</td>
<td>gridded data</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td>YES?</td>
<td>YES?</td>
<td>YES?</td>
<td>Works best on well-defined objects</td>
</tr>
<tr>
<td>SAL</td>
<td>not available in MET</td>
<td>properties of objects</td>
<td>gridded data</td>
<td>NO</td>
<td>YES</td>
<td>NO</td>
<td>YES</td>
<td>NO?</td>
<td>YES</td>
<td>YES?</td>
<td>Designed for fixed domains</td>
</tr>
<tr>
<td>Cluster analysis</td>
<td>not available in MET</td>
<td>proximity of objects</td>
<td>gridded data</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO?</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>Covers all possible scales</td>
</tr>
<tr>
<td>FQM</td>
<td>not available in MET</td>
<td>phase errors</td>
<td>gridded data</td>
<td>NO?</td>
<td>YES</td>
<td>NO</td>
<td>NO?</td>
<td>NO?</td>
<td>NO?</td>
<td>YES</td>
<td>Works for any kind of field</td>
</tr>
</tbody>
</table>
Classes of spatial verification methods seem to have different aims

Scale decomposition and fuzzy neighborhood methods
- Focus on skill quantification
- What is the forecast skill at small scales? Large scales? Low/high intensities?
- What scales and intensities have reasonable skill?
- Different fuzzy methods emphasize different aspects of skill

Object-based methods
- Focus on describing the error
- What is the error in this forecast?
- What is the cause of this error (wrong location, wrong size, wrong intensity, etc.)?

Field verification methods (morphing)
- Focus on describing phase errors
- Absence of a feature in the observations leads to peculiar convergence of distortion vectors
When can each type of method be used?

<table>
<thead>
<tr>
<th>Method</th>
<th>Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale decomposition and fuzzy neighborhood methods</td>
<td>- Whenever high density observations are available over a reasonable domain</td>
</tr>
<tr>
<td></td>
<td>- When knowing scale- and intensity-dependent skill is important</td>
</tr>
<tr>
<td></td>
<td>- When comparing forecasts at different resolutions</td>
</tr>
<tr>
<td>Object-based methods</td>
<td>- When rain blobs are well defined (organized systems, longer rain accumulations)</td>
</tr>
<tr>
<td></td>
<td>- When it is important to measure how well the forecast predicts the properties of systems</td>
</tr>
<tr>
<td></td>
<td>- When size of domain &gt;&gt; size of rain systems</td>
</tr>
<tr>
<td>Field verification (morphing)</td>
<td>- When forecasts have a fair resemblance to the observations</td>
</tr>
<tr>
<td></td>
<td>- When knowing phase errors of the field is important</td>
</tr>
</tbody>
</table>