Assessing high resolution forecasts using fuzzy verification methods

Beth Ebert

Bureau of Meteorology Research Centre, Melbourne, Australia

Thanks to Nigel Roberts, Barbara Casati, Frederic Atger, Felix Ament, Daniel Leuenberger, Urs Germann, Mike Kay, Susanne Theis, Ulrich Damrath, Daniela Rezacova
High vs. low resolution

Which rain forecast would you rather use?

Mesoscale model (5 km) 21 Mar 2004

Global model (100 km) 21 Mar 2004

Observed 24h rain

RMS=13.0

RMS=4.6
What makes a useful forecast?

- Resembles the observations on the broader scale
- Predicts an event somewhere near where it was observed
- Predicts the event over the same area (i.e., with the same frequency) as observed
- Has a similar distribution of intensities as the observations
- Looks like what a forecaster would have predicted if she'd had knowledge of the observations
"Fuzzy" verification methods

- 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 methods

- First (?) suggested by H. Brooks at 1998 Mesoscale Verification workshop
  - Brooks et al. (1998)
  - Atger (2001)
  - Damrath (2004)
  - Casati et al. (2004)
  - Theis et al. (2005)
  - Roberts (2005)
  - Rezacova et al. (2006)
Fuzzy verification framework

Fuzzy methods use one of two approaches to compare forecasts and observations:

- single observation – neighborhood forecast
- neighborhood observation – neighborhood forecast
Fuzzy verification framework

Treatment of forecast data within a window:

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

May apply to observations as well as forecasts (neighborhood observation-neighborhood forecast approach)

* Event defined here as a value exceeding a given threshold, for example, rain exceeding 1 mm/hr
Example: Fractions skill score
(Roberts and Lean 2005)

Compares fractional coverage in forecast with fractional coverage in observations

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

\[
= 1 - \frac{1}{N} \sum_{i=1}^{N} P_{\text{fcst}}^2 + \frac{1}{N} \sum_{i=1}^{N} P_{\text{obs}}^2
\]
Example: Multi-category contingency table (Atger 2001)

Compares occurrence of event in forecast with observed occurrence of event

Hit = at least one forecast event in vicinity of observed event
Moving windows

Accumulate scores as windows are moved through the domain

observation
forecast
## 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 2005)</td>
<td>NO-NF</td>
<td>Similar frequency of forecast and observed events</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>
<tr>
<td><strong>Intensity-scale</strong> (Casati et al. 2004)</td>
<td>NO-NF</td>
<td>Lower error than random arrangement of obs</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>
</tbody>
</table>

*NO-NF = neighborhood observation-neighborhood forecast, SO-NF = single observation-neighborhood forecast
Multi-scale, multi-intensity approach

- Forecast performance depends on the scale and intensity of the event

![Fractions skill score diagram](image)
Case study

- Verification of 2 km resolution precipitation forecast of 1 hr rainfall in Switzerland using MeteoSwiss Alpine Model (aLMo)

(data courtesy of Daniel Leuenberger, MeteoSwiss)
Upscaling

Decision model – Useful forecast resembles observations when averaged to coarser scales.
Fuzzy verification framework

- Good performance
- Poor performance

- Anywhere in window – ETS
- Joint probability – ETS
- Multi-event rank table – HK
- Prognostic approach – BSS
- Fuzzy logic – ETS
- Uncertainty scale – SS
- Spatial scale (deg)
- Temporal scale (deg)
- Conditional square root of FPPS
- Area related RMSE
Aggregate results for 24 h period

- **Upscaling - ETS**
  - Spatial score (log): Good performance
  - Spatial score (log): Poor performance

- **Fuzzy logic - ETS**
  - Spatial score (log): Good performance
  - Spatial score (log): Poor performance

- **Intensity-scale - SS**
  - Spatial score (log): Good performance
  - Spatial score (log): Poor performance

- **Fractions skill score - FSS**
  - Spatial score (log): Good performance
  - Spatial score (log): Poor performance

- **Proc. part. hindcast - ETS ratio**
  - Spatial score (log): Good performance
  - Spatial score (log): Poor performance

- **Cond. square root of RPS**
  - Spatial score (log): Good performance
  - Spatial score (log): Poor performance

- **Area related RNSE**
  - Spatial score (log): Good performance
  - Spatial score (log): Poor performance

Legend:
- **Good performance**
- **Poor performance**
Advantages of fuzzy verification

- Knowing which scales have skill suggests the scales at which the forecast should be presented and trusted
- Suitable for discontinuous fields like precipitation
- Can give good results for forecasts that verify poorly using exact-match approach
- Results match with our intuition
- Can be used to compare forecasts at different resolutions
- Multiple decision models and metrics
  - Direct approach → verification of intensities
  - Categorical approach → verification of binary events
  - Probabilistic approach → verification of event frequency
Many verification possibilities

- **categorical scores**
  - POD, FAR, ETS, etc.

- **probabilistic methods**
  - BS, RPS, reliability, ROC, etc.

- **continuous scores**
  - RMSE, MAE, etc.

---

Neighborhood observation – neighborhood forecast
(modeler viewpoint)

Single observation – neighborhood forecast
(user viewpoint)
Thank you!