Object identification techniques for object-oriented verification

Mike Baldwin
CIMMS/University of Oklahoma
Baldwin’s presentation on object-oriented verification

Harold Brooks

NSSL
Issues

- Object identification – how many objects do you see?
- How to characterize and measure differences between objects?
- Dealing with different numbers of observed and forecast objects
Automated rainfall object identification

- Contiguous regions of measurable rainfall (similar to Ebert and McBride 2000)
Connected component labeling

- *Pure* contiguous rainfall areas result in 34 unique “objects” in this example
Expand areas by 15%, connect regions that are within 20km

• Results in 5 objects
Useful characterization

• Attributes related to rainfall intensity and auto-correlation ellipticity were able to produce groups of stratiform, cellular, linear rainfall systems in cluster analysis experiments

• However, autocorrelation calculation is SLOW
New auto-correlation attributes

- Replaced ellipticity of AC contours with max-min correlation at specific lags (50, 100, 150km, every 10°)
Attributes

• Area (km²), lat, lon
• Mean, std dev (σ) of precip (mm) within object
• Difference between max & min correlation at 50, 100, 150km lags (Δcorr)
• Orientation angle (θ) of max correlation at 50, 100, 150km lags (E-W = 0°, N-S=90°)
• Each object is characterized by 11 attributes, with a wide variety of units, ranges of values, etc.
How to measure “distance” between objects

• How to weigh different attributes?
  – Is 250km spatial distance same as 5mm precipitation distance?

• Do attribute distributions matter?
  – Is 55mm-50mm same as 6mm-1mm?

• How to standardize attributes?
  – $X'=(x-\text{min})/(\text{max-}\text{min})$
  – $X'=(x-\text{mean})/\sigma$
  – LEPS
Decided to use LEPS

- Distance = 1 equates to difference between largest and smallest object for a particular attribute
- Linear for uniform dist (lat, lon, \( \theta \))
- Have to be careful with \( \Delta \theta \)
- L1-norm: \( d(x, y) = \sum_{i=1}^{n} |x_i - y_i| \)
## NSSL/SPC Spring Program 2004

<table>
<thead>
<tr>
<th></th>
<th>WRF-NMM</th>
<th>WRF-NCAR</th>
<th>WRF-CAPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horz/ vert grid</td>
<td>4.5km/ 35 lvls</td>
<td>4.0km/ 35 lvls</td>
<td>4.0km/ 51 lvls</td>
</tr>
<tr>
<td>Physics</td>
<td>MYJ PBL Ferrier micro, GFDL rad</td>
<td>YSU PBL, Lin et al. micro, Dudhia-RRTM rad</td>
<td>YSU PBL, Lin et al. micro, Dudhia-RRTM rad</td>
</tr>
<tr>
<td>Init cond</td>
<td>Eta (interp 40 km)</td>
<td>Eta (interp 40km grid)</td>
<td>Eta + ADAS + Level II</td>
</tr>
</tbody>
</table>

**Observed ppt = Stage II (radar-only) 4km 1h accum**
- Comparison for ~1 month (May 10 – Jun 4)
Object ID and characterization

- Remapped each model to same grid as Stage II, common domain for all
- Run object ID, get attributes
- Create database of objects meso-\(\alpha\) scale and larger \([\sim (200 \text{ km})^2]\)
How to match observed and forecast objects?

$O_1 = \text{missed event}$

$O_2$

$O_3$

$F_1$

$F_2 = \text{false alarm}$

$d_{ij} = \text{‘distance’ between } F_i \text{ and } O_j$
How to match observed and forecast objects?

$O_1 = \text{missed event}$

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

$O_2$  

$O_3$  

Objects might “match” more than once…

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

$F_1$  

$F_2 = \text{false alarm}$

$F_2$  

$F_1$
Estimate of $d_T$ threshold

- Compute distance between each observed object and all others at the same time
- $d_T = 25^{th}$ percentile = 2.5
- Forecasts have similar distributions
Example of object verf

Object identification procedure identifies 2 forecast objects and 2 observed objects.
<table>
<thead>
<tr>
<th>Attributes</th>
<th>Fcst_1</th>
<th>Fcst_2</th>
<th>Obs_1</th>
<th>Obs_2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area</td>
<td>70000 km²</td>
<td>285000</td>
<td>135000</td>
<td>70000 km²</td>
</tr>
<tr>
<td>mean(ppt)</td>
<td>0.97</td>
<td>0.32</td>
<td>0.45</td>
<td>0.60</td>
</tr>
<tr>
<td>σ (ppt)</td>
<td>1.26</td>
<td>0.44</td>
<td>0.57</td>
<td>0.67</td>
</tr>
<tr>
<td>Δcorr(50)</td>
<td>1.17</td>
<td>0.27</td>
<td>0.37</td>
<td>0.36</td>
</tr>
<tr>
<td>Δcorr(100)</td>
<td>0.99</td>
<td>0.42</td>
<td>0.54</td>
<td>0.52</td>
</tr>
<tr>
<td>Δcorr(150)</td>
<td>0.84</td>
<td>0.48</td>
<td>0.58</td>
<td>0.49</td>
</tr>
<tr>
<td>θ(50)</td>
<td>173°</td>
<td>95°</td>
<td>171°</td>
<td>85°</td>
</tr>
<tr>
<td>θ(100)</td>
<td>173°</td>
<td>85°</td>
<td>11°</td>
<td>75°</td>
</tr>
<tr>
<td>θ(150)</td>
<td>173°</td>
<td>85°</td>
<td>11°</td>
<td>65°</td>
</tr>
<tr>
<td>lat</td>
<td>40.2°N</td>
<td>47.3°N</td>
<td>39.9°N</td>
<td>44.9°N</td>
</tr>
<tr>
<td>lon</td>
<td>92.5°W</td>
<td>84.7°W</td>
<td>91.2°W</td>
<td>84.5°W</td>
</tr>
</tbody>
</table>
Distances between objects

- After transforming raw attributes to probability space (observed CDF: LEPS)
- Using L1-norm (Manhattan distance)

<table>
<thead>
<tr>
<th>Match</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fcest_1, Obs_1</td>
<td>1.48</td>
</tr>
<tr>
<td>Fcest_2, Obs_1</td>
<td>2.74</td>
</tr>
<tr>
<td>Fcest_1, Obs_2</td>
<td>2.75</td>
</tr>
<tr>
<td>Fcest_2, Obs_2</td>
<td>1.39</td>
</tr>
<tr>
<td>Obs_1, Obs_2</td>
<td>2.18</td>
</tr>
<tr>
<td>Fcest_1, Fcest_2</td>
<td>3.81</td>
</tr>
</tbody>
</table>
Average distances for matching fcst and obs objects

- 1-30h fcsts, 10 May – 03 June 2004
- Eta (12km) = 2.12
- WRF-CAPS = 1.97
- WRF-NCAR = 1.98
- WRF-NMM = 2.02
With set of matching obs and fcsts

- Nachamkin (2004) compositing ideas
  - errors given fcst event
  - errors given obs event
- Distributions of errors for specific attributes
- Use classification to stratify errors by convective mode