Data Formats and Types

Tressa L. Fowler
What you can do with MET depends on what type of data you have.
The **format** (grid, point) of your data determines your MET tool(s).

The **type** (continuous, binary) of your data determines the analyses to use within each tool.
Data Formats
Gridded Forecasts
(2D or 3D)

Dataset: d31 RIP: realtime tefe
Post: 00:00 UTC Fri 08 Dec 06
Valid: 06:00 UTC Sat 09 Dec 06 (01:00 EST Sat 08 Dec 06)

Temperature at k-index = 50

copyright 2010, UCAR, all rights reserved.
Point Observations
(2D or 3D)
Gridded Observations
(2D or 3D)

Past 24-hour accumulated precip. (water equiv inches)

Analysis valid 1200 UTC  Tue 09 Dec 2008

NCEP "Stage IV" analysis

copyright 2010, UCAR, all rights reserved.
(High resolution) Gridded Data for use with Neighborhood Methods

- Observed
- Forecast

Intensity threshold exceeded where squares are blue

Fraction = 6/25 = 0.24

slide from Mittermaier
Gridded data to transform into Objects
Examine spatial error field at different scales using wavelets.
<table>
<thead>
<tr>
<th>Data</th>
<th>MET Tool</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gridded Forecasts</td>
<td>Grid stat (traditional or neighborhood)</td>
</tr>
<tr>
<td>Gridded Observations</td>
<td>Wavelet Stat</td>
</tr>
<tr>
<td></td>
<td>MODE</td>
</tr>
<tr>
<td>Gridded Forecasts</td>
<td></td>
</tr>
<tr>
<td>Point Observations</td>
<td>Point Stat</td>
</tr>
</tbody>
</table>

copyright 2010, UCAR, all rights reserved.
Data Types
Types of Forecasts

• **Continuous**
  – Wind speed
  – Temperature

• **Categorical (includes Binary)**
  – Rain / No Rain
  – Hurricane Category 1 - 5

• **Probabilistic**
  – Prob of freezing precip

• **Ensembles**
Types of Observations

• Continuous
  – Wind speed
  – Temperature
  – Wind direction*

• Categorical (includes binary)
  – Rain / No Rain
  – Hurricane Category 1 - 5
Sometimes you may not be sure what type of data you have

• Example: Pressure in 10mb increments
  – If you have fewer than 10 distinct values, you have categories, not continuous values.

• Example: Accumulated precip amounts
  – You have a continuous variable, but you just want to verify occurrence.
  – Use thresholds to create categories from continuous measures.
<table>
<thead>
<tr>
<th>Data type</th>
<th>Analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Continuous</strong> forecasts,</td>
<td>Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Bias</td>
</tr>
<tr>
<td><strong>Continuous</strong> observations</td>
<td></td>
</tr>
<tr>
<td><strong>Continuous</strong> forecasts,</td>
<td>Receiver Operating Characteristic (ROC) curve, Kolmogorov-Smirnov</td>
</tr>
<tr>
<td><strong>Categorical</strong> observations</td>
<td></td>
</tr>
<tr>
<td><strong>Categorical</strong> forecasts,</td>
<td>Contingency table statistics and skill scores, frequency bias</td>
</tr>
<tr>
<td><strong>Categorical</strong> observations</td>
<td></td>
</tr>
<tr>
<td><strong>Probabilistic</strong> forecasts,</td>
<td>Brier score, ranked probability score (RPS), reliability diagram</td>
</tr>
<tr>
<td><strong>Categorical</strong> observations</td>
<td></td>
</tr>
</tbody>
</table>
Introduction to Standard Verification

Tressa L. Fowler
Basics

• Match up forecasts and observations at points.

• Calculate differences, sums, or counts over all the points.

• Summarize these things as statistics.
Matching Points to Grids

- Observation points are unlikely to fall exactly on forecast grid points.

- Match in horizontal space via choice of methods:
  - Closest
  - Interpolate
  - Function of surrounding points, e.g.
    - Min of closest 4
    - Median of closest 25

- Match in vertical by interpolating between level above and below.
Matching Grids to Grids

• Must use some converter to put forecasts and observations on the same grid.

  – Example: copygb
Time

• If your forecasts and observations are not at the same time, you may need to define a time window for your observations.

Forecast Time

↓

Obs       Obs

Observation Window

copyright 2010, UCAR, all rights reserved.
Now you have a bunch of forecast / observation pairs

How well do they match?

<table>
<thead>
<tr>
<th>F</th>
<th>O</th>
<th>F</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.32</td>
<td>0.03</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0.51</td>
<td>0.00</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.42</td>
<td>0.48</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0.08</td>
<td>0.14</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0.20</td>
<td>0.23</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0.75</td>
<td>0.33</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
How well do they match?

Continuous Example

\[
RMSE = \sqrt{\frac{\sum (F_i - O_i)^2}{n}}
\]

\[
r = \frac{\sum (F - \bar{F})(O - \bar{O})}{\sqrt{\sum (F - \bar{F})^2 \cdot \sum (O - \bar{O})^2}}
\]

\[
MAE = \frac{\sum |F_i - O_i|}{n}
\]

Copyright 2010, UCAR, all rights reserved.
How well do they match? Categorical Example

<table>
<thead>
<tr>
<th></th>
<th>Observed Event</th>
<th>Observed Non-event</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Forecast</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Event</td>
<td>Count = 532</td>
<td>Count = 219</td>
</tr>
<tr>
<td></td>
<td>(Hits)</td>
<td>(False Alarms)</td>
</tr>
<tr>
<td>Non-event</td>
<td>Count = 393</td>
<td>Count = 1,627</td>
</tr>
<tr>
<td></td>
<td>(Misses)</td>
<td>(Correct No’s)</td>
</tr>
</tbody>
</table>

\[
GSS = \frac{\text{hits} - \text{hits}_{\text{random}}}{\text{hits} + \text{misses} + \text{false alarms} - \text{hits}_{\text{random}}}
\]

\[
POD = \frac{\text{hits}}{\text{hits} + \text{misses}}
\]

Copyright 2010, UCAR, all rights reserved.
How well do they match? Probabilistic Example

\[ BS = \frac{1}{T} \sum (p_i - o_i)^2 \]

Reliability = \[
\frac{1}{T} \sum n_i (p_i - \bar{o_i})^2
\]

Resolution = \[
\frac{1}{T} \sum n_i (\bar{o}_i - \bar{o})^2
\]

copyright 2010, UCAR, all rights reserved.
References for information about verification