OPTIMIZING METAR NETWORK DESIGN TO REMOVE POTENTIAL BIAS IN VERIFICATION OF CLOUD CEILING HEIGHT AND VISIBILITY FORECASTS

Eric Gilleland

National Center for Atmospheric Research
Research Applications Division
INTRODUCTION

Method demonstrated for thinning a network design for verification of forecasts of cloud ceiling height and visibility for use by the general aviation (GA) community.

Data: Meteorological Aeronautical Report (METAR) data from surface stations.

Why thin?: Placement of METAR stations may bias verification analyses. For example, areas with many METAR stations may be rewarded or penalized more than areas with fewer METAR stations.
Outline

• Data

• Strategy
  – Inspecting Pair-wise Correlations vs. Distance
  – Coverage Designs
  – Percent Agreement

• Examples of Some Results

• Ongoing Work
Data used here are hourly data collected from 1590 METAR stations from October 1, 2002 through October 31, 2003.
Both datasets are theoretically continuous, but neither \textit{really} is.

<table>
<thead>
<tr>
<th>Flight rules</th>
<th>Cloud ceiling height</th>
<th>Visibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor Visibility</td>
<td>LIFR</td>
<td>&lt; 500 feet</td>
</tr>
<tr>
<td></td>
<td>IFR</td>
<td>&lt; 1000 feet</td>
</tr>
<tr>
<td>Clear Skies</td>
<td>MVFR</td>
<td>&lt; 3000 feet</td>
</tr>
<tr>
<td></td>
<td>VFR</td>
<td>&gt; 3000 feet</td>
</tr>
</tbody>
</table>
1. Inspect the relationship of distance against (pair-wise) correlations.
2. Based on results of Step 1, use a coverage design algorithm to find optimal designs of various sizes (nested). (see Nychka (1998) or Johnson (1990))
3. Use a percent agreement score (PA) to determine a reasonable design size.
Step 1: Inspecting Correlations

From a plot of distance against pair-wise correlations, much information is gleaned.

- Are the data correlated at all?
- Are most stations correlated, but a few not correlated (or less correlated)?
- Is there structure to the correlations? If so, can a correlation function, $\rho(h)$, be fit to the data?
- If a correlation function can be fit, $1 - \rho(h)$ can be used as a dissimilarity metric in a coverage design.
The same as the previous, but for ceiling height.
Example B: Cloud ceiling height.
Step 2: Coverage Designs

For a given set of candidate points, $C$, denote the set of design points as $D$ where $D \subset C$.

A distance metric between any point $x$ and a particular design, $D$, is

$$d_p(x, D) = \left[ \sum_{x' \in D} \phi^p(x, x') \right]^{1/p},$$

where $p < 0$ is a parameter, and $\phi$ is a distance or dissimilarity metric.

Note that the above is an average of “distances” between a given point $x \in C$ and each point $x' \in D$. 
An overall cover criterion is an $L_q$ average of cover points in the design region. Namely, 

$$\left[ \sum_{x \in C} d_p(x, D)^q \right]^{1/q}$$

where $q > 0$ is a parameter and $d_p(x, D)$ is as above.
Step 2: Coverage Designs

- Large negative values of $p$ tend to yield designs that are more spread out.
- The coverage criterion involves an averaging over the coverage surface. Averages of the prediction variance should correspond to coverage; especially if $\phi(\cdot) = \rho(\cdot)$.
- For $q = 1$ the average value of the surface is restored.
- When $q = \infty$ the coverage criterion is the maximum value of the surface.
- As $q \rightarrow \infty$ and $p \rightarrow -\infty$ the result gives a classic minimax design.
Step 2: Coverage Designs

- Criterion (1) is minimized over several space-filling designs of a given size to obtain a “coverage design” from among the class of space-filling designs.
- Coverage designs are generated using a “swapping” algorithm.
- It is possible to fix points in the design.
- Generally, the initial design is chosen at random, and starting design may affect outcome.
- Criterion (1) is guaranteed to converge (Nychka (1998) or Johnson (1990)).
- Here, the function `cover.design` from the R (http://www.R-project.org) package `fields` (Nychka et al.) is used to perform the algorithm.
Step 3: Percent Agreement score (PA)

A coverage design gives a “best” design for a given size. It does not say anything about the “best” design size. One reasonable approach to finding the “best” size for this particular problem is to look at how the percent agreement (PA) score is affected for different design sizes.
## Step 3: Percent Agreement score (PA)

For only 2 stations:

<table>
<thead>
<tr>
<th></th>
<th>Event</th>
<th>No event</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event</td>
<td>$a$</td>
<td>$b$</td>
<td>$a + b$</td>
</tr>
<tr>
<td>No event</td>
<td>$c$</td>
<td>$d$</td>
<td>$c + d$</td>
</tr>
<tr>
<td></td>
<td>$a + c$</td>
<td>$b + d$</td>
<td>$a + b + c + d$</td>
</tr>
</tbody>
</table>

$$PA = \frac{a + d}{a + b + c + d} \cdot 100$$
Step 3: Percent Agreement (PA)

- For each design, compute the percent agreement between stations in $D$ and all stations in $C$ (using nearest neighbors).
- Find the smallest design size that has about the same distribution of PA as all greater designs.
Results

Verification Comparison: Past results

Verification of the National C and V algorithm from 1-14 January 2003 on the 20 and 48 stations.

<table>
<thead>
<tr>
<th>stations</th>
<th>48</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>POD</td>
<td>66%</td>
<td>68%</td>
</tr>
<tr>
<td>FAR</td>
<td>10%</td>
<td>8%</td>
</tr>
<tr>
<td>Bias</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

- The POD and FAR both improved slightly (by 2%) when using fewer stations (possibly due to removing many stations in the SFO bay area; a region where making correct C and V forecasts is likely very difficult).
- Poor forecasts in this area are punished less using the 20 stations than with the 48 stations.
- Verification statistics are still quite close, and the bias is unchanged indicating that nothing was likely lost by using 20 stations instead of 48.
## Verification Comparison: Past results

Results for January 2003 on a subset of 56 stations (chosen using this strategy) in New England region.

<table>
<thead>
<tr>
<th></th>
<th>104 stations</th>
<th>56 stations</th>
</tr>
</thead>
<tbody>
<tr>
<td>POD</td>
<td>0.87</td>
<td>0.87</td>
</tr>
<tr>
<td>FAR</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>Bias</td>
<td>0.97</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Results are rounded, but clearly very close to being the same.
Ongoing Work

- Separate day and night.
- Do each month separately for each region.
- Do the analysis for more subregions.
REFERENCES


