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.

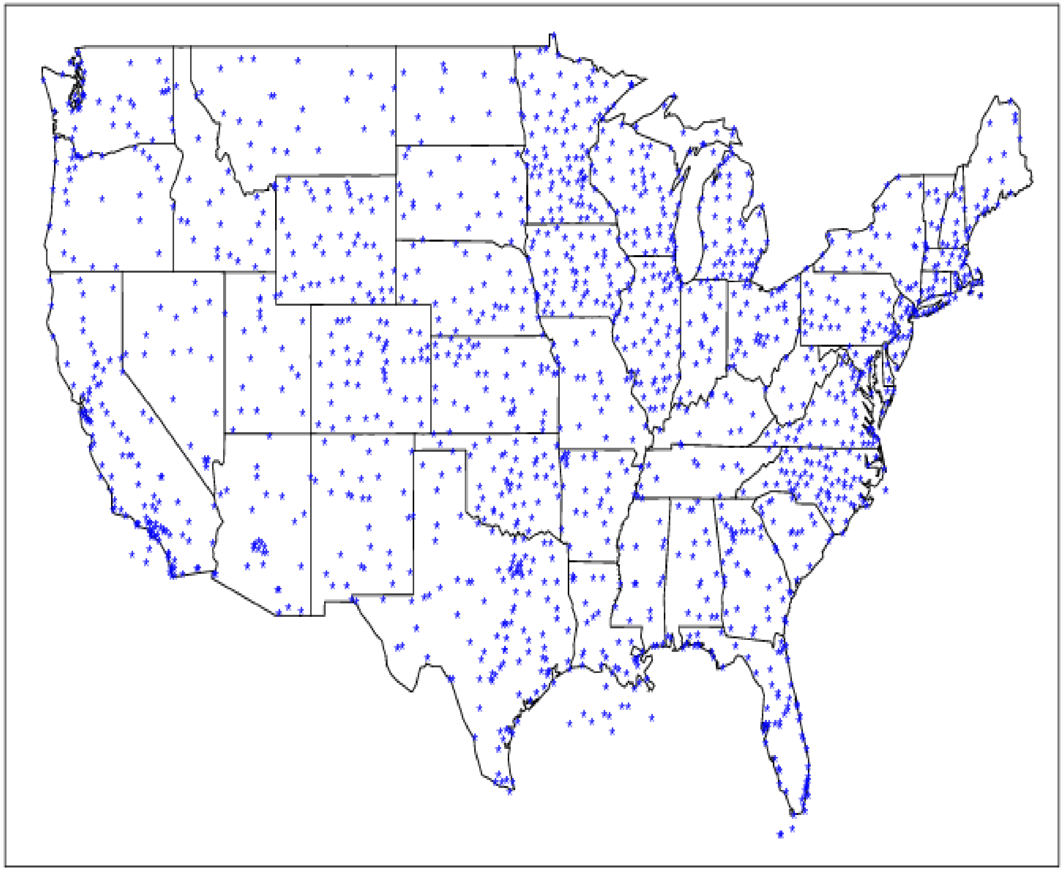
Data: Meteorological Aeronautical Report (METAR) data from surface stations.

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

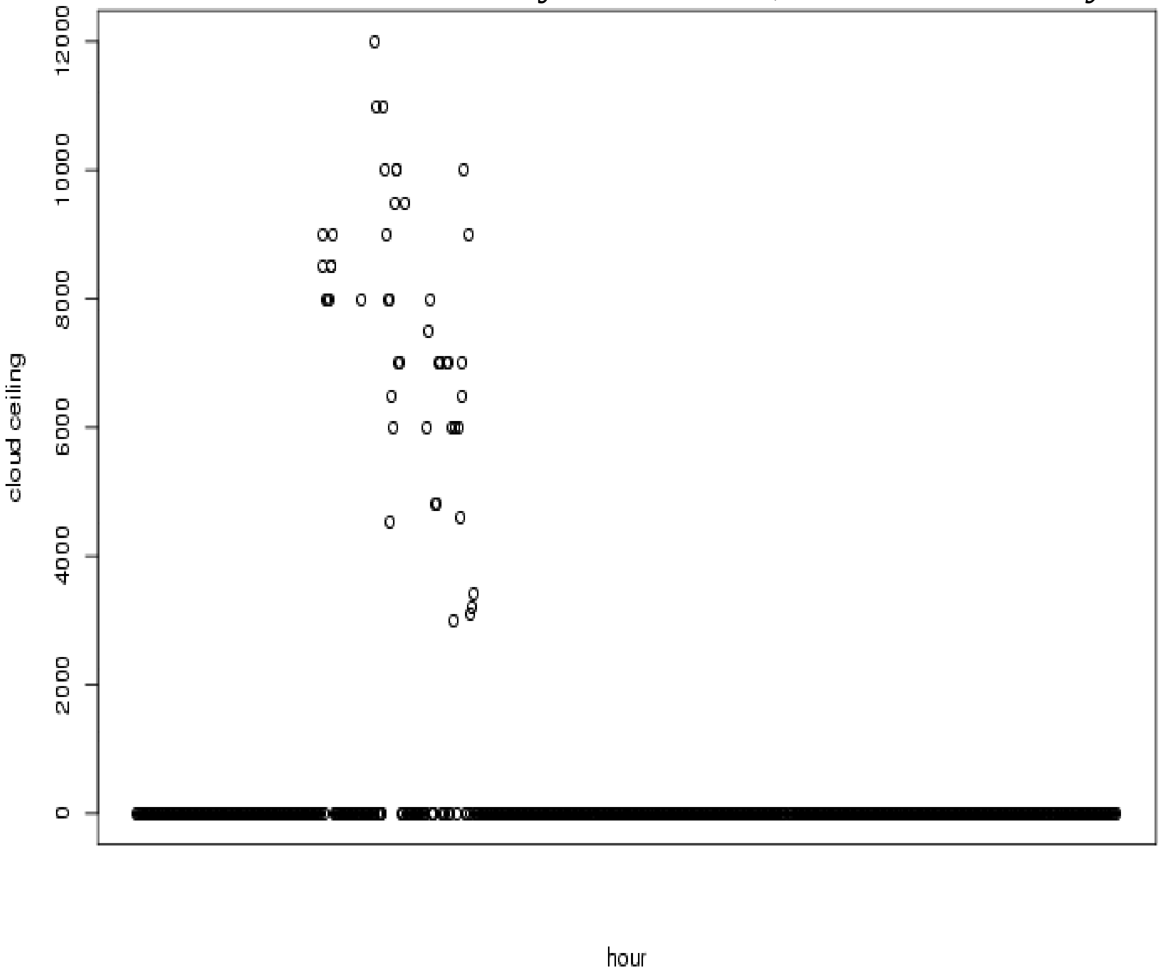
DATA



Data used here are hourly data collected from 1590 METAR stations from October 1, 2002 through October 31, 2003.

DATA

Both datasets are theoretically continuous, but neither *really* is.



Flight Rules

Discreteness in the data suggests using categories. For example, flight rules: Low Instrument Flight Rules (LIFR), Instrument Flight Rules (IFR), Marginal Visual Flight Rules (MVFR) and Visual Flight Rules (VFR).

	Flight rules	Cloud ceiling height	Visibility
Poor Visibility	LIFR	< 500 feet	< 1 mile
	IFR	< 1000 feet	< 3 miles
Clear Skies	MVFR	< 3000 feet	< 5 miles
	VFR	> 3000 feet	> 5 miles

Strategy

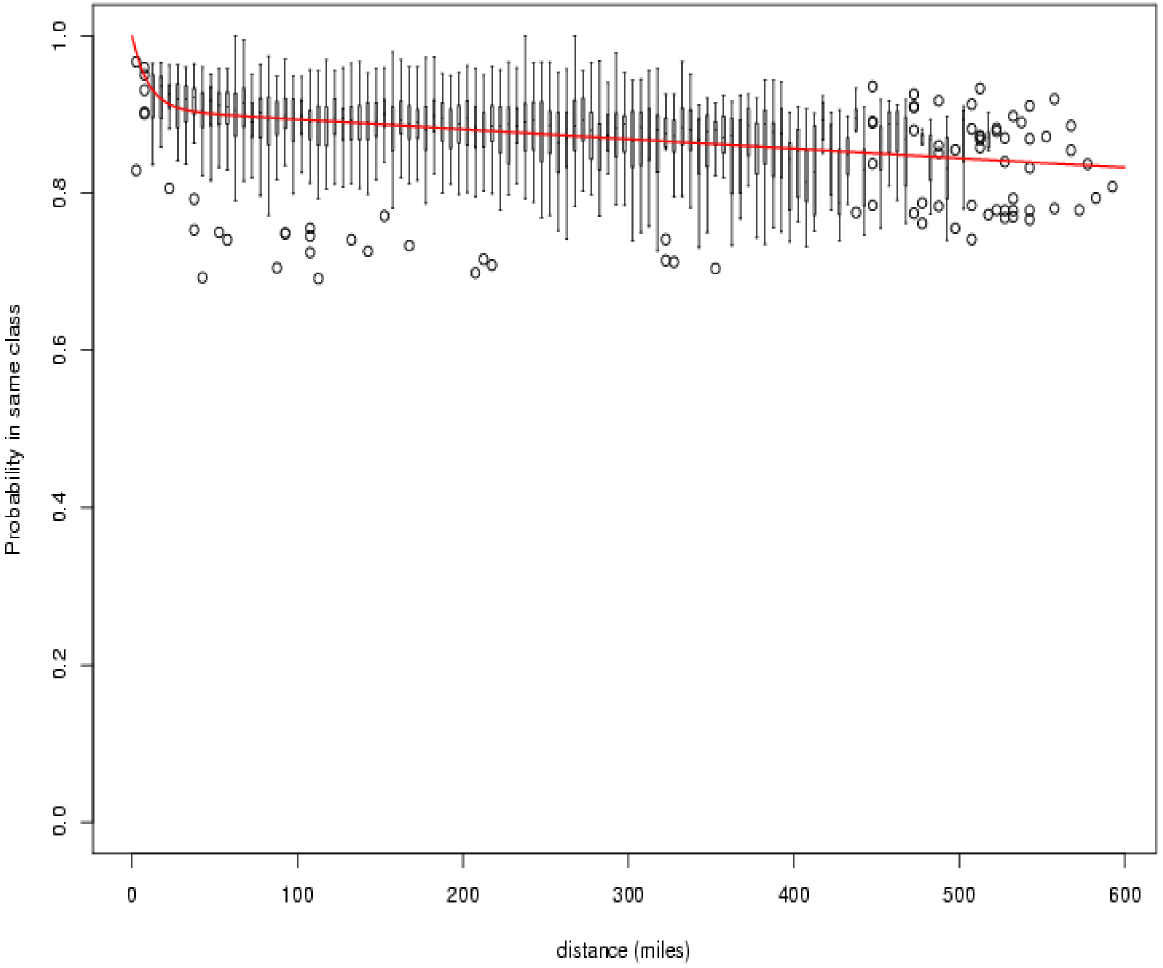
1. Inspect the relationship of distance against (pair-wise) correlations.
2. Based on results of Step 1, use a coverage design algorithm to find optimal designs of various sizes (nested). (see Nychka (1998) or Johnson (1990))
3. Use a percent agreement score (PA) to determine a reasonable design size.

Step 1: Inspecting Correlations

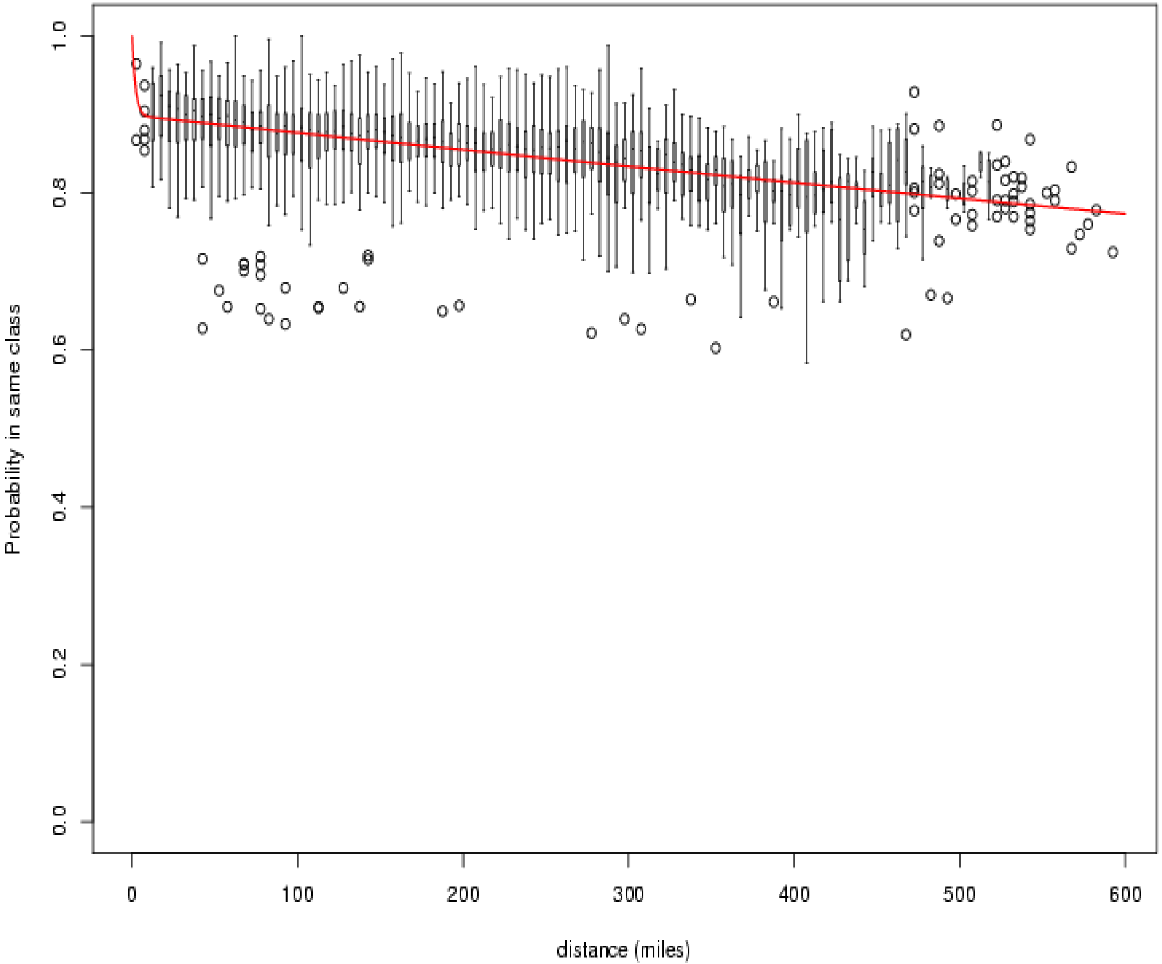
From a plot of distance against pair-wise correlations, much information is gleaned.

- 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, $\rho(\mathbf{h})$, be fit to the data?
- If a correlation function can be fit, $1 - \rho(\mathbf{h})$ can be used as a dissimilarity metric in a coverage design.

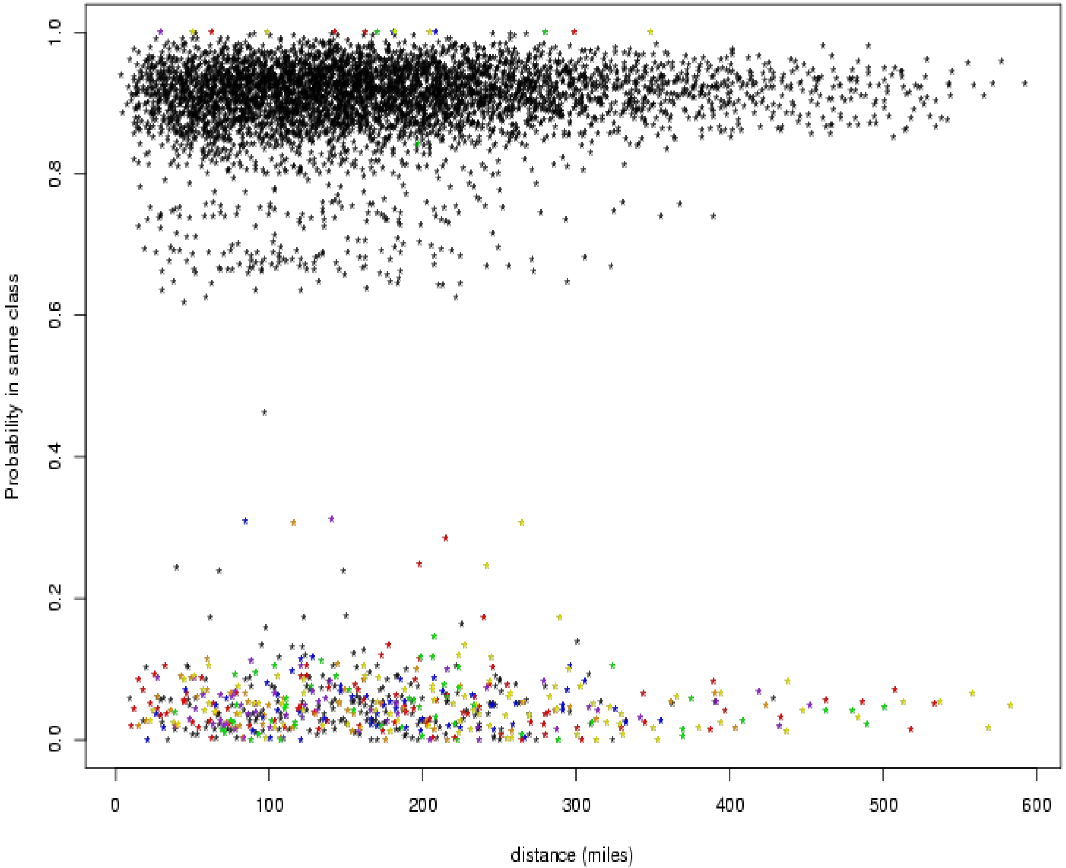
Example A: Box plots of distance against visibility correlations for New England (Oct. 1, 2002 through Dec. 22, 2002).



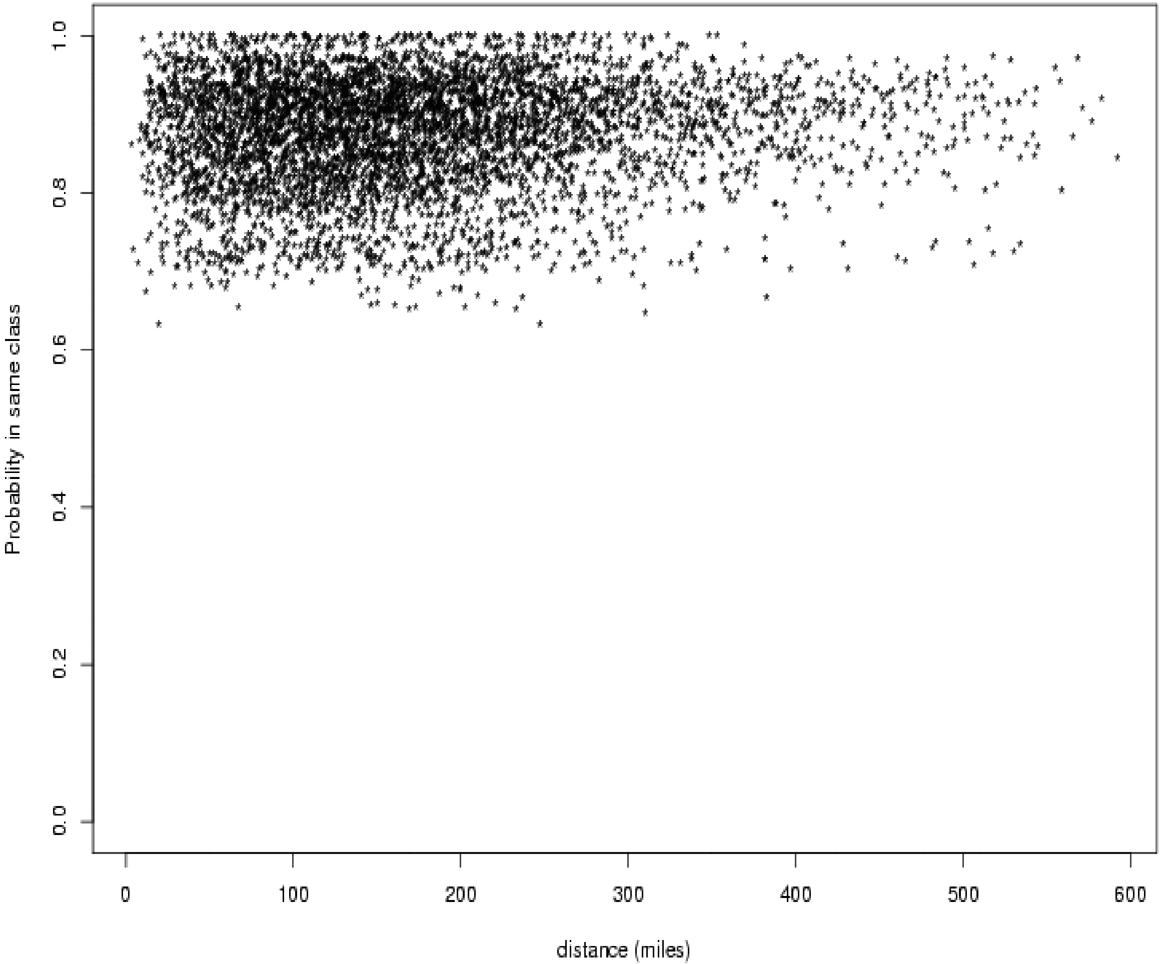
The same as the previous, but for ceiling height.



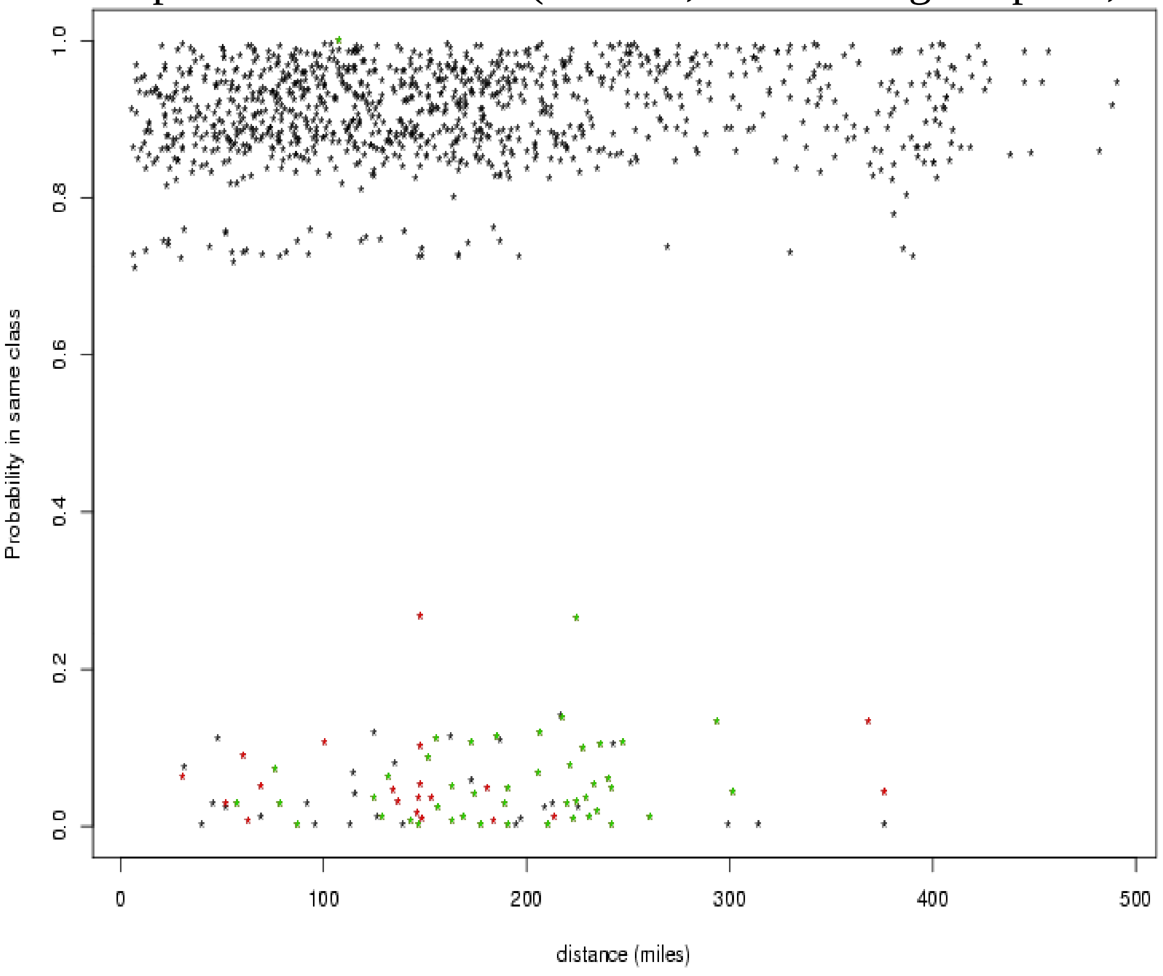
Example B: New England (March 22, 2003 through June 22, 2003) visibility.



Example B: Cloud ceiling height.



Example C: N. California (June 22, 2003 through Sep. 22, 2003).



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 $\mathbf{x}$ and a particular design, D, is

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

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

Note that the above is an average of "distances" between a given point $\mathbf{x} \in C$ and each point $\mathbf{x}' \in D$.

Step 2: Coverage Designs

An overall cover criterion is an L_q average of cover points in the design region. Namely,

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

where $q > 0$ is a parameter and $d_p(\mathbf{x}, D)$ is as above.

Step 2: Coverage Designs

- Large negative values of p tend to yield designs that are more spread out.
- The coverage criterion (1) 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 $q \longrightarrow \infty$ and $p \longrightarrow -\infty$ 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:

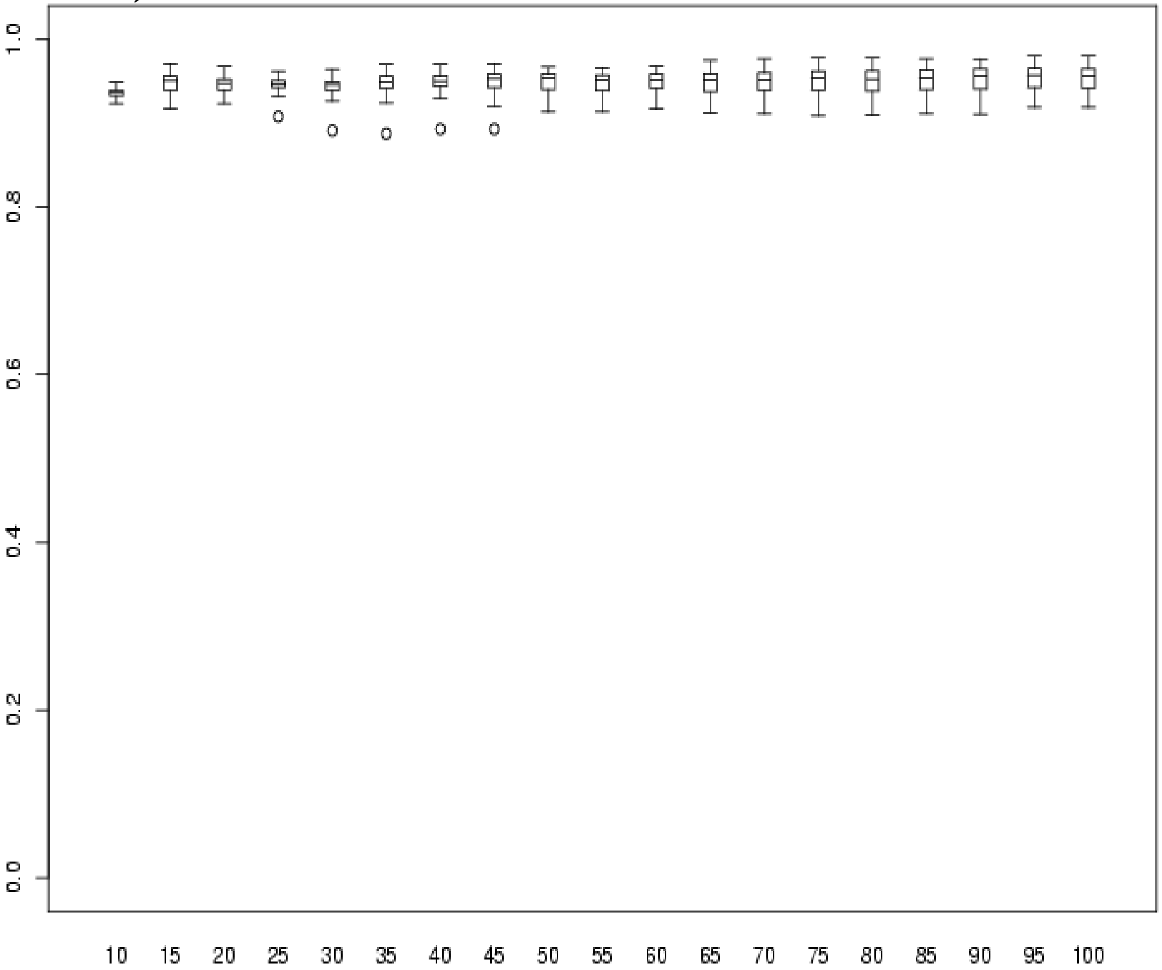
	Event	No event	Total
Event	a	b	$a+b$
No event	c	d	$c+d$
	$a+c$	$b+d$	$a+b+c+d$

$$\text{PA} = \frac{a+d}{a+b+c+d} \cdot 100$$

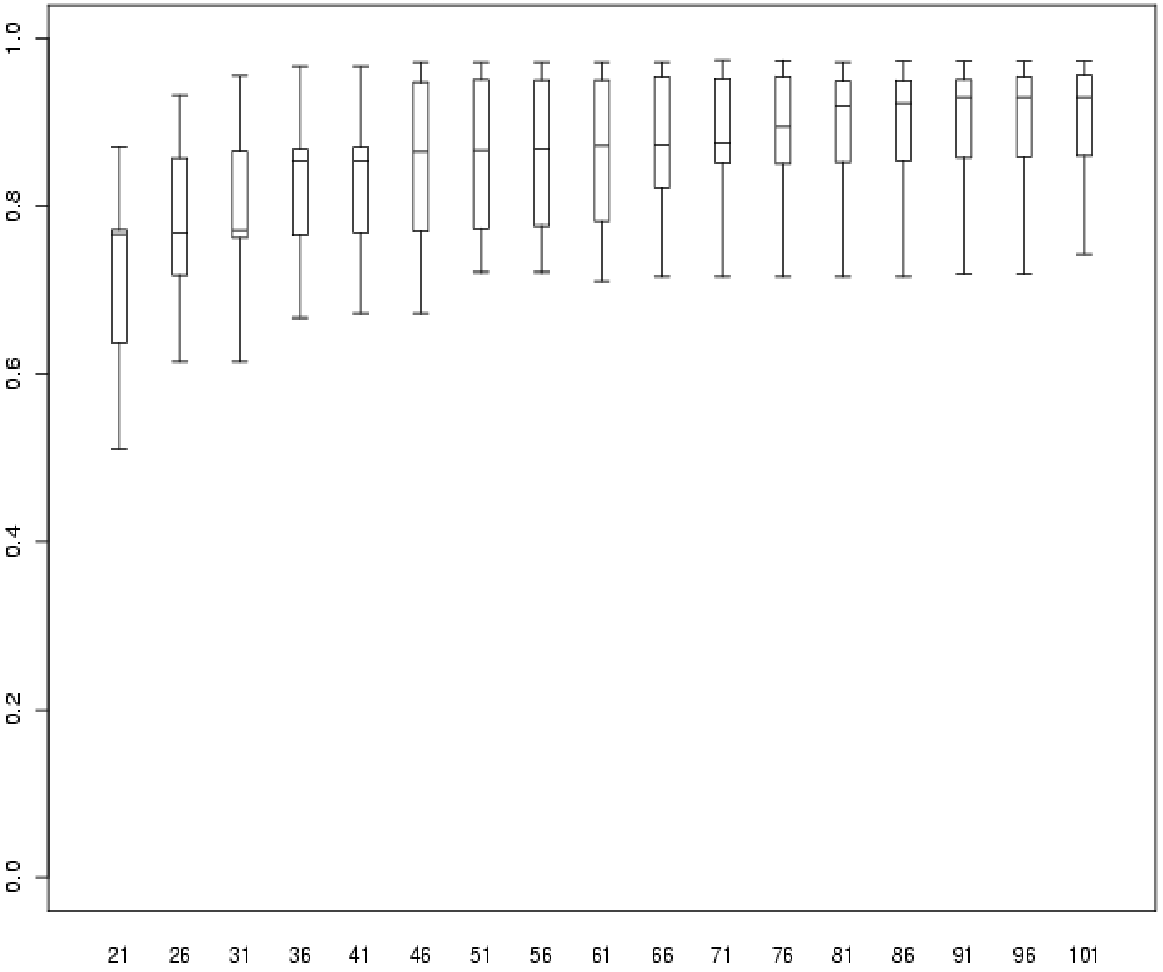
Step 3: Percent Agreement (PA)

- 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.

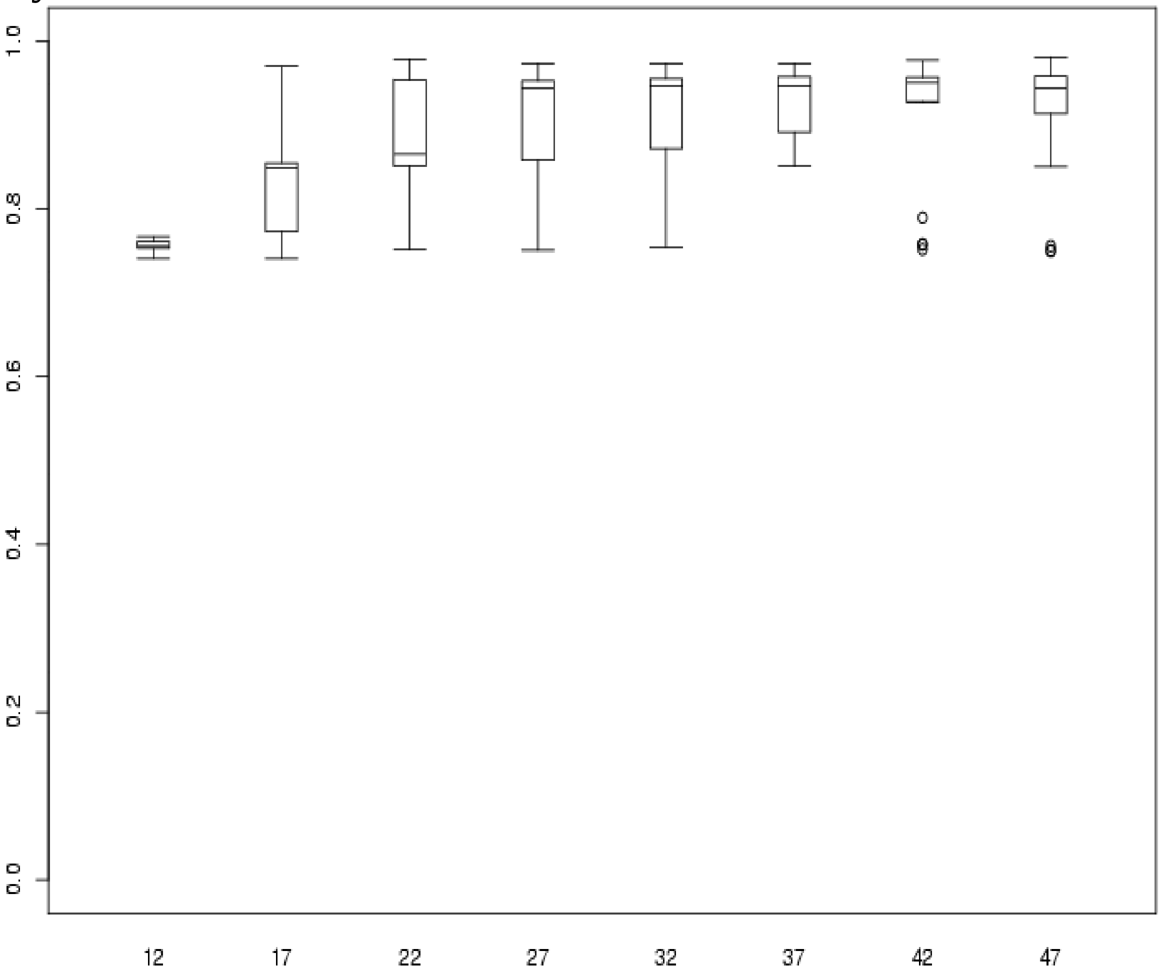
Example A: New England Oct. 1 through Dec. 22, 2002. Visibility (Ceiling very similar).



Example B: New England March 22, 2003 through June 22, 2003.

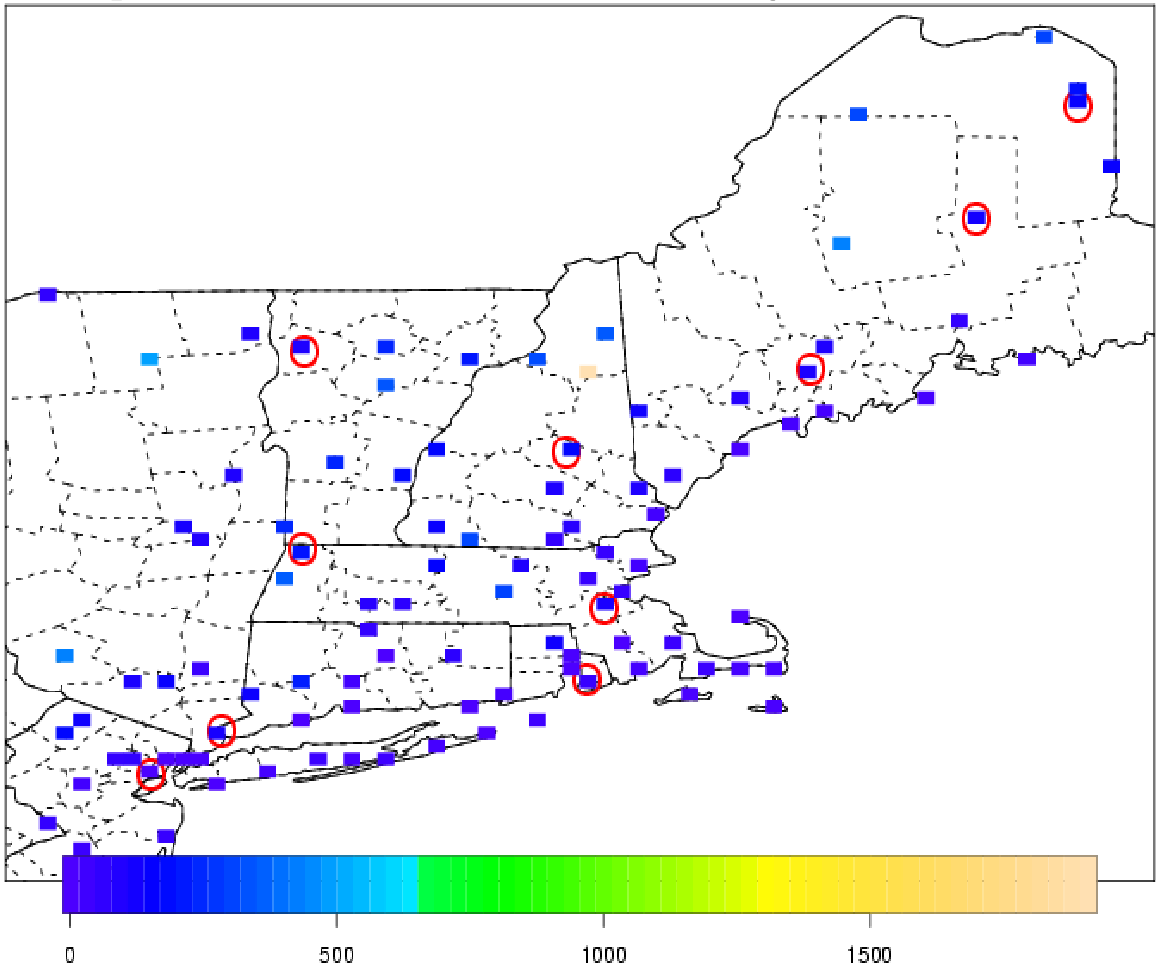


Example C: N. California subregion June 22, 2002 through Sep. 22, 2003 visibility.

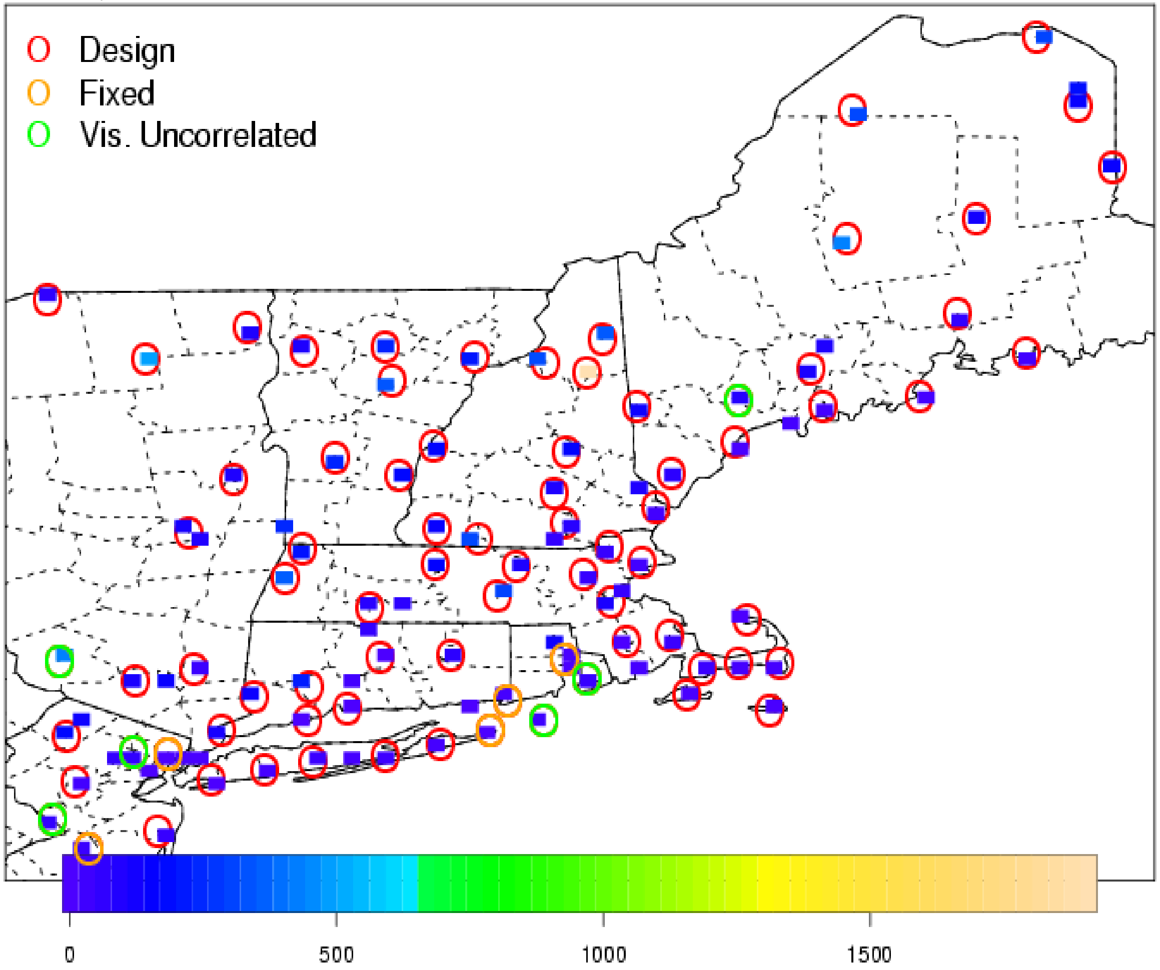


Results

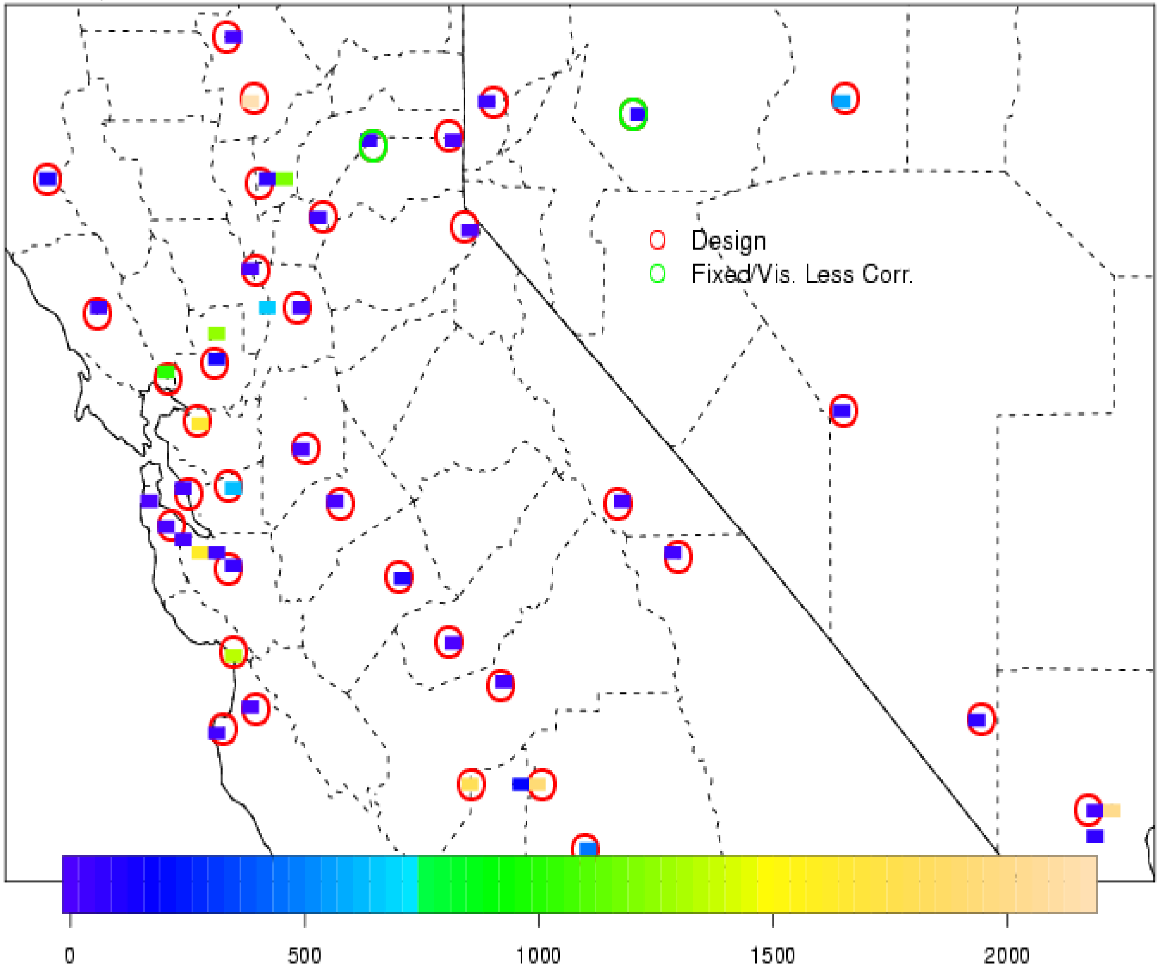
Example A: 10 station subset for New England Oct. 1, 2002 through Dec. 22, 2002.



Example B: 81 station subset for New England March 22, 2003 through June 22, 2003.



Example C: 37 station subset for N. California June 22, 2003 through Sep. 22, 2003.



Verification Comparison: Past results

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

stations	48	20
POD	66%	68%
FAR	10%	8%
Bias	1	1

- 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.

	104 stations	56 stations
POD	0.87	0.87
FAR	0.10	0.10
Bias	0.97	0.97

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

Johnson, M.E., Moore, L.M., and Ylvisaker, D., 1990: Minimax and maximin distance designs, *Journal of Statistical Planning and Inference*, 26, 131-148.

Nychka, Douglas and Saltzman, Nancy, 1998: Design of Air Quality Monitoring Networks, *Lecture Notes in Statistics: Case Studies in Environmental Statistics, Springer*, 175 Fifth Avenue, New York, NY 10010.

Nychka, D., Bailey, B., Ellner, S., Haaland, P., O'Connell, M., Hardy, S., Baik, J., Meiring, W., Royle, J.A., Fuentes, M., Hoar, T., Tebaldi, C. and Gilleland, E., 2003: fields: a collection of programs based in R/S for curve and function fitting with an emphasis on spatial data, *http://www.cgd.ucar.edu/stats/Software/Fields/*.

R Development Core Team, (2004), R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-00-3, `http://www.R-project.org`

Wilks, Daniel S., 1995: Statistical Methods in the Atmospheric Sciences, *Academic Press*, 525 B Street, Suite 1900, San Diego, CA 92101-4495.