2.4 Selecting METARs to Verify Ceiling and Visibility Forecasts

Tressa L. Fowler*, Jamie T. Braid, and Anne Holmes
National Center for Atmospheric Research, Boulder, Colorado

1 INTRODUCTION

METAR station information is used for verification of ceiling and visibility forecasts. Unfortunately, METAR stations are sparse in many areas and quite dense in others. Additionally, the quality and type of information available from METAR stations may differ considerably depending on whether the station is automated or manual, the location, the type of instrumentation, the level of training and expertise of a manual observer, etc. Thus, in order to reduce the effect of these differences on verification, it is desirable to identify a consistent subset of METAR stations to be used for verification.

A comparison of stations with their neighbors can aid in the identification process. All METAR stations in the Continental US (CONUS) are matched with their neighbors within a radius of 20 and 40 kilometers (km). A subset of stations near the San Francisco area is also analyzed in order to facilitate comparison to the work done by Gilleland (2004) (hereafter, Gilleland). The details about the METAR observations are included in Section 2. The agreement of ceiling and visibility observations between the stations is assessed, with a detailed description of the methods in section 3. Section 4 presents the results of the agreement analyses. Disagreement may imply that one station has an incorrect observation or that the environmental conditions differ considerably over the distance between the stations. In the former case, the “bad” or inconsistent observation should be eliminated. In the latter, both observations should be included to accurately represent the small-scale variation in conditions. These and other conclusions are presented in section 5.

2 DATA

METAR observations of ceiling and visibility from the CONUS from January 1st through 30th, 2003 were used to determine flight category. Flight categories include Low Instrument Flight Rules (LIFR), Instrument Flight Rules (IFR), Modified Visual Flight Rules (MVFR), and Visual Flight Rules (VFR). The ceiling and visibility measurement resulting in each flight category are shown in Table 1. Ceiling and visibility measurements at METAR stations can be made by instruments or human observers. The reporting frequency varies from station to station. Human observations tend to be available only during daylight hours while instruments measure around the clock. The METAR stations that reported ceiling and visibility observations during this time period are shown in Figure 1.

* *Corresponding author address*: Tressa L. Fowler, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307 e-mail: tressa@ucar.edu

Table 1: Ceiling and Visibility bounds for each flight category

	Ceiling	Visibility
LIFR	Less than 500 ft	Less than 1 mi
IFR	Between 500 and 1000 ft	Between 1 and 3 mi
MVFR	Between 1000 and 3000 ft	Between 3 and 5 mi
VFR	Greater than 3000 ft	Greater than 5 mi

3 METHODS

METAR observations of flight category based on ceiling and visibility were matched to their neighbors within 20 km and 40 km. While most stations have only a few neighbors within a 20 km radius, some have many neighbor stations. The number of stations varies from time to time because not all stations report with the same frequency. Thus, a METAR station will be matched with only those stations reporting during the same time period.

Research conducted at the Naval Research Lab (NRL, 2002) suggests that comparing observations of ceiling and visibility to forecasts further than the closest 4 grid points was inappropriate. Thus, there is no reason to look further than 40 km because verification will focus on forecasts with a 20 km grid spacing.

Some stations may have no neighbors within 20 or 40 km, and thus no comparisons are possible. The 20 km radius restriction yielded 378 stations that had at least one neighbor while the 40 km radius yielded 917 stations with neighbors. The METAR stations in North and South Dakota, Montana, Nevada, central Oregon, west Texas and central Tennessee are few and far between. In these areas, it would seem sensible to use whatever observations are available for verification of forecasts.

The stations in the San Francisco area were analyzed using only the 40 km radius, as too few of the stations had neighbors within a 20 km radius.

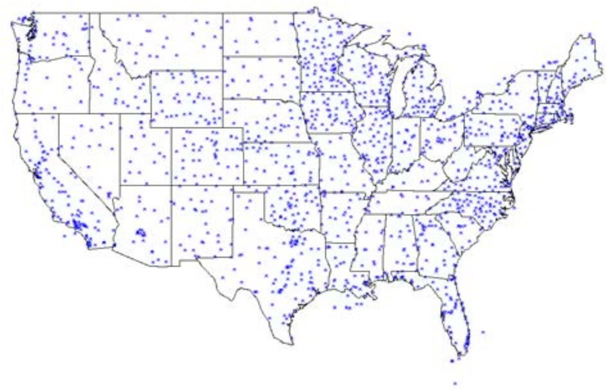

Figure 1: Locations of 1557 METAR stations reporting during January 2003.

While neighbor station comparisons are not possible, they are also not really necessary in sparse areas. In part, the neighbor comparisons will be used to eliminate stations that duplicate information from nearby stations. If there are no nearby stations, then comparisons are not meaningful. A nearest neighbor comparison could have been employed in place of the 20 and 40 km radius comparison, but for many stations the nearest neighbors are quite a distance away. This distance would have made the results more difficult to interpret. The comparisons will also help to identify "bad" or "unusual" stations, but this is not possible in areas with sparse stations coverage where there is usually only a single observation.

Table 2: Contingency table values used to compute measures of agreement between target and neighbor METAR stations.

Target METAR station	Neighbor METAR station		
	Ceiling < 1 Kft and vis < 3 mi (IFR or LIFR)	Ceiling < 1 Kft and vis > 3 mi (MVFR or VFR)	Total
Ceiling < 1 Kft and vis < 3 mi (IFR or LIFR)	a	b	a+b
Ceiling > 1 Kft and vis > 3 mi (MVFR or VFR)	c	d	c+d
Total	a+c	b+d	n

The observation at each station was compared to the observations from the surrounding stations. The percent agreement between the target station and its neighbor(s) was calculated overall and separately for ceiling and visibility events (LIFR and IFR conditions) and non-events (MVFR and VFR conditions). The agreement between the target and its neighbors is used as a measure of consistency between neighboring stations. Stations with high percent agreement are consistent with their neighbors. Those stations that fail to agree with their neighbors (i.e. have low percent agreement) may have some type of error in the observation or may represent very different atmospheric conditions due to altitude, terrain, microclimate and other variations.

The percent agreement between a target station and its neighbors is calculated from the standard 2x2 contingency table shown in Table 2. The rows and columns represent the ceiling and visibility events (IFR or worse) and non-events (MVFR or better) at the target and neighbor stations, respectively. The overall percent agreement (PA) measures the agreement for both categories together, and is represented by the following equation.

$$PA = (a+d)/n \quad \text{Eqn. 1}$$

The event (PA_E) and non-event (PA_N) percent agreement measures the conditional agreement from the target station, given that the neighbor station observed an event or non-event. The percent agreement for events and non-events are calculated from the equations 2 and 3, respectively.

$$PA_E = a/(a+c) \quad \text{Eqn. 2}$$

$$PA_N = d/(b+d) \quad \text{Eqn. 3}$$

The reason for calculating the event and non-event agreement separately is that in many locations non-events may far outnumber events. In this case, the overall and non-event percent agreement may be high while the event percent agreement may be low.

The percent agreement statistics for events and non-events are more commonly known as probabilities of detection (POD) for events and non-events. However, this nomenclature is typically applied when there is a single observation of "truth" being compared to a forecast. Here, the comparison is between two different observations, and thus "percent agreement" more accurately describes what is being measured. For more information on the properties of POD, percent agreement and other statistics calculated from contingency tables, see Wilks (1995) and Brown and Young (2000).

4 RESULTS

Results for the CONUS using both a 20 and 40 km radius to define neighbors are given in Section 4.1. Results for stations within the San Francisco area using a 40 km radius are presented in Section 4.2.

4.1 Results for CONUS

The percent agreement between the CONUS METAR stations and their neighbors within 20 and 40 km is plotted in Figures 2 and 3, respectively. The red colored locations have the lowest agreement and the blue have the greatest, as shown in the color bars below the figure. The central portion of the country seems to have a high percent agreement between stations while the west coast and northeast regions have some stations with lower percent agreement between stations. Comparison of the two figures illustrates that stations had higher agreement with their neighbors within 20 km than 40 km. Given the spatial variability of ceiling and visibility events, this comes as no surprise.

The percent agreement for non-events looks quite similar to the overall percent agreement. This result is not unexpected since non-events are more frequent than events. Thus, non-events make a greater contribution to the overall percent agreement than events.

The percent agreement for events shown in Figures 6 and 7 tells quite a different story than the previous four figures. The color scales used in Figures 6 and 7 are different than those used in the first four figures. This was necessary because the percent agreement for events is generally much lower than for overall or for non-events. Use of the same color scale for all figures resulted in some figures failing to distinguish between stations.

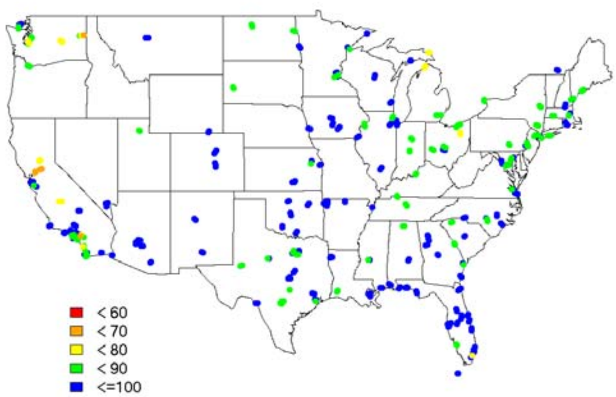

Figure 2: Overall percent agreement (PA) for METAR stations as observed by neighboring stations within 20 km.

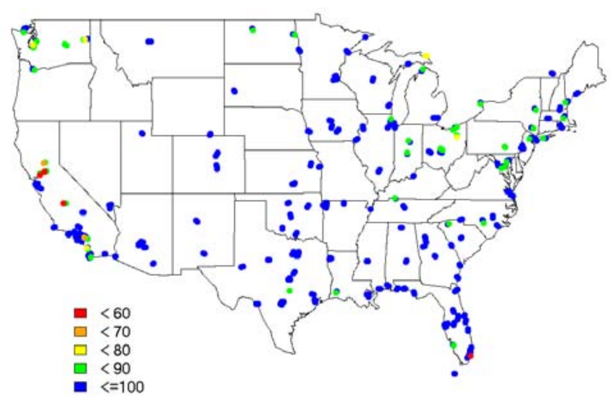

Figure 4: Percent agreement for ceiling and visibility non-events (PA_N) for METAR stations as observed by neighboring stations within 20 km.

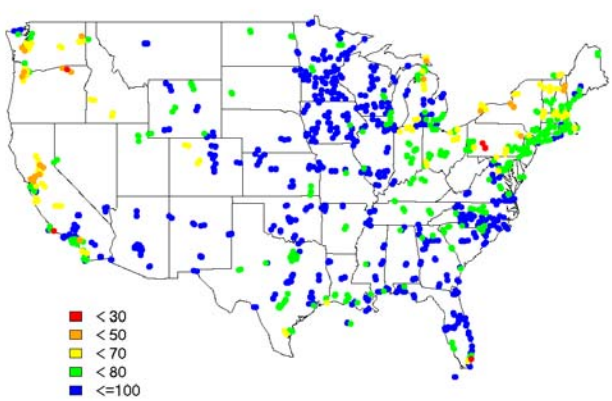

Figure 3: Overall percent agreement (PA) for METAR stations as observed by neighboring stations within 40 km.

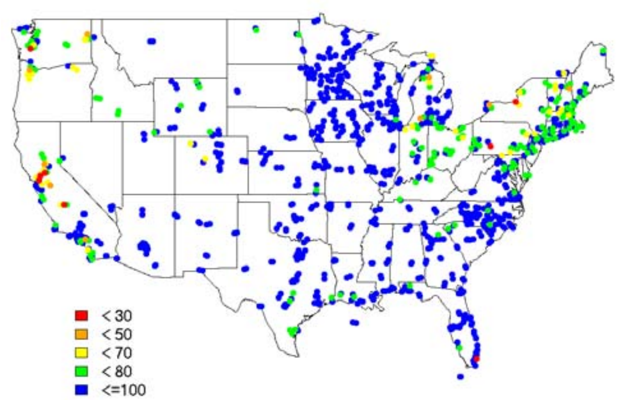

Figure 5: Percent agreement for ceiling and visibility non-events (PA_N) for METAR stations as observed by neighboring stations within 40 km.

METAR station observations were much less likely to agree with their neighbors during events than non-events. The agreement does not appear to be regional in nature, as all areas show an assortment of colors. For instance, the front range of Colorado has some stations with very high agreement (blue) and some with very low agreement (red). The 40 km plot (Figure 7) shows a high concentration of yellow and green dots (agreement between 50 and 80 percent) in the Northeast and Great Lakes regions.

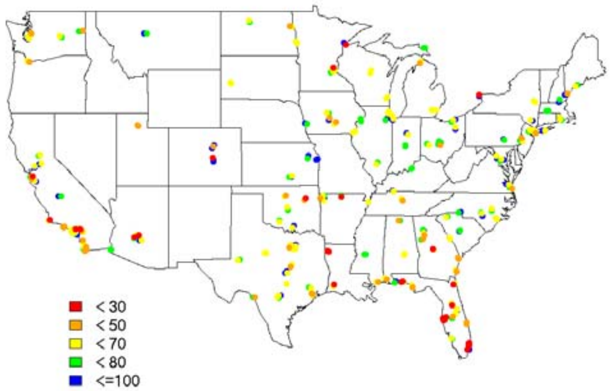

Figure 6: Percent agreement for ceiling and visibility events (PA_E) for METAR stations as observed by neighboring stations within 20 km.

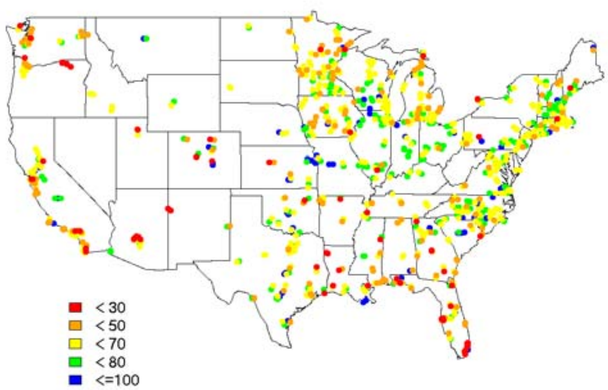

Figure 7: Percent agreement for ceiling and visibility events (PA_E) for METAR stations as observed by neighboring stations within 40 km.

4.2 Results for the San Francisco Area

The results for the San Francisco Bay area were also examined separately, in order to facilitate comparison with the work done by Gilleland. The overall (not shown) and non-event percent agreement (Figure 8) look very similar, and indicate fairly high rates of agreement among stations surrounding the San Francisco Bay. The inland stations tend to agree less well with their neighbors. However, the percent agreement statistics (Fig. 9) for ceiling and visibility events for the stations near the Bay Area are considerably lower, and a couple of stations in Oakland have very low agreement. The cause of disagreement between the Oakland stations and their neighbors will need to be investigated. However, it is possible that there are clear conditions in Oakland while neighboring stations across the Bay have low ceilings and/or visibility.

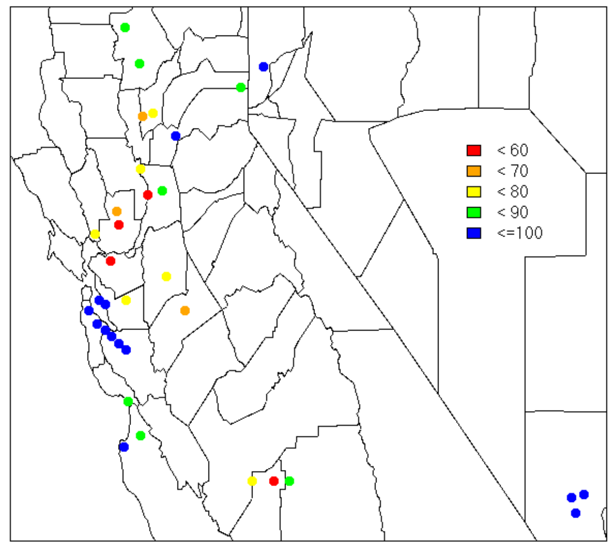

Figure 8: Percent agreement for non-events (PA_N) of METAR stations within a 40 km radius in San Francisco area.

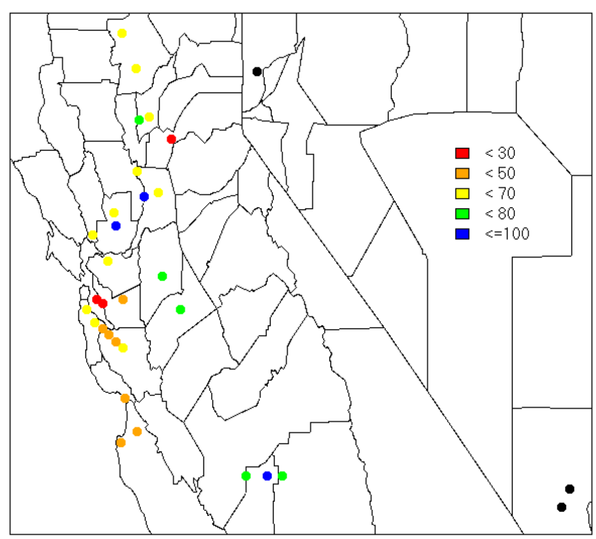

Figure 9: Percent agreement for ceiling and visibility events (PA_E) of METAR stations within a 40 km radius in San Francisco area.

Two coverage designs (i.e. selected subsets of stations) are presented by Gilleland. The first design includes a station in Oakland and one south of the bay near San Jose but no other stations in the immediate vicinity of the San Francisco Bay. Notably, the stations on the San Francisco peninsula are all excluded. The second design includes one station on the peninsula in addition to one in Oakland and one near San Jose. Based on the agreement statistics, this second design seems preferable to the first, as the conditions in Oakland, on the peninsula, and near San Jose during ceiling and visibility events may frequently be different.

Some of the stations, especially those in Nevada, analyzed by Gilleland do not appear in Figures 8 and 9. These stations did not have any neighbors within 40 km, and thus there are no percent agreement statistics to plot for those stations. The stations included here were all included in Gilleland.

5 CONCLUSIONS AND FUTURE WORK

During ceiling and visibility events, stations tended to agree less well with their neighbors than during non-events. This appeared to be true both in the CONUS and among the subset of stations in the San Francisco area. Stations in the northeastern part of the US tended to agree less well with their neighbors during non-events than stations in other parts of the country, but seemed to agree somewhat better during events than stations in other parts of the country. Many of the stations that did not agree well with their neighbors were located along the coasts. Few conclusions can be drawn about the central portion of the country due to the sparseness of stations in this area.

Based on the analyses in this paper, the second coverage design in Gilleland represents a subset of stations that preserve information about small scale differences while eliminating stations that have similar information to their neighbors that are included in the design.

Overall, the observations at METAR stations agreed well with the observations taken at their neighbors. Should this agreement prove to be consistent over different seasons, then selection of a set of METAR stations should prove quite simple as the selected stations should give much the same information as those that are excluded from verification. Those few stations that fail to agree well with their neighbors will need to be examined to determine the cause of the disagreement.

More seasons need to be examined, so the verification will not be biased toward seasonal (in this case, winter) characteristics. Additionally, a balance must be struck between eliminating spatial biases and keeping stations that show important small scale variability. Although different sets of stations may be best for different seasons, it is essential to select one set for consistent use by the automated verification system, the Real-Time Verification System (RTVS) (Mahoney *et al*., 1999). This system will perform verification of the ceiling and visibility forecasts using the set of selected stations. Since the system is automated, it is best not to change it with each season. Also, results from different seasons can be compared if the same set of stations is used for verification. However, if different sets of stations are used, then seasonal results are not comparable.

ACKNOWLEDGEMENTS

NCAR is sponsored by the National Science Foundation. This research is in response to requirements and funding by the Federal Aviation Administration (FAA). The views expressed are those of the authors and do not necessarily represent the official policy of the FAA.

REFERENCES

Brown, B. G., and G. S. Young, 2000: Verification of icing and icing forecasts: Why some verification statistics can't be computed using PIREPs. Preprints, 9th Conference on Aviation, Range, and Aerospace Meteorology, Orlando, FL, 11-15 Sept., American Meteorological Society (Boston), 393-398.

Gilleland, E., 2004: Improving Forecast Verification through Network Design. Preprints, 17th Conference on Probability and Statistics in the Atmospheric Sciences, Seattle, WA, 11-15 January, American Meteorological Society (Boston).

Mahoney, J.L., J.K. Henderson, B.G. Brown, and Craig S. Hartsough, 1999: Benefits of using FSL's Real-Time Verification System (RTVS) at the NWS/NCEP Aviation Weather Center. *Preprints, 8th Conference on Aviation*, Range, and Aerospace Meteorology, Dallas, 10-15 January. American Meteorological Society, Boston, 95-99.

Naval Research Lab (NRL), 2002: Grid point usage for verification of model ceiling forecast. Summary report to FAA AWRP Ceiling and Visibility PDT.

Wilks, D., 1995: Statistical Methods in the Atmospheric Sciences. Academic Press, San Diego.