USE OF AUTOMATED OBSERVATIONS FOR VERIFICATION OF TURBULENCE FORECASTS

Barbara G. Brown[1], Jennifer L. Mahoney[2], Robert Sharman[1], Jamie Vogt[1], and Judy Henderson[2,3]

October 2000

1. Introduction

Two types of automated observations of turbulence conditions are available or are becoming available: automated vertical accelerometer (AVAR) observations and *in situ* measurements of eddy dissipation rate (EDR). The AVAR observations, which have been available for several years on a number of United Airlines aircraft, provide measurements of an aircraft's vertical acceleration without correction for aircraft characteristics or motions. The *in situ* EDR observations are based on a transformation of the observed vertical acceleration to obtain a measurement of atmospheric turbulence that is independent of aircraft characteristics and motions. Software required to provide the *in situ* observations is currently being installed on a number of commercial aircraft, and is available from more than 65 United Airlines aircraft in a test mode.

Ideally, these data will be very valuable for evaluation of current and new systems for forecasting turbulence conditions. Currently only pilot reports (PIREPs) and, to a limited extent, AVAR observations have been used for verification of these forecasts. Because the EDR observations differ in many respects from PIREPs, it is important to evaluate characteristics of these reports and to consider possible approaches for their future inclusion in forecast verification studies. In particular, the form of the observations may lead to the desirability of computing different types of statistics. Moreover, the spatial and temporal characteristics of the observations may lead to certain caveats regarding their use in verification studies.

---

[1] Research Applications Program, National Center for Atmospheric Research, Boulder, CO 80307-3000
[2] Forecast Systems Laboratory, Environmental Research Laboratories, National Oceanic and Atmospheric Administration, Boulder, CO
[3] Joint collaboration with the Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO

Thus far, the AVAR observations have only been considered as observations of no turbulence, when the AVAR measurement falls within a specified range of values (generally, within $\pm$ 20% of the acceleration of gravity, 9.8 ms^{-2}). Recently, however, we began to explore the possibility of inferring aircraft movements – changes in direction and altitude – from the time series of locations of AVAR observations. By filtering out time periods when the aircraft was maneuvering, the remaining AVAR observations should represent time periods that are unaffected by such motions, and it should be possible to interpret these observations as indicating true turbulence effects. As part of the effort to verify the potential for using these filtered AVAR observations, we also felt it was important to evaluate the relationship between the selected AVAR observations and co-located PIREPs. In addition, we believed that it is important to understand characteristics of the AVAR observations prior to further investigation of the use of the *in situ* EDR observations as verification data.

Thus, this report considers several different aspects related to the use of automated observations for verification of turbulence forecasts: (i) relationships between AVAR observations and PIREPs; (ii) statistical characteristics of recent *in situ* EDR observations; and (iii) concepts for use of the EDR observations for verification of turbulence forecasts.

2. Relationships between AVAR observations and PIREPs

AVAR observations and PIREPs for the period November 1998 through March 1999, and November 1999 through March 2000 were included in this study. The AVAR observations were filtered to remove observations that might have been associated with aircraft maneuvers (i.e., vertical or horizontal changes in direction; acceleration). Only AVAR and PIREP observations above 20,000 ft were included. The AVAR observations were matched to PIREPs reporting Yes turbulence or No turbulence that were located within a particular time window and spatial distance around the AVAR observation. The time/space windows that were considered include (i) within 70 km and 10 min; and (ii) within 40 km and 2 hr. In all cases the elevation window was 1,000 ft. The PIREPs also were subdivided by weight, so that only heavy aircraft (greater than 27,000 lb empty weight) were included in some analyses.

Figures 1 and 2 show histograms of the AVAR observations as a function of the turbulence severity reported by the co-located PIREP. The AVAR values have been converted to absolute acceleration relative to the acceleration of gravity, so that larger values should represent more intense turbulence. However, it is clear from Figs. 1 and 2 that there is very little, if any, relationship between the AVAR observations and the turbulence severity reported by the PIREPs, regardless of which space/time matching constraints were used, and whether the comparisons were restricted to heavy aircraft.

Turbulence reports from PIREPs also were compared to the turbulence severities reported by nearby PIREPs, using the same matching process, and these results are shown in Figs. 3-4. It is reassuring to observe from these figures that co-located PIREPs generally report quite similar levels of turbulence severity. In contrast to the results shown in Figs. 1 and 2, the PIREP severity values shown in Figs. 3-4 are very strongly related. In particular, the peak frequencies in the histograms match the severity reported by the co-located PIREP.

The results in Figs. 1-2 led to a great deal of concern regarding what the AVAR observations represent and how they should be used. In particular, the distribution of AVAR values when a co-located PIREP reports no turbulence is about the same as the distribution when a PIREP reports moderate or severe turbulence. These results indicate that the AVAR observations *should not be used even to represent no-turbulence conditions*.

This result is contrary to what has been assumed about the AVAR observations over the last several years. However, in subsequent discussions with Carl Knable (United Airlines), it was discovered that each AVAR observation represents the maximum vertical acceleration measured over the last twelve seconds before the data is downlinked. That is, it does not represent the full ten minutes between observation times, as has been assumed in all previous analyses. Because the sampling period (twelve seconds) is so short, it is not surprising that the AVAR values have little relationship to other measurements of turbulence. These results suggest that the AVAR observations should not be used for verification of turbulence forecasts. Fortunately, the EDR observations are collected over the full data rate of the vertical acceleration sensor (480 times per second), so these observations will not be subject to the same types of concerns (Carl Knable, personal communication).

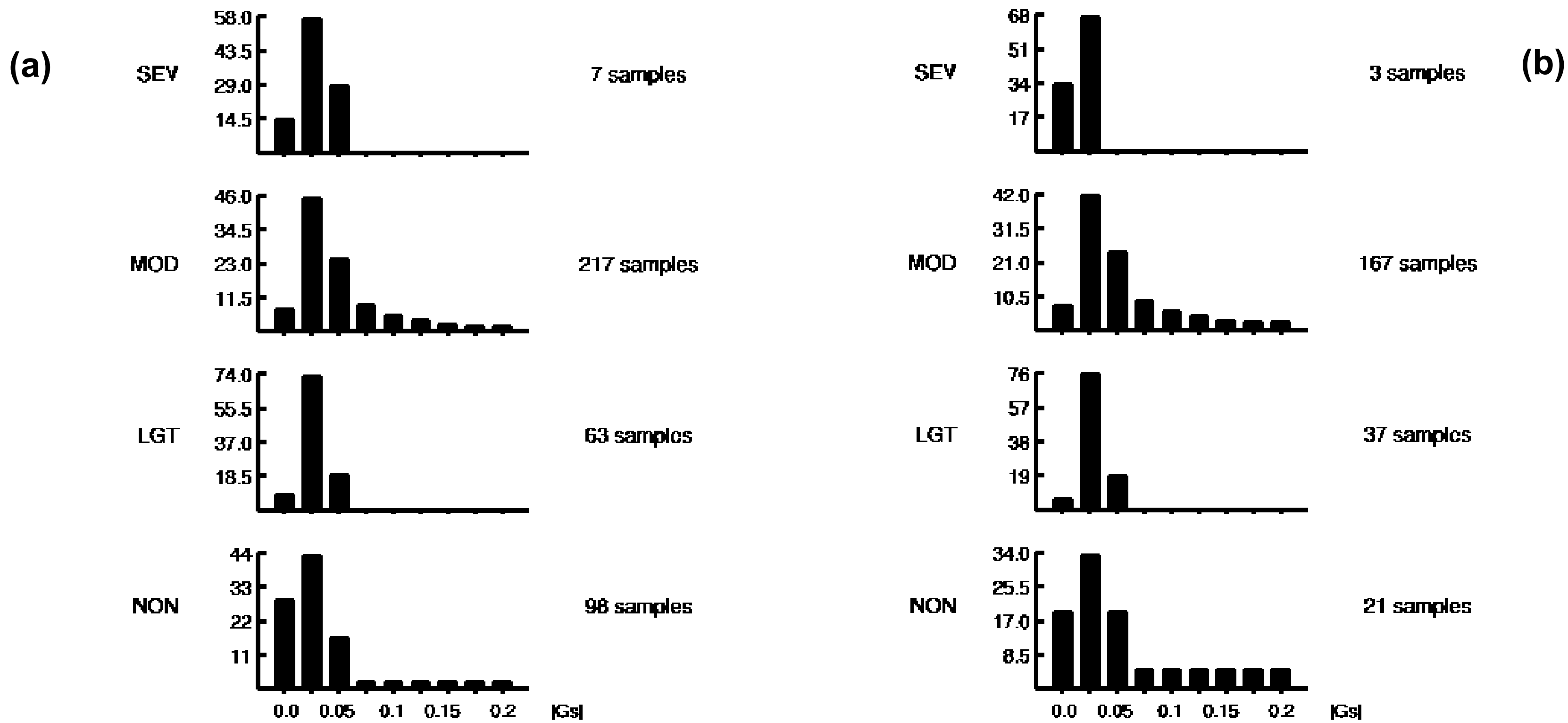

Figure 1. Frequencies (y-axis) of AVAR values (x-axis) associated with different levels of turbulence severity reported by near-by PIREPs. Comparison is based on PIREPs and AVAR observations within 10 min and 70 km of each other, for (a) all aircraft weights; and (b) heavy aircraft.

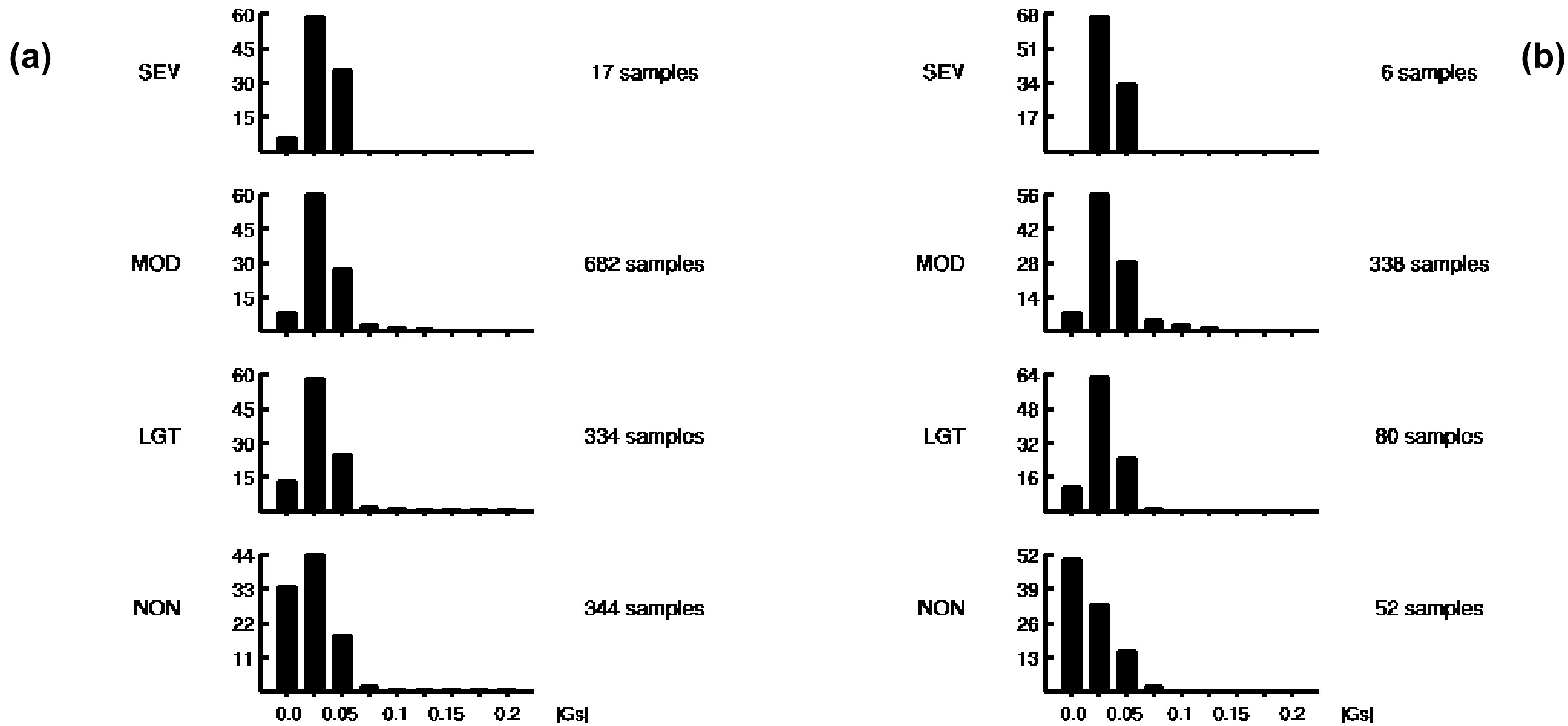

Figure 2. As in Fig. 1, for PIREPs and AVARs within 2 hr and 40 km of each other.

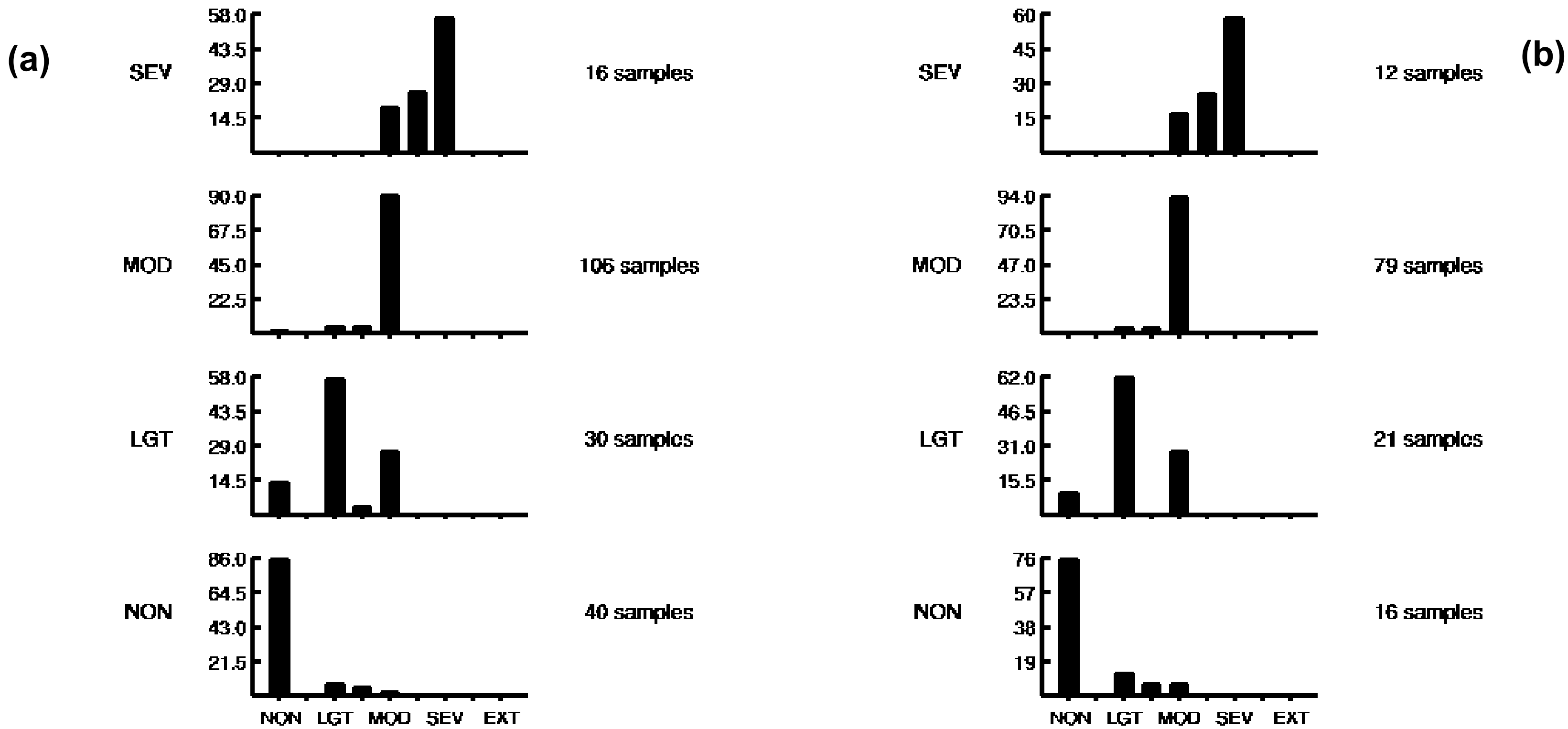

Figure 3. As in Fig. 1 for co-located PIREPs, located within 10 min and 70 km of each other.

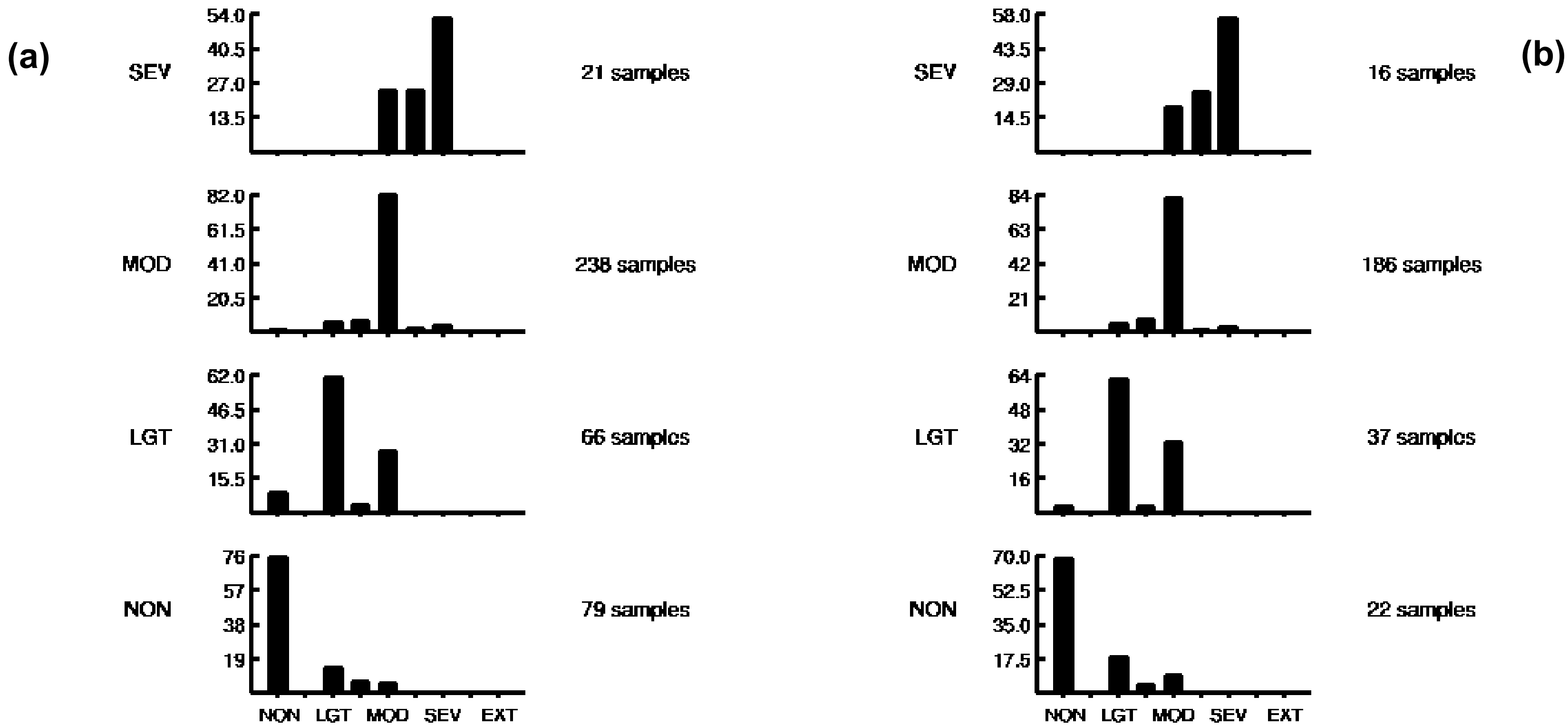

Figure 4. As in Fig. 1, for co-located PIREPs, located within 40 km and 2 hr of each other.

The good news from this analysis is the fact that the PIREPs seem to report turbulence severity quite coherently. As shown in Figs. 3-4, it is unusual for one PIREP to indicate moderate-or-greater severity while another PIREP in the vicinity reports no turbulence. This result suggests that reporting errors and subjectivity in PIREPs may be less of a problem than has previously been assumed.

3. Statistical characteristics of EDR observations

EDR values that have been collected over the last nine months are considered in this section, with emphasis on reports from January and August 2000. Data from January are of interest because it is the center of the winter months, when upper-level turbulence is of greatest concern. The August data are of interest because August is the most recent month with a complete set of observations. A total of 59,948 reports were included in the dataset for January, and 190,496 were included for August. This difference in total counts appears to be associated with a change in the data rate; however, we are continuing to investigate this question.

The EDR values are in units of turbulent kinetic energy. The observations include peak (MAX-EDR) and median (AVE-EDR) values for every minute, collected over a 16-min period before being downlinked from the aircraft. Both the MAX-EDR and AVE-EDR values are binned into categories before they are downlinked, starting with a minimum value of 0.05, with increments of 0.10 up to a maximum of 0.75. Table 1 shows the counts and percentages of reports in each catgory for each month. As shown in Table 1, almost all of the MAX-EDR and AVE-EDR reports are in the 0.05 category, which represents no (or minimal) turbulence. Thus, frequencies in the other categories are quite small. A larger percentage of observations of MAX-EDR than AVE-EDR observations are in the larger categories. In fact, there were no AVE-EDR observations in the 0.75 category. Table 1 also suggests that the relative frequencies of reports are fairly similar between the January and August datasets.

The frequencies of different values of MAX- EDR for January and August 2000 are also provided in histograms in Figs. 5 and 6, respectively. In these plots, the 0.05 category was ignored since the large number of reports in this category would swamp the diagram. Of course, these small values are very important since they represent the no-turbulence category, which is crucial for verification. Figures 5 and 6 also illustrate the dramatic drop in the frequencies for values of MAX-EDR greater than 0.15. The apparent secondary frequency peak for the 0.75 category occurs because larger values are all binned into this category.

Table 1. Frequencies of different categories of EDR in datasets for January and August 2000.

EDR category	AVE-EDR				MAX-EDR			
	January		August		January		August	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent
0.05	58,363	97.4	185,274	97.3	56,120	93.6	177,488	93.2
0.15	1,272	2.1	4,167	2.2	2,978	5.0	9,606	5.0
0.25	131	0.2	709	0.4	426	0.7	2,040	1.1
0.35	47	0.1	180	0.1	109	0.2	686	0.4
0.45	45	0.1	75	0.0	46	0.1	248	0.1
0.55	49	0.1	55	0.0	36	0.1	143	0.1
0.65	41	0.1	36	0.0	35	0.1	82	0.0
0.75	0	0	0	0	198	0.3	203	0.1

Figs. 7-20 present maps of the MAX-EDR reports for January and August. Each map includes the location of each report, including all flight levels, for a specific EDR category. The maps for the 0.05 category (Figs. 7-8) indicate quite wide coverage by the EDR reports. The distribution of reports not only follows the flight paths, as would be expected, but also consistently covers the central part of the country, excluding the northern-tier states, the Southwest, and the Gulf Coast states. As the intensity increases to 0.25 (Figs. 11 and 12), the MAX-EDR reports become more concentrated around major airports; this feature is particularly noticeable in the August map, where there are more reports in total. The relationship between report locations and the locations of major airports seems to lessen as the threshold value increases. In some cases (e.g., Fig. 21), most of the observations seem to fall along a single flight track. It will be desirable to investigate cases like this in greater depth to determine if they are associated with a single flight and to determine the cause of the large values.

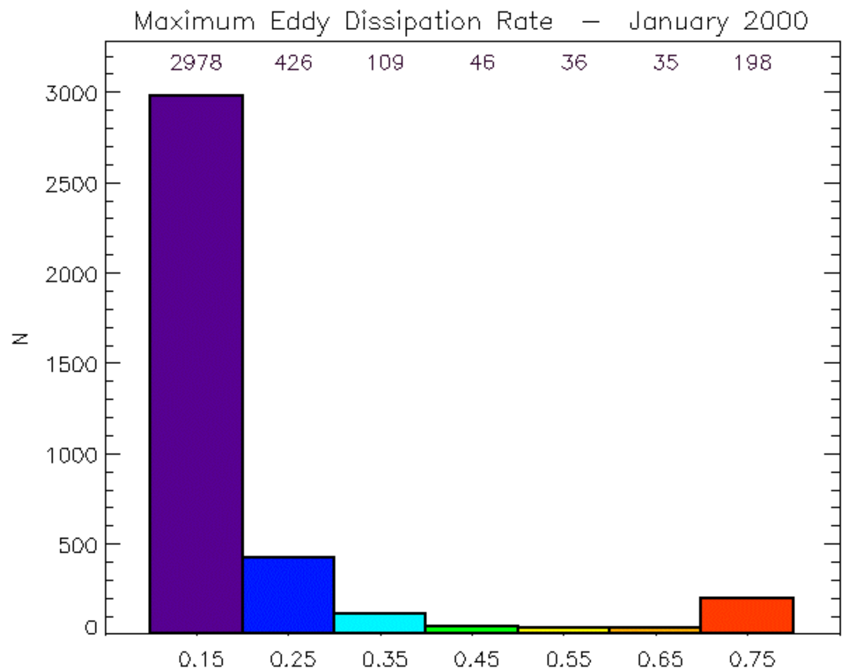

Figure 5. Histogram of MAX-EDR values for January 2000. Each bar represents an intensity category of MAX-EDR values. Frequency is plotted on y-axis with EDR categories along the x-axis.

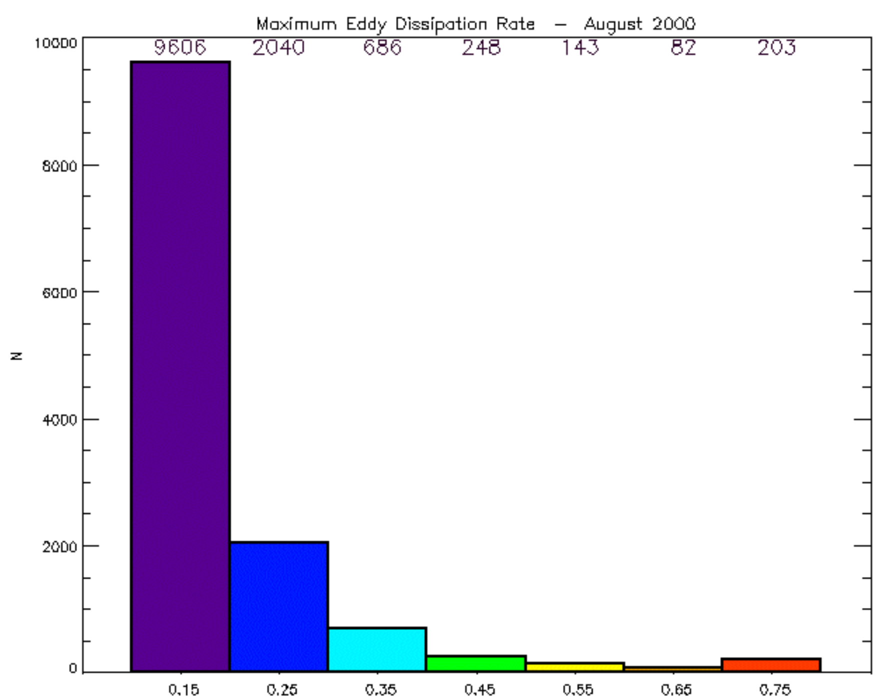

Figure 6. As in Fig. 6, for August 2000.

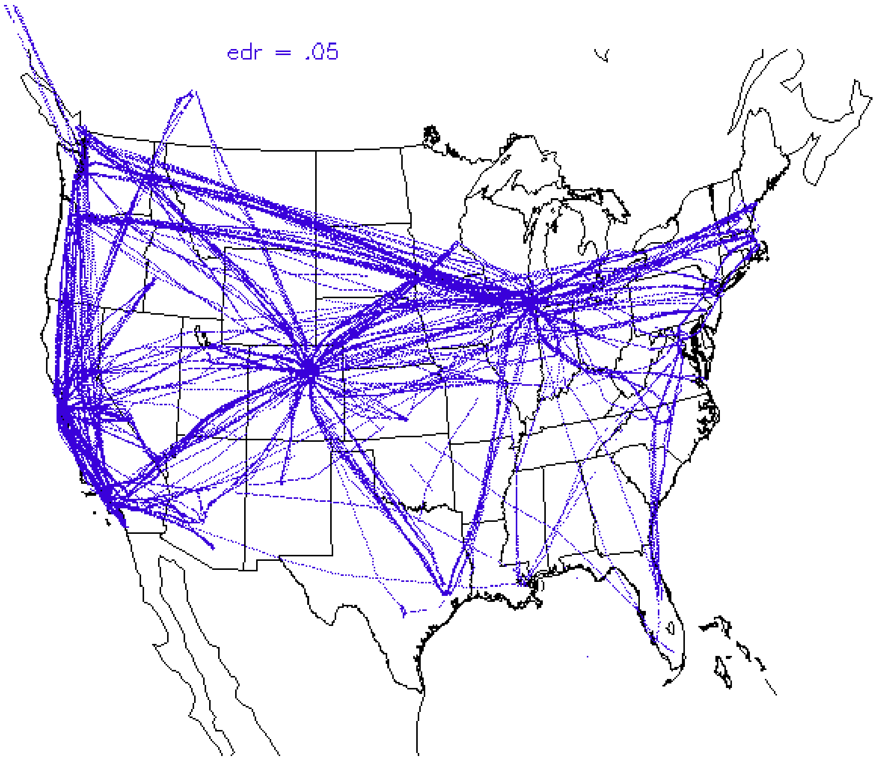

Figure 7. **Map of MAX-EDR values for January 2000. Dots indicate the locations of MAX-EDR reports with a value of 0.05.**

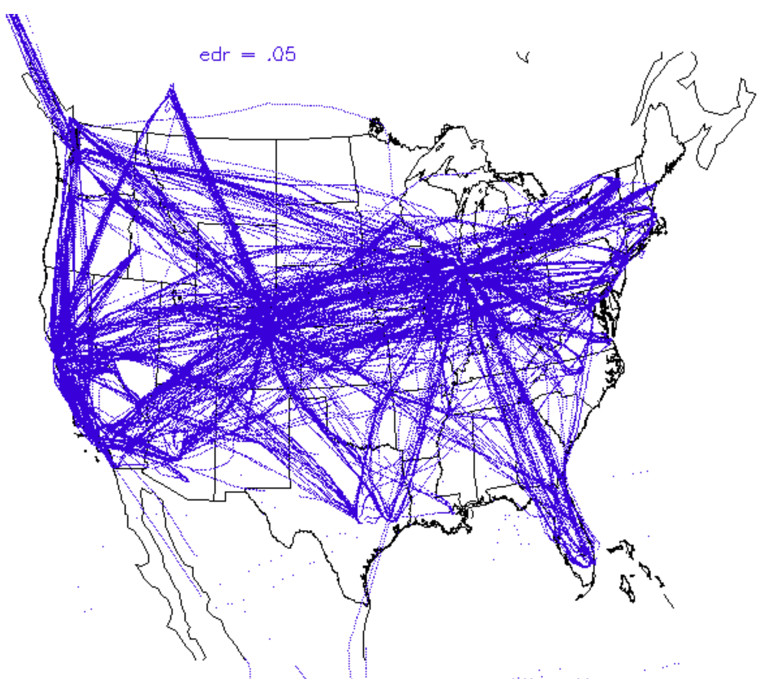

Figure 8. As in Fig. 7, for August 2000.

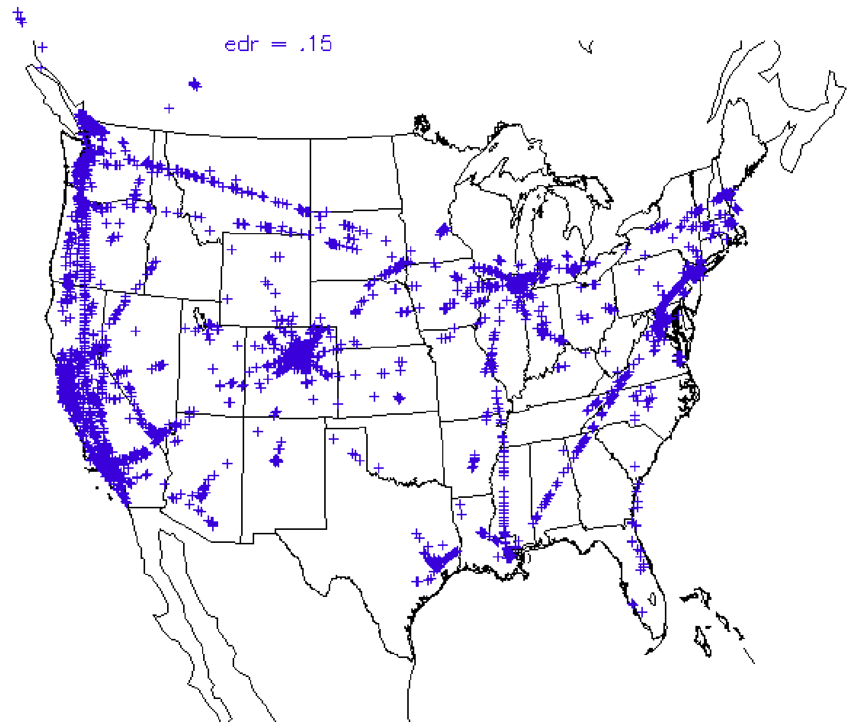

Figure 9. Map of MAX-EDR values for January 2000. Plus symbols indicate the locations of MAX-EDR reports with a value of 0.15.

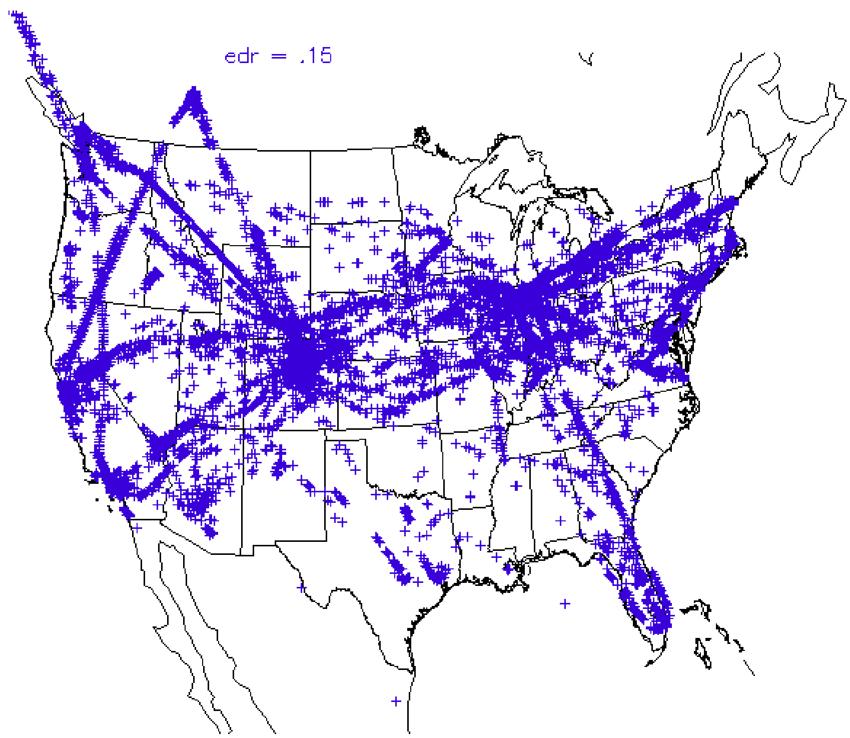

Figure 10. As in Fig. 9, for August 2000.

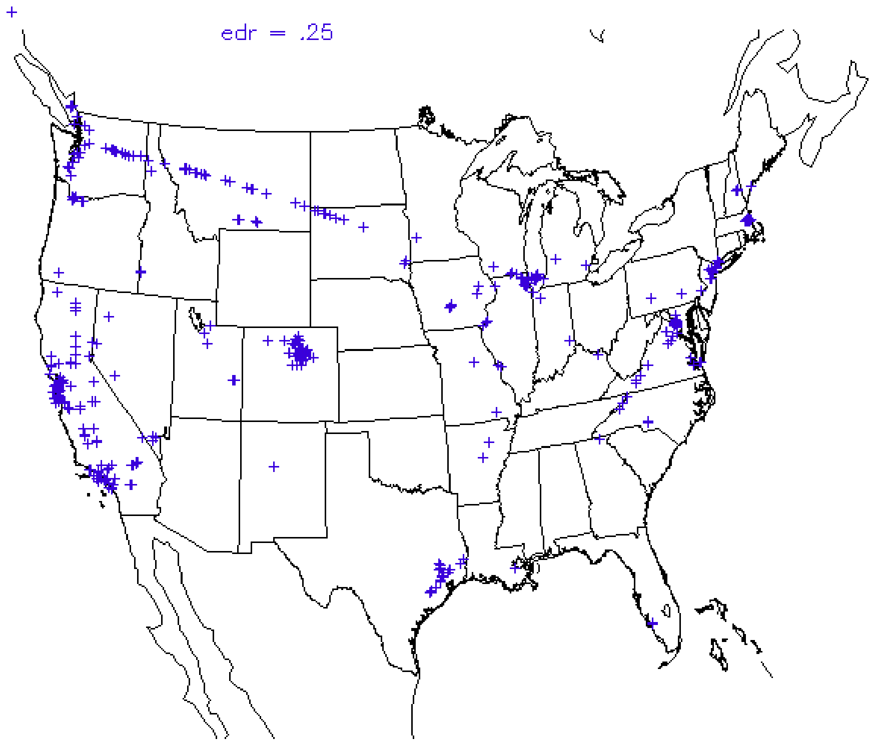

Figure 11. Map of MAX-EDR values for January 2000. Plus symbols indicate the locations of MAX-EDR reports with a value of 0.25.

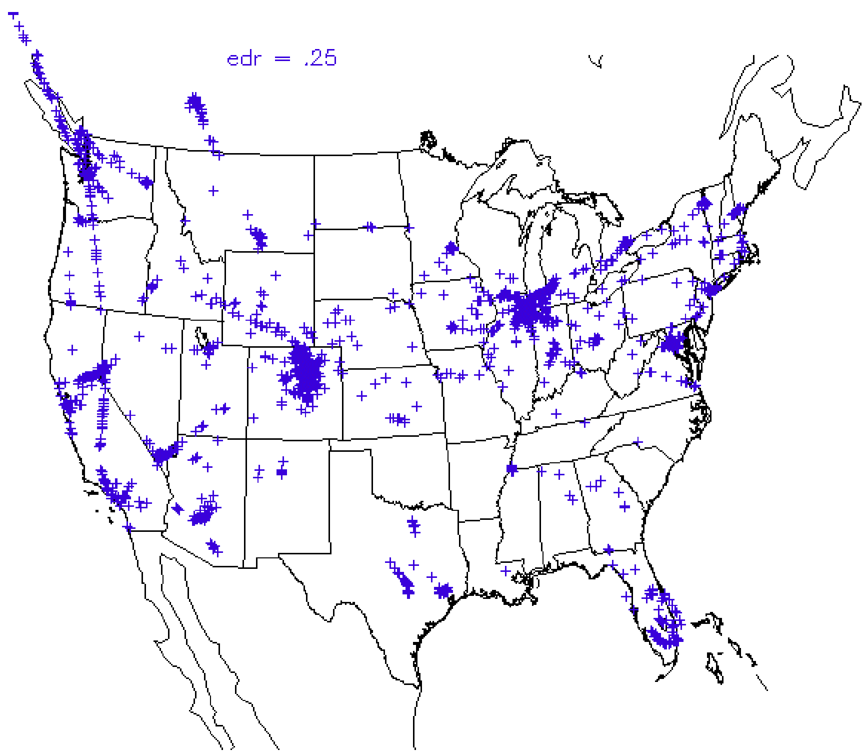

Figure 12. As in Fig. 11, for August 2000.

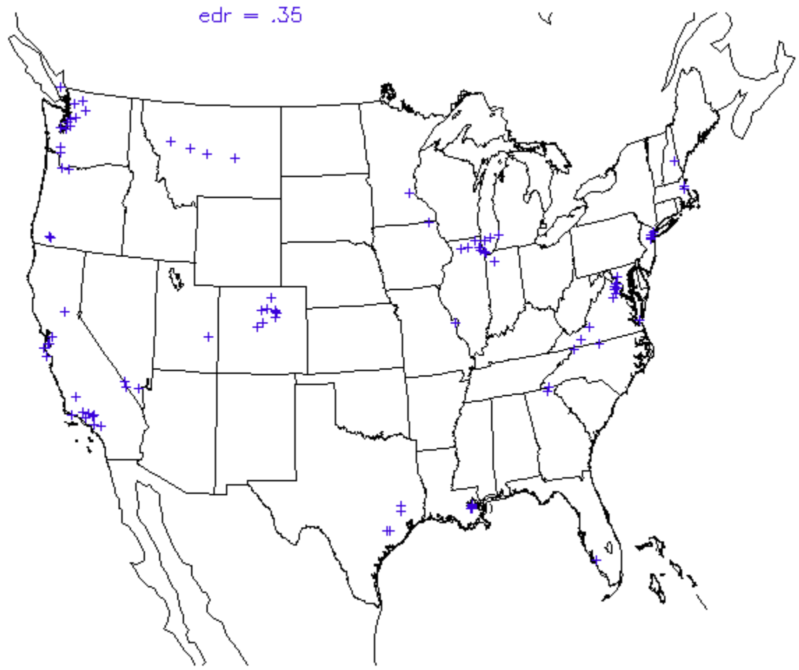

Figure 13. Map of MAX-EDR values for January 2000. Plus symbols indicate the locations of MAX-EDR reports with a value of 0.35.

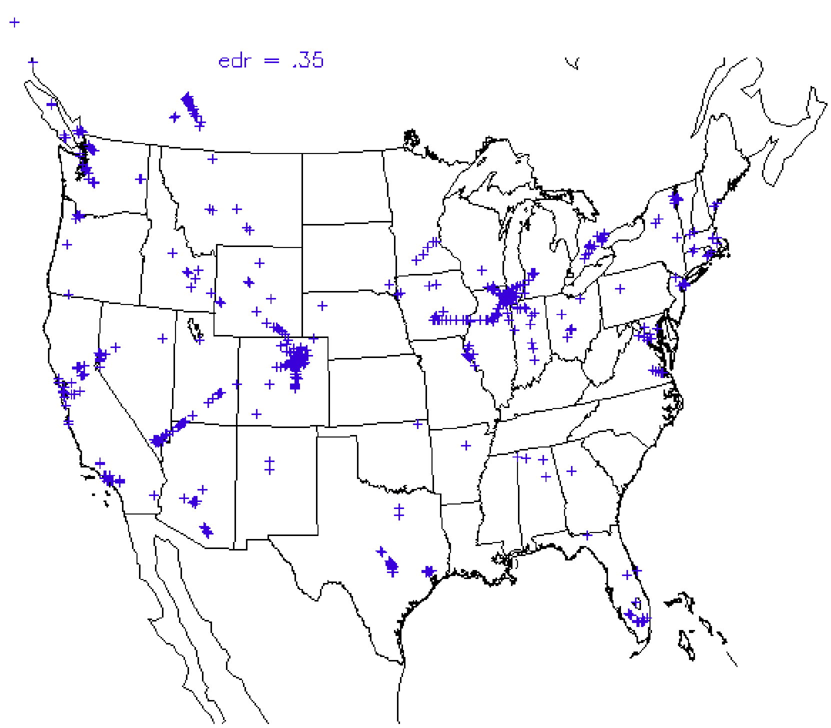

Figure 14. As in Fig. 13, for August 2000

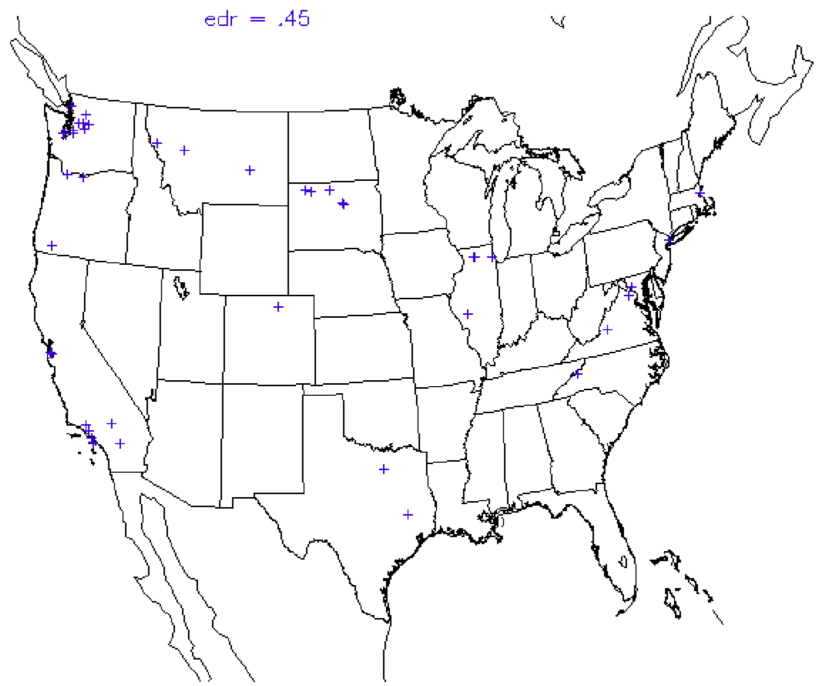

Figure 15. Map of MAX-EDR values for January 2000. Plus symbols indicate the locations of MAX-EDR reports with a value of 0.45.

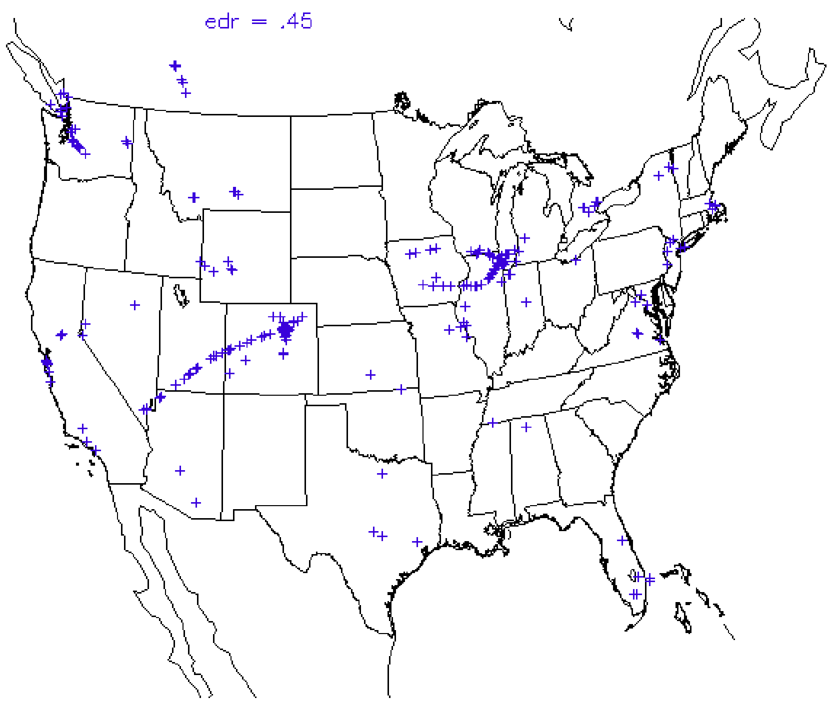

Figure 16. As in Fig. 15, for August 2000.

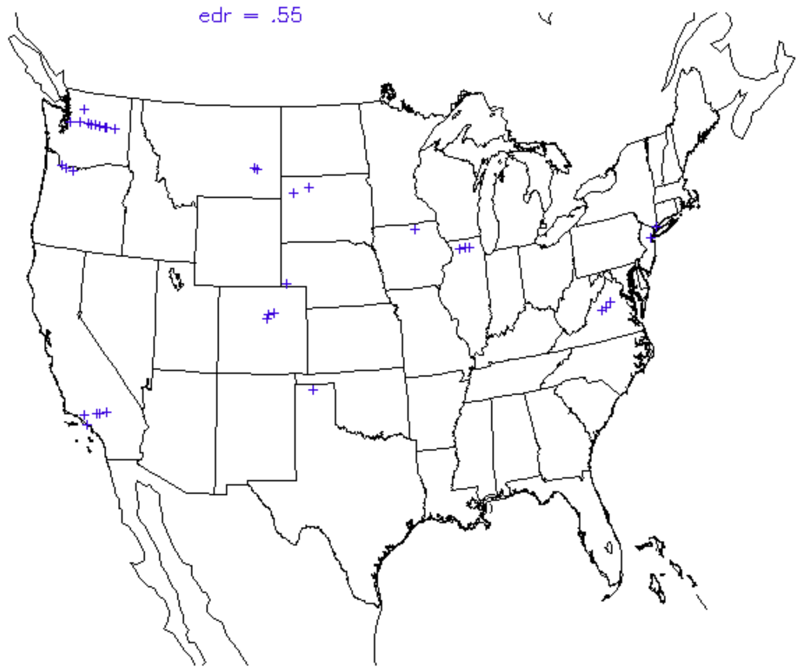

Figure 17. Map of MAX-EDR values for January 2000. Plus symbols indicate the locations of MAX-EDR reports with a value of 0.55.

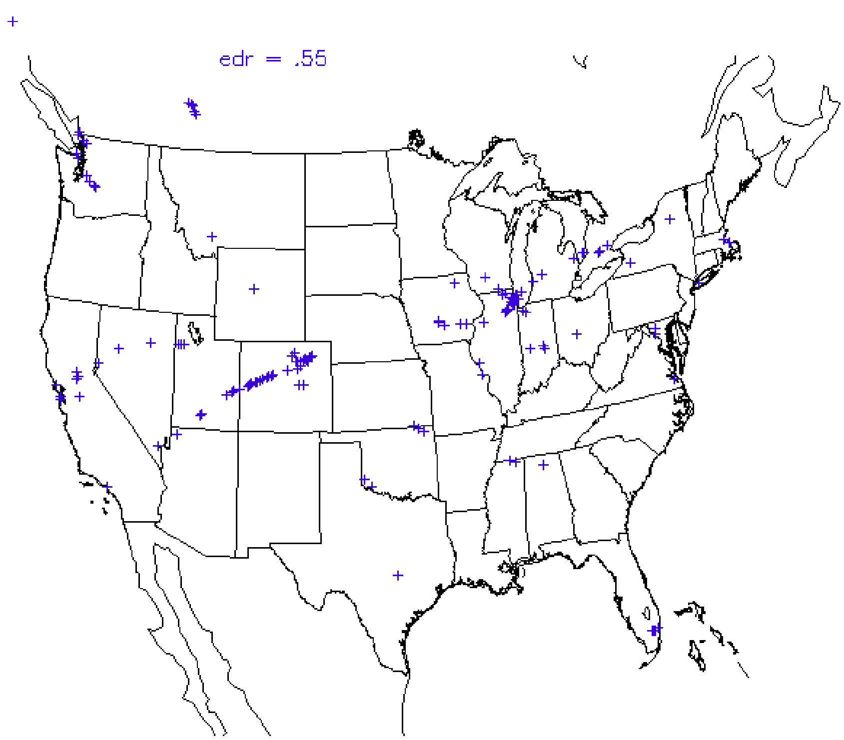

Figure 18. As in Fig. 17, for August 2000.

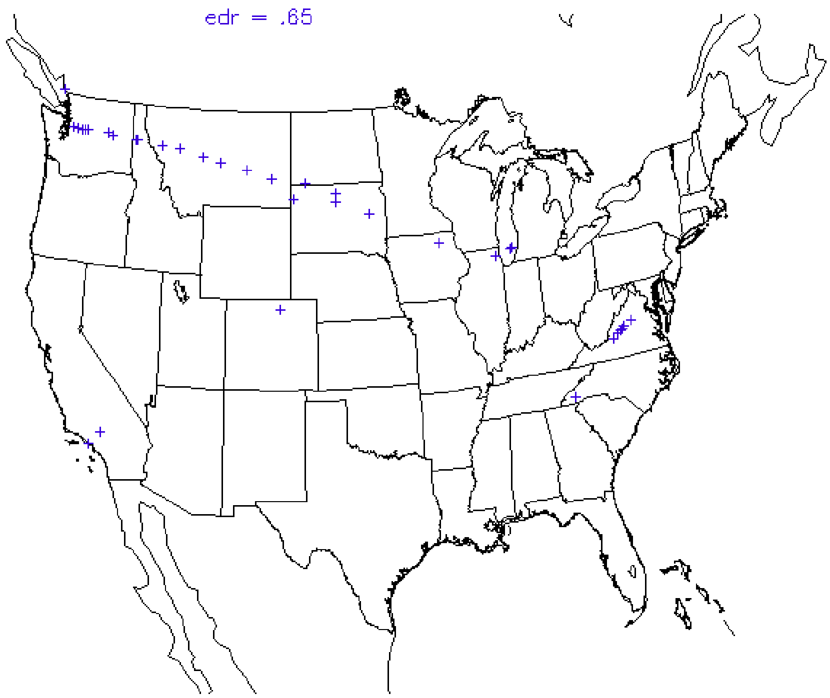

Figure 19. Map of MAX-EDR values for January 2000. Plus symbols indicate the locations of MAX-EDR reports with a value of 0.65.

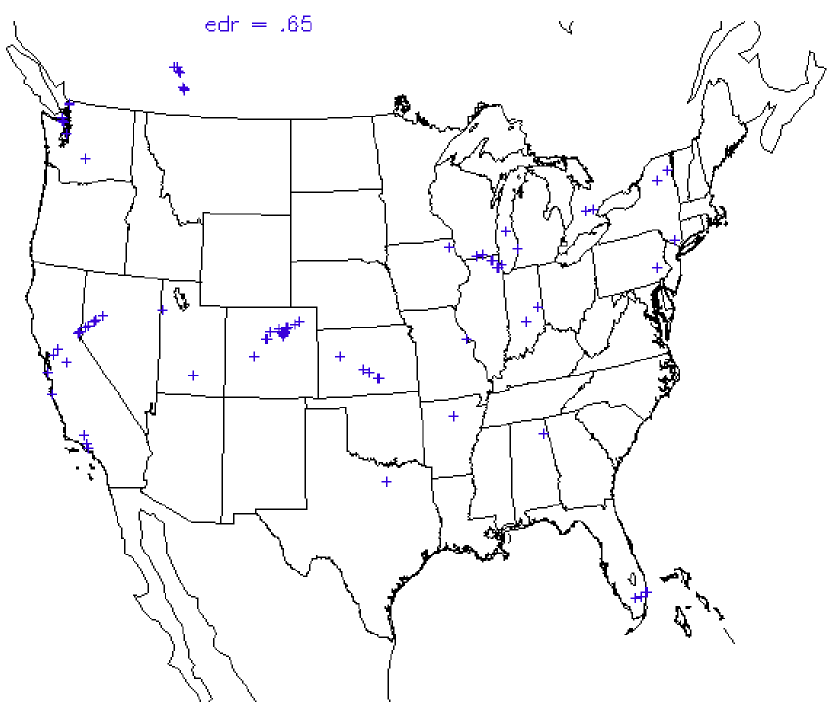

Figure 20. As in Fig.19, for August 2000.

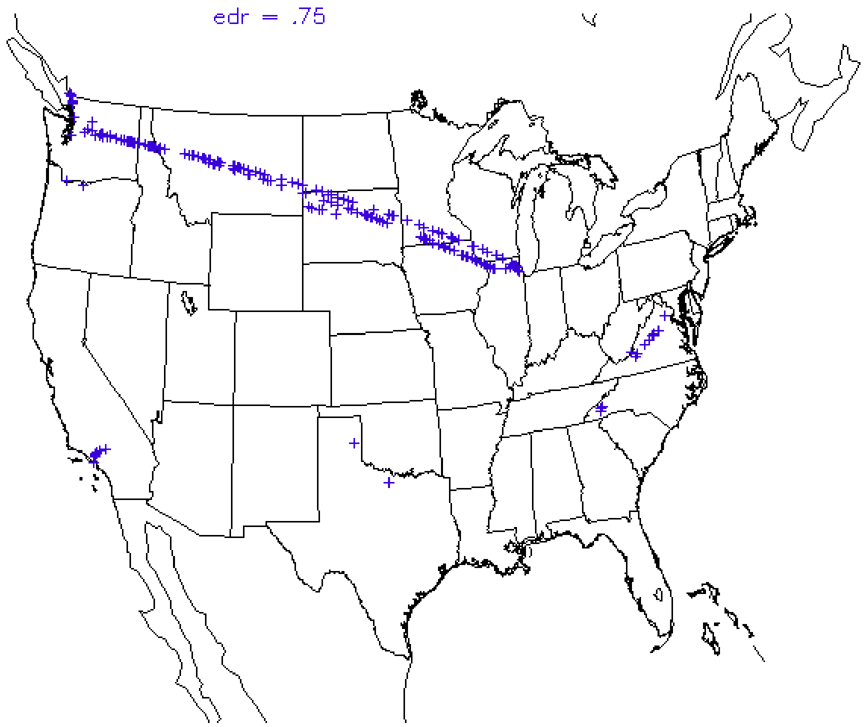

Figure 21. Map of MAX-EDR values for January 2000. Plus symbols indicate the locations of MAX-EDR reports with a value of 0.75.

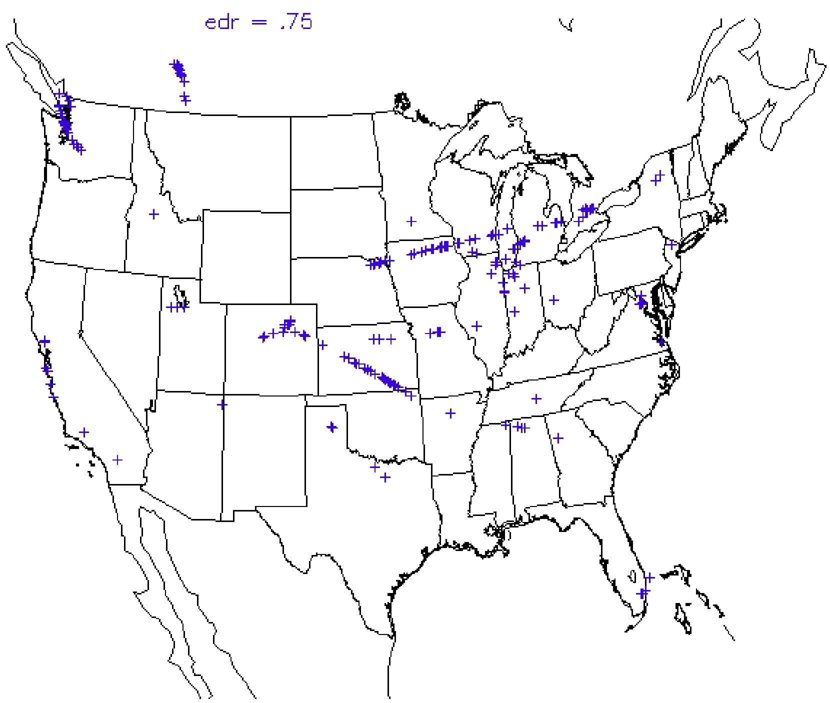

Figure 22. As in Fig. 21, for August 2000.

4. Verification methods based on EDR observations

The nature and distribution of the EDR reports creates the opportunity for new approaches for verification of turbulence forecasts. In particular, the reports include objective information about turbulence severity, which can be compared to the forecast values provided by algorithms and forecast systems. In addition, the basic methods that have been used previously (i.e., based on PIREPs) can also be applied to the EDR data.

As described earlier, the *in situ* EDR data provide two different measures of turbulence – MAX-EDR and AVE-EDR. In the process of developing methods for using the EDR reports for verification, it will be important to investigate the relationships between these two measures. It is possible that in the future some combination of the two types of observations will be selected as an appropriate measure of turbulence for verification of the turbulence forecasts. However, at least until it is clear what each measure represents, both measures should be used for forecast verification.

The basic verification methods that should be considered for use with the EDR data include (i) treatment of the reports as Yes/No values, by applying thresholds to the observations; and (ii) comparison of all categories of reports to categorized forecast values. These two approaches, which are closely related, are discussed in more detail in the following paragraphs.

The verification approach that we currently use to evaluate turbulence forecasts, based on using PIREPs as the verification data, can essentially be applied directly using the EDR data. In particular, PIREPs are currently categorized by severity, much as the EDR values are categorized. The EDR categories that might be used include 0.15 and greater, and 0.25 and greater, with the 0.05 category used to represent no-turbulence conditions. Unfortunately, because there are relatively few EDR values larger than 0.25, it may be difficult to directly make use of the more extreme categories. However, as the *in situ* system becomes available on a larger number of aircraft, these larger categories may be observed more frequently. In that case, it would be possible to directly verify the forecasts for those categories.

A major issue that is of concern when PIREPs are used for verification of turbulence forecasts involves the non-systematic nature of the reports. This characteristic of the reports limits the verification statistics that can be computed and, in particular, makes it inappropriate to compute the false alarm ratio and various other verification measures. We expect that the EDR observations will be more systematic than PIREPs. However, they still will only represent a specific portion of the forecast grid

– where the larger aircraft carrying the EDR software are flying. Thus, it is not clear at this point that verification approaches based on the EDR data will be freed from this constraint, at least in the short term.

The EDR observations also can be compared directly to the values computed for the forecasts. Ideally, if the actual EDR values were available (i.e., rather than the categories), it would be valuable to compute the correlation coefficient measuring the strength of the relationship between the forecasts and observations. However, computing this statistic is problematic when the data are grouped into so few categories. Thus, the most appropriate approach will be to create multi-category contingency tables, in which the forecast values are grouped into the same number of categories as the observations. As in the 2x2 contingency table, the entries in the multi-category tables will represent the number of times that a particular observed value was associated with a particular forecast category. The distributions of forecast values associated with different categories of EDR can be displayed, to measure the ability of the forecasts to discriminate between the categories. Similarly, the distributions of the EDR values can be displayed as a function of the forecast categories. These conditional-distribution diagrams will provide information concerning systematic errors and calibration of the forecasts.

Finally, it will be important to continue to utilize PIREPs for verification, along with the EDR data, for the foreseeable future. There are several reasons for this. First, all evaluations of turbulence forecasts up to the present have been based on these data, so PIREP-based analyses are needed for continuity and comparability. Second, PIREPs measure turbulence at different flight levels and by different aircraft than those used to obtain the EDR measurements. Thus, they provide a broader depiction of observed turbulence. However, the PIREPs should be considered separately from the EDR observations, rather than in combination with them, since they do represent a very different type of observation.

5. Summary and conclusions

This report has considered several aspects of the use of automated observations for the verification of turbulence forecasts. In particular, the AVAR observations have been compared to co-located PIREPs; basic statistical characteristics of the EDR data have been examined; and some implications of the characteristics of the EDR data for verification approaches have been described.

Results of the analyses of the AVAR data indicate that these data should not be used for verification. In particular, the AVAR values do not seem to be related at all to the severity of the observed turbulence. This result is true even when the values are limited to the range in which no turbulence is expected. It is likely that this result is due to the short time window in which the vertical acceleration is observed. In contrast, co-located PIREPs reported very similar levels of turbulence severity, which is good news for current verification studies that are based on turbulence PIREPs.

The EDR observations are very frequent and provide wide (though not complete) coverage over the continental United States. Nearly all of the EDR values indicate minimal or no turbulence; higher categories of EDR (0.25 and greater) are relatively rare. Methods for using the EDR for verification include the use of current methods (i.e., based on the evaluation of Yes/No forecasts), as well as enhanced methods that directly consider all of the EDR categories.

Acknowledgments

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 and position of the U.S. Government.