P6.5 On the Use of Eddy Dissipation Rate Observations in Verification

Brian P. Pettegrew[1,2], Andrew F. Loughe[1,2], Jennifer L. Mahoney[1],
Judy K. Henderson[1] and Joan E. Hart[1,2]

[1]NOAA/Earth Systems Research Laboratory
Boulder, Colorado
[2]Cooperative Institute for Research in Environmental Sciences
Boulder, Colorado

1. INTRODUCTION

As part of the ongoing efforts within the Forecast Verification Section (FVS) in the Global Systems Division (GSD) at the Earth Systems Research Laboratory (ESRL), the use of in-situ Eddy Dissipation Rate (EDR) observations are being tested as an independent observation source for verification of turbulence forecasts. Until recently, the primary source of atmospheric turbulence observations available for the verification of turbulence forecasts has been voice Pilot Weather Reports (PIREPs). Verification of turbulence is difficult due to the limited amount of moderate-or-greater (MOG) reports. Further reduction in severe-or-greater observations make verifications of a turbulence forecast all the more challenging.

PIREPs, by nature, are subjective due to the human uncertainty presented in reporting (Kane et al. 1998). While valuable to air traffic flying in areas of reported weather hazards, PIREPs were not intended to provide research quality observations. Further, it has been shown that PIREPs are biased towards positive reports and are not systematic (Brown et al. 1997). PIREPs, in general are also sporadic both temporally and spatially. While the temporal rate of PIREP dissemination is unknown, Schwartz (1996) found that about 585 turbulence PIREPs are reported daily.

EDR reports are relatively new to the field of turbulence verification and are constantly being analyzed as a replacement and/or a supplement observation to the PIREP. EDR observations of turbulence are measurements of vertical motions and aircraft response to those motions taken in-situ (Cornman et al. 1995, 2004). According to Cornman et al. (2004), EDR reporting is designed to provide "routine and quantitative measurements of atmospheric turbulence – including null reports." Measurements of EDR are taken and reported on a minutely basis. This frequency of reporting creates a high density of NULL reports creating an over-sampling issue, and subsequent bias towards non-events, and reducing resolution of turbulence events for verification.

The goal of this study is to test the sensitivity of verification scores against differing EDR observation sampling rates. By doing so, it can be shown that resolution of EDR observations can be gained from raw data and the bias, from over-sampling, is reduced while maintaining skill in turbulence forecasts.

This study was performed using minutely EDR observations from June 2007 through May 2008 to verify the second version of the Graphical Turbulence Guidance (GTG2) algorithm as defined by Sharman (2004). This paper will take on the following form: Section 2 will discuss the data and methodology of the study, Section 3 will provide some results and a brief discussion and Section 4 will provide a quick conclusion.

2. DATA AND METHODOLOGY

Only two primary sources of data were used for this study: 1) minutely EDR observations and 2) deterministic fields of turbulence forecasts as provided by the GTG2 algorithm. As described in section 1, the EDR observations are measurements of vertical motions in the atmosphere and an aircraft's response to this motion. According to Cornman et al. (1995, 2004), there are two principal methods through which EDR is

Corresponding author: Brian Pettegrew, CIRES/NOAA/ESRL/GSD/AB, David Skaggs Research Center, Boulder, CO 80305. E-mail: brian.p.pettegrew@noaa.gov

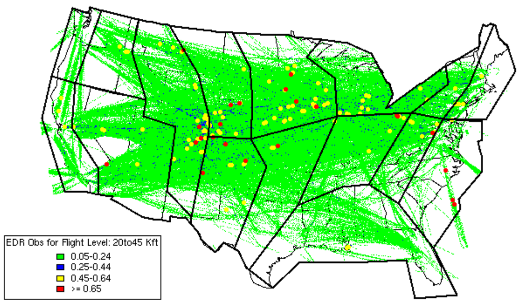

Figure 1: Spatial Distribution of EDR observations from June 2007 through May 2008 color coded by intensity. Legend in lower left corner shows intensities. Black lines outline climatologically derived regions as determined by the National Centers for Atmospheric Research (NCAR).

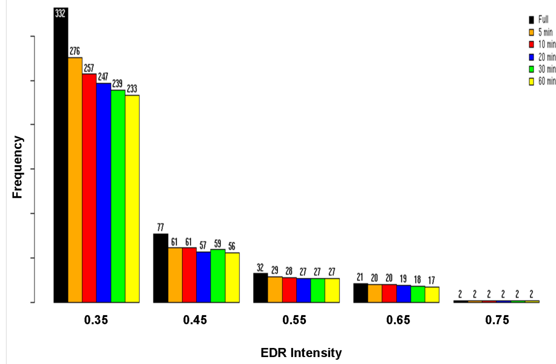

Figure 2: Distribution of Moderate-or-greater (MOG) EDR observations color coded by sub-sampling rate. Frequency of occurrence is normalized frequency. N samples observed are at the top of each bar graph with the legend in the top right.

measured. One is through the measurement of vertical accelerations and a mathematical model of the aircraft response. The second method calculates the vertical wind component in the atmosphere. As of 2006, there were about 200 United Airlines B737 and B757 aircraft equipped for EDR observations (Takacs et al. 2006). Figure 1 shows the spatial distribution of EDR observations from June 2007 through May 2008, color coded by intensity from 20,000 ft to 45,000 ft.

The GTG2 algorithm is an upgrade to the original GTG algorithm. The GTG2 uses a suite of algorithms in order to arrive at a deterministic turbulence forecast for the Continental United States (CONUS) using atmospheric variables derived from the Rapid Update Cycle (RUC) operational forecast model. One of the key improvements was the addition of the 10,000 – 20,000 ft. levels. Further improvements were made with turbulence intensity thresholds.

Formal verification evaluations of the GTG2 product (Takacs et al. 2004 and Kay et al. 2006) showed skill, particularly in discriminating between 'Yes' and 'No' turbulence observations. However, at the time, the primary in-flight observations available to evaluators were PIREPs. In an effort to integrate EDR observations into the evaluation of turbulence forecasts, Takacs et al. (2006) developed a procedure in which EDR observations within a given time and space domain of a PIREP were used, effectively matching PIREPs to EDR observations in time and space.

The study discussed in this paper discards the use of PIREPs and uses the raw EDR observations solely to verify GTG2 performance. Only the 1800 UTC issuance time and 6-hr lead time (0000 UTC valid time) were used. The verification was performed over a year from June 2007 through May 2008. For this time period, EDR observations for +/- 1 hr around 0000 UTC are gathered. A grid matching technique similar to that employed in previous evaluations was utilized where EDR observations are matched to the 12 nearest grid points of the GTG2 (4 above, 4 below and 4 at altitude). To retain the worst possible condition, the maximum turbulence as determined from the GTG2 at the location of the EDR observation is recorded. EDR observations are sub-sampled at time rates of 0, 5, 10, 20, 30, and 60 minutes. Sub-sampling in this case refers to a filtering process in which several NULL reports are removed and MOG reports are retained to gain resolution from the EDR observations dataset. A filtering technique was performed in which the maximum peak EDR observation per flight per given sampling rate (i.e. 5, 10, 20 min) was retained. Given the frequency of EDR observations, it is expected that some MOG observations might be lost due to multiple reports in a given sampling window. For instance, if two moderate reports of turbulence are observed via EDR in the same flight within a 10-min time frame, only the most recent of those moderate reports are retained. Figure 2 shows the distribution of MOG observations from EDR reports from June 2007 through May 2008 for both the full data set and the sub-sampled data sets.

Table 1: Verification statistics and formulas produced from dichotomous statistical analysis.

Statistics	**Formula**
PODy (Probability of Detection, Yes)	YY/(YY+NY)
PODn (Probability of Detection, No)	NN/(NN+YN)
Bias	(YY+YN)/(YY+NY)
False Alarm Ratio	YN/(YN+YY)
False Alarm Rate	1-PODn
True Skill Statistic (TSS)	PODy + PODn -1

From the individual matches in each sub-sampled data set, dichotomous skill scores are calculated from a 2x2 contingency table based on the MOG threshold as pre-determined by the GTG2 algorithm. In this study, a MOG intensity value of 0.35 and greater for EDR is used while the MOG intensity scale for the GTG2 deterministic forecast is 0.475. Table 1 shows scores derived from the contingency table. These statistics were stratified seasonally and monthly for the entire time period, but these results are not shown here. Regional results were also ignored due to the lack of spatial coverage of EDR observations through some regions of the CONUS. Further, a short case study demonstrates the sampling of EDR observations for a given origin and destination to show the spatial impact of sub-sampling these data.

3. RESULTS AND DISCUSSION

Initial results suggest that a sub-sampling rate of 5 min is sufficient to reduce noise in the NULL observations of turbulence. A short case study was performed to look at the sampling rate of EDR observations based on a flight path across the CONUS when sampled at the various rates. Figure 3 shows EDR observations for a flight path along a range of latitudes in the CONUS for a flight originating in New York, NY (JFK) and landing in San Francisco, CA (SFO). All observations are from the same flight as confirmed by the tail number. The observations are color coded by intensity along the flight path with blue indicating NULL observations and red indicating at least MOG turbulence along the flight path. In this case, there was only one observation of MOG turbulence along the flight path. The yellow observations represent MOG turbulence as indicated by PIREPs along the same range of latitudes.

Progressively, through the different samples, NULL observations are removed from the flight path while the MOG turbulence observation in red are preserved. The mean distance between EDR observations are above each plot. When the EDR data are not filtered, an observation is reported approximately every 15 km. When filtered down to the maximum EDR per flight per 5-min, the mean distance between observation increases to 53 km for this flight. At 10-min, the distance is 67 km. The mean distance is greater than 100 km by 30-min. Vinnichenko et al. (1968) states that a pocket, or “pancake”, of clear-air turbulence (CAT) has a horizontal length of < 60 km and a horizontal extent of a few meters at least 85% of the time. A sub-sampling of EDR observations every 5 min per flight provides an observation every 53 km which fits into the Vinnichenko dimensions allowing for sufficient sampling to physically capture possible CAT occurrences while removing an adequate number of NULL occurrences to bring some resolution to the data set. Sampling at 10-min intervals is found to be sufficient as well.

Dichotomous statistics further support the idea that filtering EDR observations are beneficial and do not degrade performance when evaluating turbulence forecasts. The dichotomous statistics were calculated based on an increasing threshold from the GTG2 product. The GTG2 deterministic threshold for MOG turbulence from 0 to 1 is 0.475. Figure 4 shows the GTG2 Bias score from all sampling rates as determined from Table 1 for the time June 2007 through May 2008. Bias scores at low thresholds are exponentially high, as expected by the over-sampling of NULL turbulence, but at the MOG threshold of 0.475, while the Bias using the full data set is reduced significantly, the Bias score using the 5-min sub-sampling is much closer to 1, an indication that resolution of MOG observations can be gained and low intensity bias’ can be removed in the data set. Further, the Bias statistics indicate less over forecasting. PODy statistics by threshold (Fig. 5) support the use of EDR observation filtering by indicating insignificant reduction in performance of the

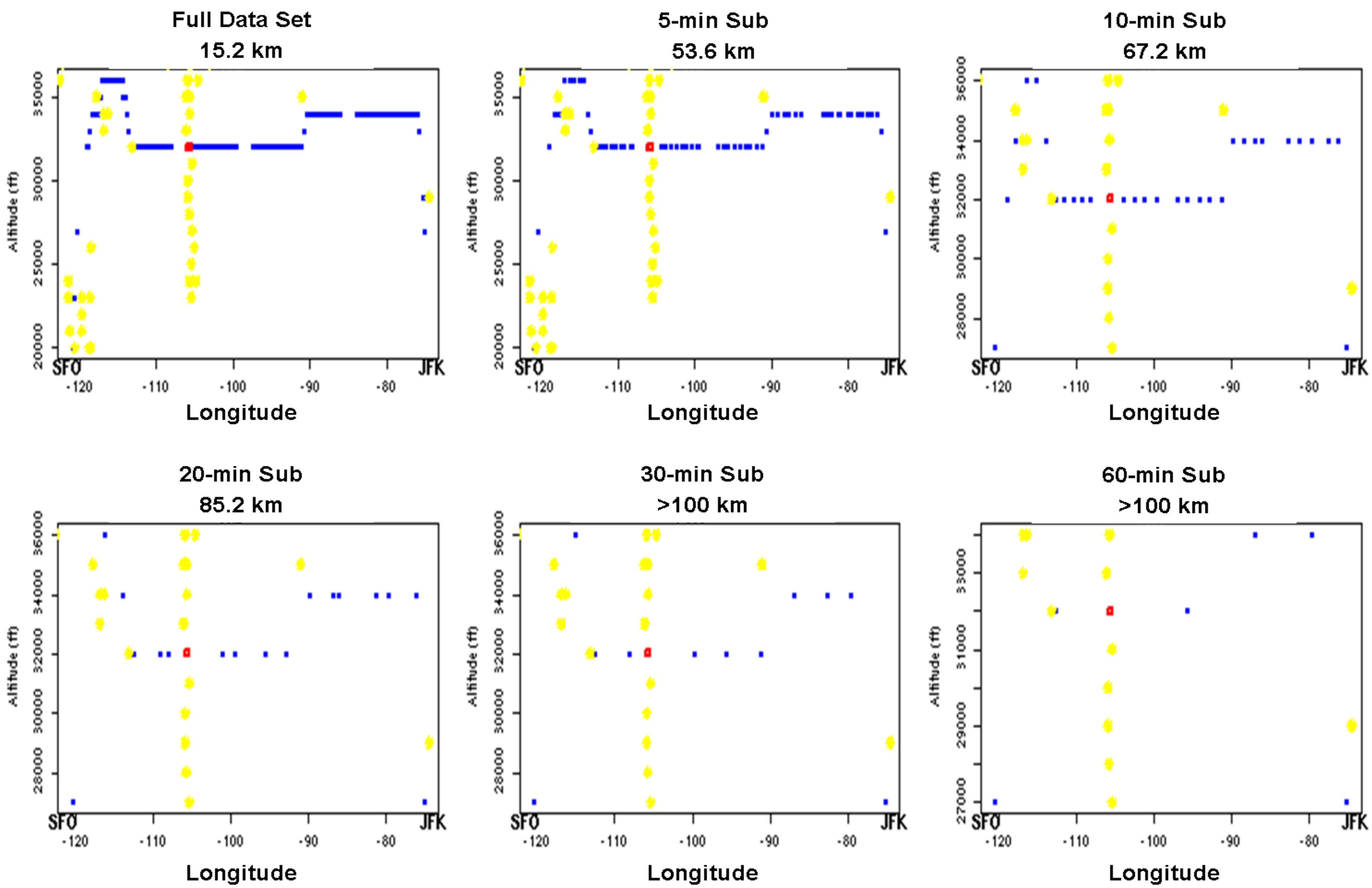

Figure 3: EDR sampling along a flight path from New York, NY (JFK) to San Francisco, CA (SFO) for each sampling rate. PIREP observations are overlaid in for the same latitude.

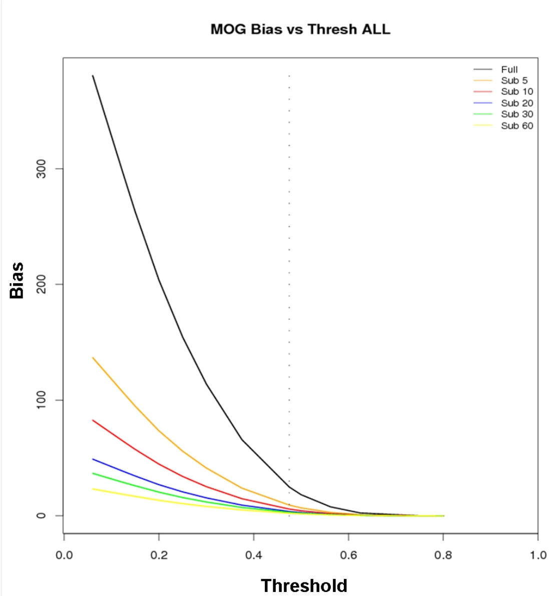

Figure 4: Dichotomous Bias score for GTG2 product using EDR observations. The GTG2 Moderate-or-greater threshold is indicated by the dashed vertical line. The legend is located in the top right.

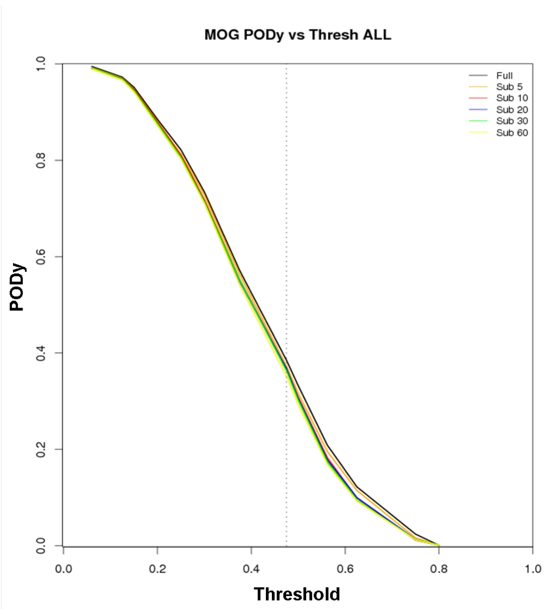

Figure 5: Dichotomous Probability of Yes Detection (PODy) score for GTG2 product using EDR observations. The GTG2 Moderate-or-greater threshold is indicated by the dashed vertical line. The legend is located in the top right.

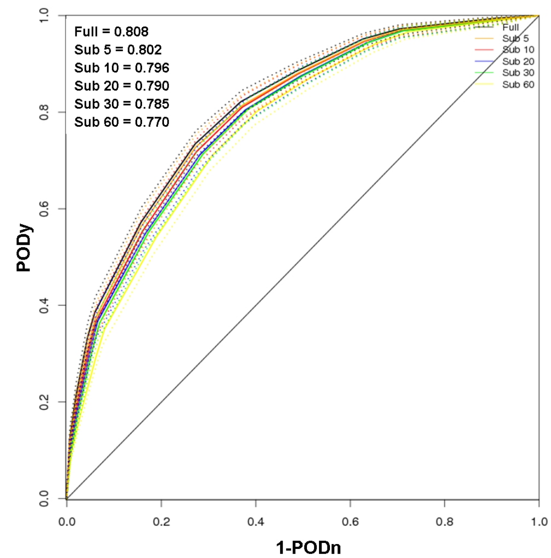

Figure 6: Receiver Operating Characteristic (ROC) plot for GTG2 product using EDR observations. The legend is located in the top right. Area under the curve (AUC) calculations are in the top left corner.

GTG2. As anticipated, PODy reduces greatly as the threshold is increased closer to 1, but by filtering NULL observations and maintaining the integrity of the MOG data set, the notion that skill is reduced is removed. The PODn (not shown) supports this claim with little difference in scores. The PODn statistic relies on the NULL observations for its calculation.

The performance of the GTG2 using the variously sampled EDR observations was also evaluated by looking at Receiver Operating Characteristic (ROC; Mason 1982) plots which measures the ability of a forecast to discriminate between an event and a non-event by plotting the PODy, or the hit rate, against the False Alarm Rate (1-PODn) at the same varying thresholds as used in Figures 4 and 5. Figure 6 shows the ROC plot for the evaluation period of June 2007 through May 2008 for the GTG2 using the variously sampled EDR observations. The slope=1 diagonal represents the threshold for skill/no skill. The area under the curve is representative of the products performance. The area under the curve for each sampling is in the top left corner. Confidence intervals are shown in dashes around the respective performances. The area under the curve is minimally affected between the full data set and the 5-min sampled dataset (0.808 and 0.802 respectively). Further, the 5-min and 10-min sub-sampling ROC curves are within the confidence interval of the full data set ROC curve. These results support the hypothesis that product performance as calculated by dichotomous statistics is not impacted by the filtering of a robust, yet over-sampled, EDR observation data set.

4. SUMMARY

The goal of this study was primarily to show that performance statistics of turbulence forecasts are not impacted by the removal of information from the verifying observation set. In this case, EDR observations, which are sampled minutely, were filtered to preserve the maximum turbulence observation per flight per varying time rate. The time rates selected were 5, 10, 20, 30 and 60 minutes.

A case study on selected flight paths showed that 5-min sub-sampling of the EDR observations was ideal in space as it increased the distance between observations on average from 15 km to 53 km, which may be representative of the physical dimensions of a CAT occurrence. Beyond 20 min, sub-sampling increases the mean sampling distance to greater than 100 km introducing the risk that the observation is missing possible turbulence occurrences, thus making it less effective in verifying turbulence forecasts.

Dichotomous statistics calculated from the verification of GTG2 using EDR showed little change from the application of the observation filter. Performance statistics such as PODy and PODn were impacted very little, especially at the MOG threshold. The greatest change occurred in the Bias.

When using EDR observations in verification of turbulence forecasts, the frequent sampling produces an abundance of NULL observations reducing the resolution of significant turbulence events. It has been shown that filtering these data sets carefully can retain the significant observations while reducing NULL observations and producing resolution in the data set, making it more attractive for verification use. In conclusion, when using an EDR observation data set for verification, the performance as indicated by skill scores are not impacted by the filtering of the data set and is less intensive in terms of verification processing.

Acknowledgements: This research is in response to requirements and funding provided by the Federal Aviation Administration. The views expressed are those of the authors and do not necessarily represent the official policy and position of the U.S. Government.

REFERENCES

Brown , B. G., G. Thompson, R. T. Bruintjes, R. Bullock, and T. Kane, 1997: Intercomparison of In-Flight Icing Algorithms, Part II: Statistical Verification Results. *Wea. Forecasting*, **12**, 890-914.

Cornman, L. B., C. S. Morse, G. Cunning, 1995: Real-time estimation of atmospheric turbulence severity from in-situ aircraft measurements. *J. Aircraft*, **32**, 171-177.

——, L. B., G. Meymaris and M. Limber, 2004: An update on the FAA Aviation weather Research Program's *in situ* turbulence measurement and reporting system. *11th Conf. of Aviation, Range and Aerospace Meteorology,* Hyannis, MA., Amer. Meteor. Soc.

Kay, M. P., J. K. Henderson, S. A. Krieger, J. L. Mahoney, L. D. Holland, and B. G. Brown, 2006: Quality Assessment Report - Graphical Turbulence Guidance (GTG) Version 2.3. Submitted to FAA Aviation Weather Technology Transfer (AWTT) Technical Review Panel.

Kane, T. L., B. G. Brown, and R. T. Bruintjes, 1998: Characteristics of Pilot Reports of Icing, *Preprints, 14th Conference on Probability and Statistics*, Phoenix, AZ, Jan. 11-16, Amer. Meteor. Soc., 90-95.

Mason, I., 1982: A model for assessment of weather forecasts. *Australian Meteorological Magazine*, **30**, 291-303.

Schwartz, B., 1996: The quantitative use of PIREPs in developing aviation weather guidance products. *Wea. Forecasting*, **11**, 372-384.

Sharman, R., J. Wolff, G. Wiener, and C. Tebaldi, 2004: Technical Description Document for the Graphical Turbulence Guidance Product 2 (GTG2). Report, submitted to the Federal Aviation Administration Aviation Weather Research Program (FAA/AWRP).

Takacs, A., L. Holland, M. Chapman, B. G. Brown, J. L. Mahoney, and C. Fischer, 2004: Graphical Turbulence Guidance 2 (GTG2): Quality Assessment Report. Submitted to FAA Aviation Weather Technology Transfer (AWTT) Technical Review Panel.

——, A., L. Holland, R. Hueftle, B. G. Brown, and A. Holmes, 2006: Using in situ eddy dissipation rate (EDR) observations for turbulence forecast verification, Preprints, *12th Conference on Aviation Range and Aerospace Meteorology*, Atlanta, GA, Jan. 28 – Feb. 3, Amer. Meteor. Soc.

Vinnichenko, N.K., Pinus, N.Z., Schmeter, S.M., and Shur, G.N., 1968: *Turbulence in free atmosphere*, L., Hydrometeoiz-dut, 336 p. [in Russian].