[Responses to the Comment by the Anonymous Referee #1]

>> We deeply appreciate the referee#1 for providing constructive comments. The manuscript is revised following the comments below.

This study attempts to map the values of a turbulence diagnostic (derived equivalent vertical gust-DEVG) to a standard measure of turbulence intensity (cube root of the eddy dissipation rate-EDR). The motivation for obtaining such a mapping is that a limited fraction of commercial aircraft report either EDR or DEVG; therefore, establishing a correspondence between the two would allow turbulence data to be collected, studied, and exploited over an expanded global area using a single, consistent metric-the aircraft-type-independent EDR. The derived relationships are based on statistical comparisons of mutually exclusive data sets for DEVG and EDR over specific world regions. Although reasonable relationships result from the analysis, because DEVG and EDR measurements were not made simultaneously on the same flights, care must be taken to assure that apples are being compared to apples. Specific comments on this and other issues are given below.

Major Comments:

1) Section 2.2: QC procedures. Quality control is absolutely critical for the DEVG data. Fits to the PDFs of DEVG (e.g., Figure 9) is at the core of this analysis. The PDF tails are very sensitive to small changes in the bin counts. However, these tails are where the raw DEVG have quite large bin counts in some regions that manifest as secondary modes in the PDF. The post-QC PDF tail is the residual of the difference between fairly large numbers. Thus, it is important to justify the QC procedure.

a. Why are there so many invalid DEVG values? Are there documented case studies that show how these errors occur? And why do they occur primarily in certain regions?

As far as we know, there is no official document on the quality issue of the derived equivalent vertical gust (DEVG) data. Instead, in the Aircraft Meteorological Data Relay (AMDAR) observing system newsletters published by the World Meteorological Organization (WMO), the horizontal distribution of the DEVG has been reported since April 2016 (https://sites.google.com/a/wmo.int/amdar-news-and-events/newsletters/volume-11-april-2016). As in Fig. A1, the DEVG indicates relatively large values over some regions of Australia, New Zealand, and Europe, which is a consistent distribution with the current study. It is noted that the WMO newsletter did not mention the quality of the DEVG. The possible reasons of suspicious DEVG values can be a power loss of electrical power contactor and a bug in the DEVG initialization logic, which is related to intermittently added 1-*g*-bias, where *g* is the acceleration due to gravity (D. Body, personal communication, November, 2019). Considering that the DEVG data merged point samples from many kinds of flight over the globe, and time series of the DEVG and other recorded variables are not available, it is difficult to clearly identify the reason of suspicious DEVG values. A further investigation of the quality control (QC) procedures of the DEVG remains for future work, after obtaining the time series of recorded variables. A statement is included in the revised manuscript. [Page 6, Line 6-8]

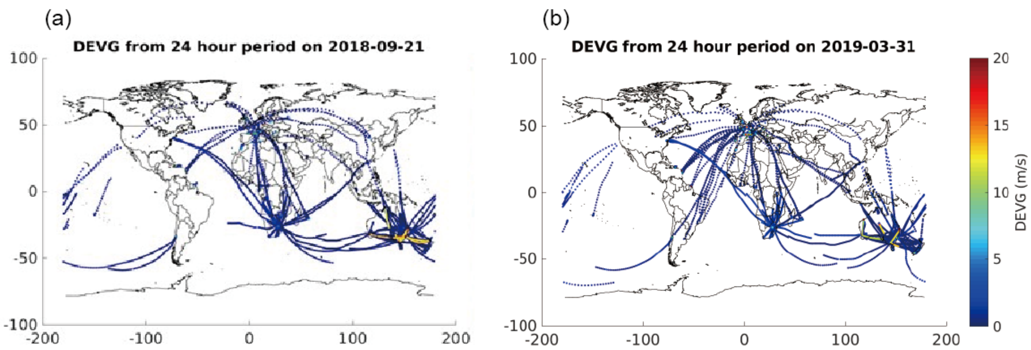

Figure A1. The horizontal distribution of the DEVG values for one day on (a) 21 September 2018 and (b) 31 March 2019, adapted from the WMO AMDAR observing system newsletter (https://sites.google.com/a/wmo.int/amdar-news-and-events/newsletters/volume-16-october-2018 and https://sites.google.com/a/wmo.int/amdar-news-and-events/newsletters/volume-17-april-2019).

b. Related to (a), can you provide physical justifications for the four steps of the QC procedure?

As done in previous studies (e.g., Gill 2014; Meneguz et al. 2016; Kim et al. 2017), the minimum flight time is considered, which is set to approximately one hour within an individual file that has the same flight tail number. As the time series of recorded variables are not available, we adopt an approach that uses a cluster of the DEVG data within a certain spatiotemporal window to increase a confidence of a turbulence event. An early version of the QC procedures in the current study is designed based on those by Gill (2014) and Meneguz et al. (2016) that used the Global Aircraft Data Set. These QC procedures to the Australian AMDAR data for four years (from 2011 to 2014) are revised through interactive discussions with two scientists affiliated in the Met Office (D. Turp and P. G. Gill, personal communications, February 2015). Again, these QC procedures are revised based on active discussions with scientists and forecasters associated in the Aviation Weather Center/NCEP (personal communications, from June to August 2018). During the QC procedures, we checked horizontal distributions of the raw and QC'd DEVG data when all moderate (MOD) and severe (SEV) turbulence events are reported. At least in the current study, the irrelevant turbulence events are discarded and a probability density function (PDF) of the DEVG follows the lognormal distribution, which is consistent with previous studies (e.g., Nastrom and Gage 1985; Frehlich 1992; Frehlich and Sharman 2004; Sharman et al. 2014; Kim et al. 2017). A statement is included in the revised manuscript. [Page 5, Line 29-33; Page 6, Line 17-19]

c. Page 5, lines 29-30 state that the threshold values used in the QC steps are empirically determined. This empirical process needs to be explained clearly in detail. How can you tell that too few or too many reports were not removed? This is crucial, because errors in this process directly affect the tails of the DEVG PDFs.

As mentioned above, the validity of QC procedures was discussed with the experts of the aircraft-based observations. In the raw DEVG data, only SEV (MOD) turbulence events were reported without MOD (LGT) turbulence event (Fig. A2), which is the most suspicious distribution. We checked all SEV and MOD turbulence events and excluded irrelevant turbulence events by comparing surrounding turbulence events. It is confirmed that the PDF of QC'd DEVG follows a lognormal distribution, as shown in Kim et al. (2017). A further investigation of the QC procedures of the DEVG remains to be accomplished in the future, after obtaining the time series of recorded variables. A statement is included in the revised manuscript. [Page 5, Line 29-33; Page 6, Line 17-19]

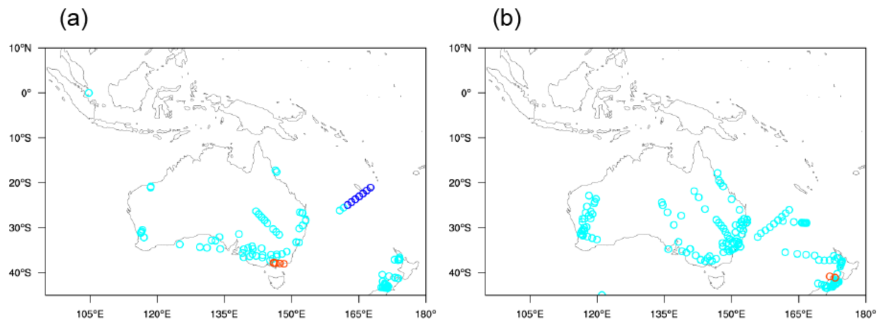

Figure A2. The horizontal distribution of the raw DEVG data for one hour of (a) 1000-1059 UTC 5 October 2015 and (b) 0100-0159 UTC 11 October 2016. The null (DEVG < 2 m s^{-1}), MOD (4.5 ≤ DEVG < 9 m s^{-1}), and SEV (DEVG ≥ 9 m s^{-1}) turbulence events are indicated as cyan, blue, and red circles, respectively.

2. Is parsing DEVG PDFs by northern or southern hemisphere the most meaningful and useful classification? There are reasons why the PDFs might differ for flights (a) over land vs. over ocean, (b) at different altitudes, (c) during different seasons, (d) during day vs. night, (e) in different latitude bands, etc.

As discussed in comment #5, the reporting frequency is changed when the aircraft passes the equator, which is indicated as an abrupt change in a data count around the equator. Considering the geographical difference in the data count, the PDFs over the Northern Hemisphere (NH) and Southern Hemisphere (SH) were computed and examined in the current study. During the revision process, we examined the PDF of the DEVG over land and ocean, different altitude ranges (15-25 kft, 25-35 kft, and 35-45 kft), seasons (spring, summer, autumn, and winter), day and night, and different latitude bands of a spacing of 20°. The mean and standard deviation of the natural logarithm of the DEVG (Table A1) are not significantly changed for aforementioned conditions, except that those in latitudes equatorward of 30° are clearly smaller than those poleward of 30°. This statement is included in the revised manuscript. [Page 8, Line 12-16]

Table A1. Values of the mean and standard deviation (SD) of the natural logarithms of the DEVG. The unit is m s^{-1}.

(a) Land or Ocean

	Land	Ocean
Mean	-1.1610	-1.1290
SD	0.7181	0.7881

(b) Altitude bands

	15-25 kft	25-35 kft	35-45 kft
Mean	-1.8025	-1.3114	-0.8702
SD	1.1149	0.7974	0.6106

(c) Seasons

	Spring	Summer	Autumn	Winter
Mean	-1.1985	-1.2953	-1.2812	-1.2448
SD	0.7776	0.8278	0.8124	0.8005

(d) Daytime or Nighttime

	Daytime	Nighttime
Mean	-1.2298	-1.2185
SD	0.7488	0.8277

(e) Latitude bands

	70°S-50°S	50°S-30°S	30°S-10°S	10°S-10°N	10°N-30°N	30°N-50°N	50°N-70°N
Mean	-1.9345	-1.2325	-0.6136	-0.8136	-0.4624	-1.6851	-1.8362
SD	0.9532	0.5945	0.4269	0.7049	0.5833	1.3877	1.2280

3. As noted in p. 2, line 22, DEVG estimates may be inaccurate during ascent or descent, and, thus, the data at cruise altitudes (> 15 kft) only are used. However, even above 15 kft, aircraft can change altitudes and direction that could affect the measurements. Why not restrict the use of data by only accepting estimates made during straight-and-level flight?

Thank you for pointing out this rather important issue. Unfortunately, the AMDAR data used in the current study do not provide the phase of flight which indicates 'level flight', 'ascending', and 'descending'. As in this study, Kim and Chun (2016) used 15 kft as the lower limit of altitude based on the time series of the DEVG and several variables recorded in the in situ flight data recorders. This statement is included in the revised manuscript. [Page 4, Line 19-21]

4. Figure 9 (and explanation in p. 7, lines 26-29). How do you justify throwing out some of the points in the PDFs for the fitting procedure?

At the highest bins, there are not enough data for reliable lognormal fits, while at the lowest bins, instrument noise may be affecting the result and the small DEVG values corresponding to nonturbulent conditions are not of practical interest. This statement is included in the revised manuscript. [Page 8, Line 8-10]

5. In Figures 1 and 8, there is an oddly abrupt change in the data count right around the equator over the central Pacific. There is much more data across a wider swath south of the equator. Is this real? What is the cause of this sharp transition?

The abrupt change in the data count around the equator can be found in the WMO AMDAR observing newsletter (Figs. A1 and A3), which is consistent with the current study (Fig. 1 of the original manuscript). For some reasons, when the aircraft passes the equator, the reporting frequency is changed from relatively lower to higher or higher to lower (Figs. A3 and A4). This difference in the reporting frequency between the NH and SH brings a sharp transition between two hemispheres. The abrupt transition across the equator may be related to systematic settings in aircraft-to-ground reporting during navigation. As the raw DEVG data indicate the abovementioned features, we would like to mention that this can be one of the characteristics of the DEVG data included in the AMDAR data and therefore the DEVG data can provide much more turbulence observations over the SH. This statement is included in the revised manuscript. [Page 4, Line 16-19]

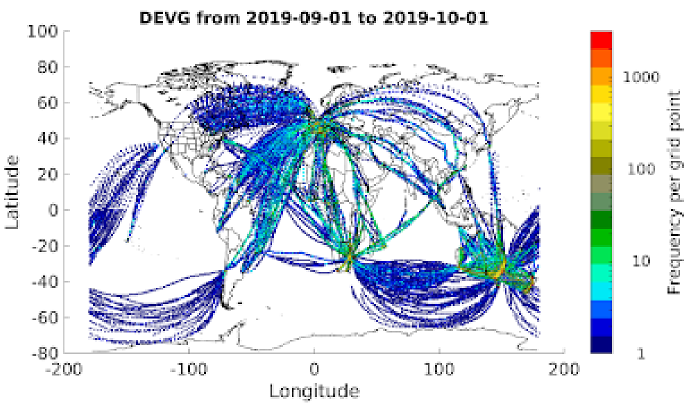

Figure A3. The horizontal distribution of the number of the DEVG data, accumulated within a 50×50 km grid box from 1 September to 1 October 2019. Adapted from the WMO AMDAR observing system newsletter (https://sites.google.com/a/wmo.int/amdar-news-and-events/newsletters/volume-18-october-2019).

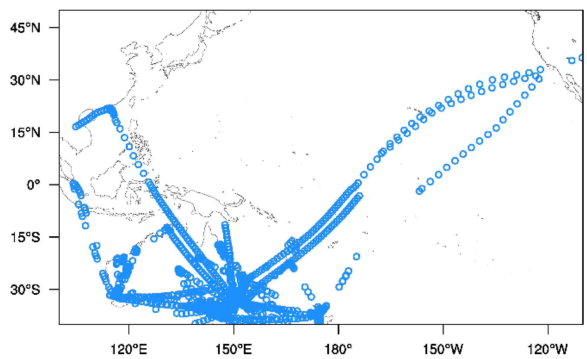

Figure A4. The horizontal location of the DEVG reports of the same flight tail number for one day (5 October 2015).

Minor Comments:

1. DEVG and EDR have different units. Is there a physical basis on which to make a unit conversion? Or is there an explanation of why it is acceptable to ignore the difference in units?

As written in the original manuscript (section 3.1), a simple mapping equation from a certain turbulence diagnostics *D* to the EDR was proposed by Sharman and Pearson (2017) using the observed in situ EDR data. First, turbulence diagnostics *D* and the EDR are standardized as:

$$X_1 = \frac{\ln(D^*) - \langle \ln(EDR) \rangle}{SD \ln(EDR)}, \tag{1}$$

$$X_2 = \frac{\ln(D) - \langle \ln(D) \rangle}{SD \ln(D)}, \tag{2}$$

where the angle brackets indicate an ensemble mean, SD is a standard deviation, D^* is the remapped EDR value corresponding to the raw turbulence diagnostics, and *EDR* is the in situ EDR values. Second, two standardized values, X_1 and X_2, are set to be equal and Eqs. (1) and (2) can be combined as:

$$\ln(D^*) = \langle \ln(EDR) \rangle + \frac{SD \ln(EDR)}{SD \ln(D)} \{ \ln(D) - \langle \ln(D) \rangle \}. \tag{3}$$

Therefore, Eq. (3) is written as the simple form as:

$$\begin{aligned} &\ln(D^*) = \ln(EDR) = a + b\ln(D), \\ &a = \langle \ln(EDR) \rangle - b\langle \ln(D) \rangle, \text{ and} \\ &b = \frac{SD \ln(EDR)}{SD \ln(D)}. \end{aligned} \tag{4}$$

In the current study, the turbulence diagnostics *D* is replaced with the DEVG value as:

$$\begin{aligned} &\ln(DEVG^*) = \ln(EDR) = a + b\ln(DEVG), \\ &a = \langle \ln(EDR) \rangle - b\langle \ln(DEVG) \rangle, \text{ and} \\ &b = \frac{SD \ln(EDR)}{SD \ln(DEVG)}. \end{aligned} \tag{5}$$

where $DEVG^*$ is the remapped EDR value corresponding to the DEVG value included in the AMDAR data. In this regard, it is acceptable to ignore a difference in units, although the EDR is a direct turbulence intensity metric, while the DEVG is not a direct turbulence intensity metric but a gust-load transfer factor.

2. Page 6, lines 6-7: It's not very informative that some MOD and SEV turbulence reports coincide with a high Ellrod1 index. Unless a statistical analysis is conducted to show a meaningful correlation, this remark should be omitted.

We agree with the reviewer and the sentence is deleted in the revised manuscript.

References

Frehlich, R., 1992: Laser scintillation measurements of the temperature spectrum in the atmospheric surface layer. *J. Atmos. Sci.*, **49,** 1494-1509, doi:10.1175/1520-0469(1992)049<1494:LSMOTT>2.0.CO;2.

Frehlich, R., and R. D. Sharman, 2004: Estimates of turbulence from numerical weather prediction model output with applications to turbulence diagnosis and data assimilation. *Mon. Wea. Rev.*, **132,** 2308-2324, doi:10.1175/1520-0493(2004)132<2308:EOTFNW>2.0.CO;2.

Gill, P. G., 2014: Objective verification of World Area Forecast Centre clear air turbulence forecasts. *Meteor. Appl.*, **21,** 3-11, doi:10.1002/met.1288.

Kim, S.-H., and H.-Y. Chun, 2016: Aviation turbulence encounters detected from aircraft observations: Spatiotemporal characteristics and application to Korean aviation turbulence guidance. *Meteor. Appl.*, **23,** 594-604, doi:10.1002/met.1581.

Kim, S.-H., H.-Y. Chun, and P. W. Chan, 2017: Comparison of turbulence indicators obtained from in situ flight data. *J. Appl. Meteor. Climatol.*, **56,** 1609-1623, doi:10.1174/JAMC-D-16-0291.1.

Meneguz, E., H. Wells, and D. Turp, 2016: An automated system to quantify aircraft encounters with convectively induced turbulence over Europe and the northeast Atlantic. *J. Appl. Meteor. Climatol.*, **55,** 1077-1089, doi:10.1175/JAMC-D-15-0194.1.

Nastrom, G. D., and K. S. Gage, 1985: A climatology of atmospheric wavenumber spectra of wind and temperature observed by commercial aircraft. *J. Atmos. Sci.*, **42,** 950-960, doi:10.1175.1520-0469(1985)042<0950:ACOAWS>2.0.CO;2.

Sharman, R. D., and J. Pearson, 2017: Prediction of energy dissipation rates for aviation turbulence. Part I: Forecasting nonconvective turbulence. *J. Appl. Meteor. Climatol.*, **56,** 317-337, doi:10.1175/JAMC-D-16-0205.1.

Sharman, R. D., L. B. Cornman, G. Meymaris, J. M. Pearson, and T. Farrar, 2014: Description and derived climatologies of automated in situ eddy-dissipation-rate reports of atmospheric turbulence, *J. Appl. Meteor. Climatol.*, **53,** 1416-1432, doi:10.1175/JAMC-D-13-0329.1.