

## ***Interactive comment on “Sampling Error in Aircraft Flux Measurements Based on a High-Resolution Large Eddy Simulation of the Marine Boundary Layer” by Grant W. Petty***

**Anonymous Referee #2**

Received and published: 19 October 2020

Based on data from a large-eddy simulation for a stratocumulus topped marine boundary layer, the author performed an ensemble of flight measurements and analyzed the convergence of the sampled fluxes towards their truth, as well as investigated the dependence of the random error on the track length. The author compares the observed random error against the theoretically derived expression by Lenschow and Stankov (1986) and found good agreement for track length of 10-30 km. Further, integral length scales of the turbulent quantities were calculated from the LES data for different flight track angles. The author shows that integral length scales depend on the flight track angle and compares these with the proposed, and still commonly-used, approaches by Mann and Lenschow (1994).

C1

The topic and the content of the paper fits well into the journal and is of high interest to the research community. Due to lack of any alternative ways to estimate the random error for flight measurements, almost everybody uses the expressions proposed by Lenschow and Stankov (1986), though it is already known that especially for shorter tracks the estimated random error often does not reflect the true uncertainty. Here, especially for shorter flight tracks, an improved random error estimation is highly desired in order to avoid misinterpretations of observations.

The paper itself is well written and results are sufficiently presented, though at some points in the text more information needs to be given. At several points in the text it is not clear how variables are calculated and from which data, i.e. from the full 3D LES data or from the sampled space-series along the flight tracks. Also, at some points the discussion is rather short and the findings are not well put into the context of previous research. For example, also the study by Schröter et al. (2000) had already analyzed how the sampled flux converges towards its truth with increasing track length. Even though the atmospheric setups are not comparable, the findings presented here should be put into the context of previous work.

After extensive review I can recommend the manuscript for publication only after major revisions have been done and extended analysis is presented. My major concern is outlined in the following.

(A) I miss one important aspect in the study. The author shows that the predicted random errors according to the Lenschow and Stankov formula matches the observed standard deviation remarkably well. However, in this study the random error using the LS86 formula is calculated based on integral length scales, correlation coefficients, and fluxes, that were inferred from the three-dimensional LES data. Though it is nice to show that the LS86 formula works well in the theoretical case where all these data is available, this approach does not reflect the reality at all. In reality, the integral length scale need to be calculated based on the flight-sampled data itself, hence, it is exposed to the same sampling errors as the flux is. Particularly for short track lengths,

C2

the errors in the integral length scales are supposed to be remarkably high. These biased values then propagate into the LS86 formula, with the consequence that the standard deviation of the predicted random error is also very high, with far reaching implications concerning the interpretation of the measurement, as the random error estimation cannot be trusted anymore for short tracks. In some situations the LS86 formula indicates high random error, but in other situations it indicates low random errors (please see the discussion in Sühring and Raasch (2013) about this). Hence, the comparison of the LS86-predicated error against the truly random error in the way you have done it here is actually not fair and does not help much in the interpretation of observed data. The manuscript would strongly profit if you add such an analysis. I would propose to add following analysis: \* How does the integral length scale and the correlation coefficient used in Eq. 5 depend on the track length? Here a direct comparison against the true value is possible since the integral length scales from the 3D LES data are available. \* How does the random error behaves when it is calculated directly from the sampled space-series, and how well does it match with the 'true' random error? Based on this, the performance of the LS86 formula can be directly shown for different track length, which would be extremely helpful for researchers.

Further comments:

Abstract - Line 2: vertical turbulent fluxes

Line 26 - 30: The paragraph is quite imprecise, it is not clear to what refinements done by Mahrt (1998) or findings by Hollinger and Richardson (2005) the author refer to. The introduction would profit if the author would elaborate this a bit more.

line 44: Please remove the word "exceptional". Though it is indeed quite a large setup, it is not exceptional any more. In the last one to two years such setups have become already standard in the LES community.

line 42-44: To the reader of the manuscript it might appear unclear what is meant by this sentence "The single most ...". I suppose the author mean that the smaller scales

C3

in the LES are filtered by the subgrid scheme, as well as by numerical errors, being not sampled adequately in an LES.

line 48-49: I disagree with the last part of the sentence. Emulating turbulence measurements in an LES always suffers from the missing subgrid-scale contributions, numerical errors etc., independent which grid resolution is used. Comparing measured turbulence spectra and spectra derived from LES-sampled data will always show a drop-off spectral energy on the smallest spatial scales (about 10 times the grid spacing as a rule of thumb), as it is also the case here (Fig. 2). The relevant question is if, and when, how much the missing subgrid-scale contributions affect the analysis of sampled turbulence data. I would recommend just to rephrase the part with the subgrid-scale flux here.

end of introduction: I miss a manuscript outline here, to guide the reader through the manuscript.

line 79-88: Beyond the fact that a marine boundary layer was simulated, what is the general atmospheric setup? Is this a boundary layer in the trade-wind zones or in a polar region (which latitude). Of course, according to Fig. 1 it becomes clear, but such information should be also given in the text.

line 93: How were these spectra calculated? Were they calculated from the emulated flights or from the 3D LES data (1D or 2D spectra)?

Fig. 2: Is it temperature or potential temperature? In Fig. 1 profiles of potential temperature are shown, later on the author only refer to temperature, though I cannot find any statement about a transformation.

line 98: With the phrasing "circular eddies" you indirectly imply that the turbulence is isotropic, which isn't the case as plenty of observations and simulation data show. I would recommend to simply remove the word "circular" here.

line 99-103: The author describes the spectra sufficiently here, though some refer-

C4

ences with respect of the minimum domain size of the LES domain are missing. However, I miss some further discussion about possible implications on further results. It is well known that at smaller wavelength the spectral resolution is bounded due to the subgrid-scale model as well as numerical dispersion and dissipation errors, causing these steep drop-off. In most cases this is no big issue as the smaller scales do not contain much energy. But also at the longer wavelengths, the spectral resolution of LES is bounded by the domain size, especially for humidity (de Roode et al. 2004). From previous studies it is known that structures grow in time, meaning that the spectral peaks move towards larger wavelengths, until the structures cannot grow anymore as they are bounded by the domain size. Somehow the spectra for  $q$  indicates this. In case this happens I suppose the integral length scales of humidity are also affected by this, though it won't change much on the overall results I guess. You already bring this up in line 135-137, but maybe it is worth to already bring this up here too.

line 91/104: Though it is only one author, the  $w\cdot$  form is repeatedly used throughout the manuscript.

Fig. 3: Again, is this absolute or potential temperature?

line 109-11: I disagree with this. The alignment with the mean wind is also visible in  $q'$  and  $T'$ .  $w'$  and  $q'$  (or  $T'$ ) correlate actually fairly well as it is typical in a convective boundary layer where the updrafts are created by buoyancy (which in turn is related to  $q'$  and  $T'$ ).

Eq. 1-2 and Fig. 7-8: Here it would be good to already mention that this quantity will be used for the integral length scale calculation. Without this background, which is not clear at this point in the manuscript, this might puzzle some readers. I am not entirely sure what these quantities are actually represent. To my understanding it is the local vertical flux at one point in time which can only be calculated from simulation data in the special case of a horizontally homogeneous boundary layer. To obtain heat fluxes in the traditional sense, time-averaging would need to be applied on top. Hence,

C5

$w'T'$  is not a real flux but a quantity used to compute integral length scales. I would recommend to make this clear in the manuscript, i.e. that these temporal local fluxes cannot be compared one to one to flux measurements from aircrafts or towers, neither with respect to the spatial pattern nor with respect to the amplitude. But in the context of the integral length scale calculation such equations and plots does make sense.

line 128: one dot too much

line 128-129 and Fig. 9: As the author have already mentioned in the text the momentum-flux profiles look quite noisy which is due to the lack of time averaging. Especially for the momentum flux longer averaging periods are required to obtain smooth profiles. Does the author have access to time-averaged profiles in order to check whether the shown momentum flux profiles changes much? According to the wind profiles shown in Fig. 1, I suppose the momentum flux profiles would decrease linearly with height. Here it would be worth to note this in the text and maybe a statement of how much the comparison against emulated flight measurements will be affected, as the truth in momentum flux is just a first guess.

line 133-134: I am wondering why the lower range for the 90%-contribution changes among the different height levels. To calculate an ogive you actually start at the smallest wavelength (or largest wavenumber) and integrate until the target wavelength. The smallest wavelength, however, is fixed by the cut-off wavelength which depends on the grid spacing, so the intervals should be always something like  $[2.5m \cdot x_1]$ ,  $[2.5m \cdot x_2]$ , etc.

Fig. 11 f-h: It is not entirely clear how the integral length scales for fluxes are obtained. I suppose you calculated  $I_f$  from the local fluxes obtained from equation 1-3, but it is nowhere written in the text. Or did you used equation 6-7 to calculate  $I_f$ ?

line 196-201: It would be nice to add the resulting length scales from equation 6-7 into Fig. 11 f-h to have this also visually.

C6

Section 3.3 and 3.4: The information content is not that big to put it into separate sections, you can easily merge it with section 3.2. But that's of course a matter of taste.

line 211: Please add the interval for  $n$  directly after the arctan-formula, this will make it easier to understand what is done here.

line 208-215: Before you start with your flight-track setup, you should first start to explain this special problem which is just because you use a stationary frozen flow field rather than a moving framework with advection where this problem will not occur unless the domain size is large enough. This will make it easier for the reader to understand what is done here.

Fig. 13: Just a minor point, but it would be nice to mention the flight direction - from southwest to the northeast, or was it the other way around?

line 234-240: I am not entirely sure what is the effect of the insufficiently sampled true value due to lack of temporal averaging, especially for the momentum flux which has been clearly not converged. But as you compare the flight-sampled data against these true flux, and both are calculated based on the same data set, I suppose that this won't have any effect on the following results. But interested readers might ask the same question, so it could be worth to spend one or two sentences here to make this more clear.

244-245: It is unclear which bias the author does mean here and how this is connected to the minimum separation distance.

line 262: Though you cannot mention it all here, at least some of these previous studies should be mentioned here directly.

Caption Fig. 17: Better to explain here directly what the black dots mean rather than to make a detour via the red curves in Fig. 14.

Fig. 17-19: please see my major comment

C7

Fig. 14-16: I assume the plot is based on data from all performed flight tracks? Actually this is not a valid approach to show the convergence of the measured flux towards its true value. For a given track length the data shows a certain spread, though, strictly speaking the data from the flight tracks is not comparable since the alignment of the flight tracks vary. In an atmosphere with fully isotropic turbulence the approach would be fine, but in an anisotropic boundary layer the scatter among the crosswind tracks might be different compared to the scatter among the parallel-wind tracks, i.e. the root-mean square might depend on the track angle. This should be stated clearly in the text so that readers are aware of this. For example, you can use different colors for the different track angles and provide the root-mean square values for each track angle separately (e.g. in an additional table), in addition to the "ensemble" root-mean square value you have already given. This might also indicate the optimal track angle with respect to the mean wind direction, supposed the statistics are sufficient.

Fig. 17-19: Is there a difference between parallel and crosswind tracks and any recommendation?

line 274-275: You are right with this when you refer to track length of about 10-30 km. But for even shorter track length the LS86 formula tends to underpredict the error as your data show. Actually, everybody knows that the random error for short track length is remarkably high, but flying only short tracks is sometimes the only way in some situations. I would suggest to be precise when you refer to "short" tracks, the understanding of "short" can be sometimes very different.

line 296: Here I miss some final conclusion. From the results shown, which version of integral length scale calculation does the author recommend?

References:

Stephan R. de Roode; Peter G. Duynkerke; Harm J. J. Jonker, 2004 Large-Eddy Simulation: How Large is Large Enough? *J. Atmos. Sci.* (2004) 61 (4): 403–421.

