# 1 General comments

First, I would like to extend my sincere thanks to both reviewers for their very thorough and thoughtful reviews. Being a newcomer to this particular area of research, the reviews drew my attention to important gaps in my familiarity with relevant prior work as well as giving me occasion to tighten up and clarify my discussions of both my methods and my findings. The revised manuscript that I will submit shortly will be greatly improved as a result.

In the following, the reviewer's comments are indicated by gray shading. My responses appear in the unshaded space below.

- Grant Petty

Regarding the scope of AMT scientific questions, the question of the publication of this study in this journal may be raised. Indeed, even if the topic is about airborne flux measurements, the study is based exclusively on results from numerical simulations. It is regrettable that no observations are used in this study, either to be confronted with the simulation output or to apply the results obtained, for example on past measurement campaigns.

According to the AMT's statement of scope, "Papers submitted to AMT must contain atmospheric measurements, laboratory measurements relevant for atmospheric science, **and/or theoretical calculations of measurements simulations with detailed error analysis** including instrument simulations." I relied on the highlighted phrase when deciding whether to submit to AMT. The other reviewer seemed to agree that it was appropriate, stating "The topic and the content of the paper fits well into the journal. . . "

The track definition used by the author can lead to flight tracks greater than the domain size thank to the cyclic boundary conditions of the LES. Nevertheless, as mentioned by the author, the finite LES domain is not able to reproduce structures greater than the domain size. Is a domain of 5.12 x 5.12 km2 is therefore large enough to study airborne sampling and eddy correlation flux estimation? With a larger domain size, the characteristics of the simulated turbulent structures may be different. With this issue of the limited size of the domain, does a LES with a larger mesh grid and a larger domain have been appropriate? Taking the example of the University of Wyoming KingAir aircraft mentioned by the author, with a true air speed of 85 m/s and a measurement frequency of 25 Hz, the sampling spatial resolution is then about 3.5m. Thus, a grid mesh 3 times larger than the one used here could be adequate.

The important point here is that I had no control over the parameters of the simulation. I didn't run the LES but rather requested the output from the model run previously published by Matheou (2018), and he was kind enough to provide it.

I agree that the domain dimensions aren't as large as would be ideal for this kind of study, but this particular simulation spanned a larger-than-average range of scales and seemed likely to be a better-than-average representation of turbulent fields at very low altitudes (e.g., 10 m or 40 m). My interest in undertaking this analysis was motivated in part by my involvement with very low- and slow-flying aircraft—e.g., ∼10 m altitude and 20 m/sec. Particularly at the lowest altitudes, I'm less concerned about the limitations imposed by the 5 km domain size. If we imagine that the goal is to accurately sample the near-surface fluxes within the limited domain, via repeated parallel passes (but without the complication of turning a physical aircraft), then what happens outside the domain seems less important.

There are many figures (19 in total), some of which seem redundant or could be concatenated. Several of them are simply mentioned in the text without being analyzed or discussed. The question of the relevance of these figures may arise, not helping to clarify the main message of the article. It obviously seems appropriate and necessary to present the simulation with the help of a few figures, however, it is only from figure n∘13 that the central purpose of the paper begins to be addressed.

I have consolidated several figures into single figures and deleted three others. There are now 11 rather than 19 separate figures.

# 2 Specific comments

## 2.1 Introduction

The works of Lenschow and Stankov (1986), Lenschow et al. (1994), and Mann and Lenschow (1994) were not only based on theoretical considerations and statistical models but also on observations. It might be useful to include in the introduction, some studies on experimental data and field campaign. In general, the introduction could be enhanced in terms of bibliographic references, such as Brooks and Rogers (1997) Cook and Renfrew (2015) or Brilouet et al. (2017).

I will add those references and try to add appropriate context. That said, as a complete newcomer to boundary layer meteorology, there is a danger that I will mistate or misinterpret something important, so my inclination is to let the cited references speak for themselves as much as possible.

Line 34: The LES is able to resolve explicitly the major part of the turbulence but it remains a sub-grid contribution. Even if with a 1.25m resolution, this contribution becomes rapidly negligible with the altitude, it might be useful to mention that total turbulence = explicitly resolved + subgrid contribution.

I now extend the sentence in question to say, “leaving only a small fraction of the total turbulent exchange to subgrid-scale parameterizations, especially at levels much above the surface.”

After line 42, it is not clear if we are still in the introduction section or if the section ”description of the method” has already started. It would be useful if the main goal of the study could be more clearly highlighted and if an outline of the article were provided at the end of the introduction before going into the details of the simulation and the method.

I have significantly reorganized the introduction and data sections, including the addition of an outline of the article.

Line 60: It is correct that using a LES to examine the aircraft flux sampling problem in MABL is unique. Nevertheless, it can be mentioned that previous studies compared LES outputs with airborne measurements such as Brilouet et al. (2020) even if the resolution was coarser.

I have added a mention of that paper a bit earlier in the introduction.

## 2.2 Data

The case study is from the field campaign DYCOMS-II, Are there any observations that might be relevant to the study?

A variety of airborne measurements were taken, as described in part by Stevens et al. (2003). I have not attempted to acquire these measurements or to independently validate the LES, which would be a major effort in its own right. Some discussion of the realism of the LES results is given by the creator of the LES model in Matheou (2018).

The case study is a nocturnal cloud-topped marine boundary layer. When the author describes the environment, a few elements describing the main characteristics of this type of stratocumulus condition could be instructive for the reader (such as zi at the cloud top, the strong inversion with entrainment at the cloud top, . . . ).

Since the top of the cloud layer coincides with the inversion———and thus the top of the boundary layer———at about 840 m (the maximum height for any non-zero cloud water anywhere in the domain is 885 m), zi at cloud top is basically 1. The strong inversion with cloud top entrainment is characteristic of marine stratocumulus clouds in this region. However, because my analysis in the paper is focused entirely on the clear-air portion of the boundary layer well below cloud base, I have chosen to omit details concerning in-cloud or cloud-top processes.

Figure 2: the figure is rather small. The units of the power spectra are not mentioned. Does it might be interesting to present normalized spectra (by the variance: $kF(k)/\sigma_X^2$)? Does the spatial wavelength is $\lambda = 1/k$ or $\lambda = 2\pi/k$?

I apologize for the small figure. I had accepted the default scale parameter provided in the AMT style template but should have increased it.

The power spectrum is initially computed as a function of inverse wavelength, meaning the wavelength $\lambda$ of a complete cycle. As indicated in the axis labels, I transformed the spectra in the plots so that the horizontal axis is $\lambda$ rather than spatial frequency. The amplitude plotted as $kF(k)$ without normalization, so the units should be variance per unit $\log(\lambda)$, with the units of the variance depending on the variable plotted. I did not normalize; doing so has no effect on the curves other than shifting them vertically. The focus in these plots is on the slope. I later realized that I omitted the conventional factor of $2\pi$, so I will correct the notation on plots to make sure it is consistent with common usage.

Line 95: It might be interesting to compare with previous works.

Again as one very new to this subject area, I am not sufficiently familiar with prior work looking at wavelengths of peak energy in turbulence to be able to quickly identify the most relevant studies. I apologize. I hope that others who do have that familiarity will be able to interpret my results within the context of previous findings and/or suggest studies that should be cited.

Lines 96-98: At 40m height (0.05 zi), this is the surface layer. How much the turbulence is explicitly resolved at this height? What is the vertical profile of TKE resolved / total TKE? Also, the surface layer may have different characteristics than the layer above. Does the Monin-Obhukhov Similarity theory (MOST) is available? It would be interesting to enhance the discussion with some references on the turbulent structure inside the surface layer such as Katul et al. (2011) or Sun et al. (2016).

The output file I received from Dr. Matheou does not include the parameterized subgrid components of the TKE or fluxes, so I can't directly answer that reasonable first question. However, I understand from casual conversations, possibly incorrectly of course, that the resolved component of the turbulence should dominate once you're much above 5 or so grid levels, where $\Delta z = 1.25$ m. The flux sampling error analysis depends only on realistic spatial statistics of the turbulence on the scales containing most of the energy. The subgrid component of the turbulence would likely correspond to a more or less linear extrapolation of the plotted spectra below approximately 10 m wavelength. As a fraction of the total variance, that extropolation doesn't add much, and I don't think it would much change the general findings in this paper either, especially at 40 m and higher.

Regarding turbulent structure of the surface layer, I'll admit that I'm not a boundary layer theoretician, so I wouldn't be qualified to undertake that discussion.

Lines 99-100: Do the spectra of temperature and specific humidity reveal more energy at longer wavelength due to the influence of mesoscale on those parameters? If the domain was larger, would the wavelengths be longer?

Satellite images clearly reveal variations in stratocumulus cloud decks on larger scales, but I don't know what fraction of the total variance in velocity, temperature, or humidity is represented by those longer wavelengths. The safest way to interpret the present paper is as simulating airborne flux measurements over a restricted 5 km domain, so that things happening beyond the lateral boundaries aren't really relevant to the basic question under consideration.

Line 101: What is the reason that the horizontal wind speed spectrum has no significant dependence on height?

Again, I'm not a boundary layer theoretician, so I can't be sure of the answer. My suspicion is that because horizontal flow isn't obstructed by the upper or lower boundaries, and because there is relatively little friction near the ocean surface, vertical mixing within this nearly neutrally stratified BL is efficient enough to maintain a fairly homogeneous spectrum.

Lines 104-105: The author has chosen four representative heights, one at 10 m and another at 40 m. Are these heights characteristic of airborne measurements?

In the paper, I now mention examples of airborne measurements at 40 m (Cook and Renfrew 2015) and at 100 m and 400 m (Desai et al. 2020, in press). 10 m is also of interest to me in light of possible future turbulence measurements from a very low-flying ultralight airplane or drone.

Figures 3-5: 3 figures are considered for 4 lines. It would be interesting to concatenate them into a single figure. It will be easier to compare the characteristics of each parameters and their evolution with the height (for example with left panels at 10 m, middle panels at 40 m and 100 m and right panels at 400 m with a parameter by row).

This is a good suggestion, and I have merged them into a single figure with rows corresponding to height and columns corresponding to variables.

Lines 106-110: Also, a link with previous work would be valuable.

While the results I describe seem consistent with my expectations, I realize now that I cannot point to a specific study describing smaller-scale turbulent structures in the neutrally stratified marine boundary layer. I would be happy to add any that are suggested to me. In the meantime, I have removed the word "expected" from my description.

Figure 6: Is this figure really essential to the article?

I have deleted that figure.

Line 114: It might be helpful to define the sensible (H) and latent (E) heat fluxes. Commonly, the E notation refers to the surface moisture flux or evaporation ($E = \rho \times w'q'$). Perhaps the LE or LvE notation is more appropriate for the latent heat flux.

I have changed the notation to $LE$.

Lines 114-119: Is the definition of sensible and latent heat fluxes and their expressions as a function of fluctuations valid at different altitudes in the boundary layer? Is it not defined only for surface exchanges? The sensible heat flux is the amount of heat exchanged between the surface and the atmosphere and the latent heat flux represents the energy released or absorbed during a phase change. I may be mistaken and in that case, I apologize for this unwelcome comment.

The actual exchange of both heat and latent heat *at the surface* is of course a non-turbulent (diffusive, molecular) process. Turbulent transport within the atmospheric becomes the dominant process once you get a short distance above the surface. The eddy covariance method is inherently a measurement of vertical turbulent flux through a plane *at the level of the instrument*, whether close to the surface or higher up. Depending on stratification, storage, etc., it may or may not be an adequate representation of the surface flux.

Figures 7 and 8: These figures are not described or analyzed in the article. Are they essential to the article?

I have deleted those figures.

Line 130: It would be interesting to explain the TKE profile and how this is expected, in terms of the processes involved, given the case study under consideration. Here again, a connection with previous studies on this subject would be appreciated.

Once again, I'm not a boundary layer theoretician, and I'll admit to not being familiar with previous studies on that particular subject. That said, it seems to me that with viscous dissipation of TKE occurring only at the smallest scales (and therefore presumably being relatively slow), and with this neutrally stratified boundary layer readily mixing in the vertical, it might make sense that TKE would be almost uniformly distributed through that depth once you get well below the radiative forcing at the top of the cloud layer.

## 2.3 Integral length scales

In this section, the work of Lumley and Ponofsky (1964) could enhance the bibliography as a pioneer on these issues.

While I have this book in my private list for future reading, I haven't ever held a copy in my hand, and I'm currently unable to access a library copy due to the COVID shutdown. I hesitate to cite a source without stating what I'm specifically citing it for. My apologies.

Line 140: It is the first time, since the introduction, that the random error is mentioned. As this is the main focus of the article, wouldn't it be a good idea to highlight it further? The current design of the article suggests that it is secondary to the integral scales.

I will expand the discussion of the random error and introduce parts of that discussion earlier in the manuscript.

Line 149-150: To introduce the random error in a simplified point of view, is the equation 1 of Lenschow and Stankov can be relevant?

As a definition, yes. See above.

The spatial correlation $\rho_{w\psi}$ is defined twice (line 156 and line 164).

Thank you.

Line 158: In order to specify the experimental difficulties in estimating the integral length scale, the study of Durand et al. (2000) could be instructive.

I have added that reference.

Figure 12: Even if the random error definition contains the correlation $\rho_{w\psi}$ is the figure really essential to the article? Simulated aircraft measurements

I think it's worthwhile to show these since the values are utilized in the error determinations. In particular, I make reference in a couple of places to the fact that $\rho_{w,U}$ crosses zero near 400 m to explain why momentum flux is small there and the error factor becomes very large.

Lines 208-209: This sentence perfectly summarizes the main topic of the study. Isn't it a bit late? This message does not appear clearly enough throughout the article.

In the revision, I have tried to make that clearer in the introduction.

Lines 246-247: As mentioned in the general comments, I have some concerns about the domain size with respect to the characteristic scales of fluxes that can be observed during airborne measurement campaigns. Consequently, the results that will arise from this study seem difficult to be transposed to measurement campaigns.

Yes, the domain size is a limitation. More nearly ideal would be a 50 km domain with 1 meter resolution, but that is not currently feasible, and I am in any case working here with someone else's LES output. I do not currently have a better source of model output for this particular kind of analysis, but I believe the analysis offered here is a small step in the right direction, provided that the unavoidable limitations, especially with respect to large-scale contributions to fluxes, are kept in mind.

Line 249: Another way to check Taylor's hypothesis, for airborne measurements, the true air speed (here V = 85 m/s) can also be compared to the intensity of the turbulence $(\overline{u'^2})^{1/2}$. If $V \gg (\overline{u'^2})^{1/2}$ then the statistical properties of the turbulence field are assumed to be unchanged over the considered time interval.

This may be true, but I haven't heard it before, and I don't know whom I could cite as a source for that relationship. I will continue thinking about why this statement is valid, as I'm not immediately seeing the reason. If true, then we're talking about the standard deviation of $u'$, which was shown in Fig. 6c (now deleted in the revised manuscript) to range from around 0.5 to 1.0 m/sec.

## 2.4 Results

Figures 14-16: These three figures could be concatenated into one. Moreover, even if these figures are at the core of the study, they are barely detailed and analyzed (Figure 15 is barely mentioned).

I agree with this suggestion and have combined them into one figure.

Line 261: Including bibliographic references would be valuable.

I will look for appropriate references.

Figures 17-19: In order to facilitate the understanding of the figures, it can be useful to keep the empirical RMS error in red rather than changing the color. Are the parameters in blue necessary? If so, would it be better to include them in a table? As the minimum track length L10 for 10% relative accuracy is one of the main results, would it be a useful to group them together, for each flux and each altitude, in a table?

Again good suggestions. I have combined the three figures into a single figure, changed the colors, and moved the parameter and L10 values to tables.