Comment on amt-2021-86

Anonymous Referee #1

Referee comment on "Novel approach to observing system simulation experiments improves information gain of surface–atmosphere field measurements" by Stefan Metzger et al., Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2021-86-RC1, 2021

General comments: This is a well-written and interesting analysis of two hours of LES data used to determine optimal aircraft observing strategies for measuring surface latent and sensible heat fluxes when used in combination with a network of surface measurements. One concern is that the recommendations in terms of aircraft track angles and density of tracks is likely to be a function of atmospheric conditions, and the 2 hours analyzed only provide one very limited snapshot of the range of possible conditions. For example, the longitudinal wind streaks present in the LES simulation are known to be a function of wind speed and stability, and in a deep convectively driven boundary layer will not be present. Second, the authors make no mention of clouds, especially boundary layer clouds, which are frequently present in the upper-midwest in summer. Because the manuscript makes no mention of clouds at all, I assume that they did not occur in the simulation. Boundary clouds will certainly change the entire turbulent structure of the ABL. Thus, the results that the authors present are for a very limited range of meteorological conditions. Given the extensive computational resources needed just for this single 2h analysis, it probably was impossible to include these other situations. However, at a minimum, it is essential that the authors point out this limitation to their analysis at the beginning of the manuscript. As currently written, I felt I was misled through most of the manuscript into believing that a more comprehensive analysis of meteorological environments were being evaluated, only to find out later that this was not the case.

Line 99. Please describe what a dispersive flux is.

Fig. 3c. What do the white areas in the figure denote?

Line 224. What is meant by "super-sampled"?

Line 225. How were the candidate OSDs determined? How was the number of such OSDs to be used determined?

Line 231. What are the conditions that prevent ERF from providing a result at a given location?

Line 234. How are the area fluxes determined when there are missing cell data in the ERF domain? Is the ERF spatial average just the average of those cells that have data?

Line 236. Is the single score an average for all meteorological conditions? For example, I would suspect that the optimal flight tracks might be very different for days with boundary layer cumulus versus clear sky, or early morning/late afternoon transition times compared to mid-day. (OK, later I see that only one 2 hour mid-day period has been analyzed. It would be helpful to the readers if this was mentioned earlier in the analysis, even in the abstract).

Line 282. Were there clouds on this simulated day?

Section 2.5. Lots of complex details of the methodology are given here, but what I do not see discussed in general terms is how one defines the aircraft fluxes (usually an average along one or more flight legs) and then incorporates that data to be able to derive highly resolved spatial maps of the fluxes. A couple of sentences describing the basic principles behind this methodology at the start of the section would be beneficial for readers who are not experts in the technique.

Line 305. A couple of additional sentences here describing the ERF methodology would be useful. If length of the manuscript is a limitation, I would suggest removing some of the philosophical discussion in the introduction in to order to leave some room here.

Line 351. How was the number of 13 OSDs determined? Was it limited solely by computer resources, or was there any analysis of incremental benefits tapering off with increasing numbers of OSDs?

Line 370. The noontime value of 400m for the ABL depth seems surprisingly low. In retrospect, is the day that was simulated here representative of the boundary layers actually observed during the field program? And how was this particular day selected for the analysis? (OK, I see on line 615 that this is due to an error made in the initialization of LES humidity profile. This error should be mentioned briefly here on line 370, so that readers immediately understand the reason for the low ABL height, instead of wasting time wondering about it while working through the rest of the manuscript).

Line 434. I am surprised that the range of LST in figure 10a is only 0.1K!   Is this range meant to reflect the actual range of LSTs over the 10x10km domain?

Line 519. The phrase "doubling the scientific return" seems a little grandiose. A more accurate statement would be something along the lines of doubling the accuracy of the spatial sensible and latent heat flux estimates. The scientific return of the measurements taken during the CHEESEHEAD field campaign will be determined many years down the road when all of the analyses of the data set are completed and the papers published. In addition, in view of the fact that the analysis covered only one particular meteorological condition, "doubling the scientific return" seems really an exaggeration.

Line 678. "und" should be "and".