Author reply to Anonymous Referee #1 comments on the manuscript

AMT-2021-86

"Observing system simulation experiments double scientific return of surface-atmosphere synthesis"

by S. Metzger et al.

We thank Anonymous Referee #1 for the valuable feedback, which will substantially improve the manuscript. Please find below the referee comments recited in *blue italics font*, followed by our point-by-point replies in black upright font.

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.

Author reply:

- Many thanks for sharing these concerns! Indeed we initially aimed to simulate multiple days, as well as morning and afternoon periods. However, as we were racing against the clock to get the actual experiment running, it proved challenging to correctly emulate all candidate observing system designs in all LES runs. As we reached a critical cutoff point we needed to focus on transitioning to the ERF analyses with a single LES run in order to provide numerical results in time for supporting the CHEESEHEAD19 experiment design. With more lead time and computational resources it is possible to realize more LES runs.

- In front of that background, we chose to focus on an LES run that is both typical for the region/season and one that likely generates significant heterogeneity (without the added expense of dealing with clouds). Further considering that for safety reasons the real-world flights were to take place on mostly cloud-free days, we believe the selected case provides a useful case study for how much airborne flight leg choices influence our ability to address CHEESEHEAD19 science objectives.
- While studies with multiple LES runs are available in the literature (e.g., Sühring et al., 2018), it does not appear untypical for case studies to encompass a single LES run, either (e.g., Xu et al., 2020).

<u>Proposed changes to the manuscript:</u>

- Clearer separation throughout the manuscript between (i) the novel, general Numerical Simulation - Environmental Response Function (NS-ERF) approach to OSSEs and (ii) the specific CHEESEHEAD 19 airborne case study in the form of LES-ERF.
- Mention in the Abstract, Introduction, Materials and methods, and Discussion that the CHEESHEAD19 case study is limited to two hours of LES data with a single meteorological setting and omission of clouds. Include the specific limitations outlined by the reviewer in the Discussion section.

Line 99. Please describe what a dispersive flux is.

<u>Author reply:</u> Dispersive fluxes refer to the scalar transported by standing eddies or spatially organized time-invariant convection cells (e.g., Margairaz et al., 2020; Raupach and Shaw, 1982).

<u>Proposed changes to the manuscript:</u> Add above description of dispersive fluxes.

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

<u>Author reply:</u> Sect. 2.5 Environmental Response Functions explains: "*We also limited the ERF projection to interpolate but not extrapolate outputs, i.e. to only populate grid cells with driver combinations in the range of the virtual measurements. By doing so, the resulting data coverage lets us directly estimate how well we sampled the domain for upscaling.*"

<u>Proposed changes to the manuscript:</u> Add corresponding explanation to Fig. 3c caption.

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

<u>Author reply:</u> We were trying to colloquially describe "observing system designs".

<u>Proposed changes to the manuscript:</u> Contextualize "super-sampled" with "13 different observing system designs".

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

Author reply: The subsequent Sect. 2.3 Design hypotheses and candidate observing system designs details the rationale for creating the OSDs.

Proposed changes to the manuscript: Insert cross-reference to Sect. 2.3.

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

Author reply: Sect. 2.5 Environmental Response Functions explains: "*We also limited the ERF projection to interpolate but not extrapolate outputs, i.e. to only populate grid cells with driver combinations in the range of the virtual measurements. By doing so, the resulting data coverage lets us directly estimate how well we sampled the domain for upscaling.*"

Proposed changes to the manuscript: Add explanation and cross-reference "*…the percentage of grid cells across the study domain that ERF was able to reconstruct within the range of the virtual driver measurements (Sect. 2.5)*".

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?

Author reply: That is correct.

Proposed changes to the manuscript: add clarification "*…horizontal average over all reconstructed grid cells…*".

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).

Author reply: Agreed, please see our responses to above general referee comments.

Proposed changes to the manuscript: Please see our proposed changes in response to above general referee comments.

Line 282. Were there clouds on this simulated day?

Author reply: No, please see our responses to above general referee comments.

Proposed changes to the manuscript: Please see our proposed changes in response to above general referee comments.

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.

> Author reply: Thanks for pointing out the need for providing a brief overview!
>
> Proposed changes to the manuscript: From line 312 add a new overview paragraph with a sequential outline of the methodological steps. Point out principle differences to traditional tower flux (here 1 min vs. traditional 30 min resolution) and airborne flux (here 100 m vs. traditional ~10 km resolution) data processing. Describe how these differences facilitate joining tower and airborne fluxes into space- and time-aligned datasets for machine learning and subsequent flux map projection.

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.

> Author reply: Agreed, please see our response to above referee comments on Sect. 2.5.
>
> Proposed changes to the manuscript: Please see our proposed changes in response to above referee comments on Sect. 2.5.

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?

> Author reply: Sect. 2.3 Design hypotheses and candidate observing system designs details the rationale for creating the OSDs. A value engineering analysis to determine an optimal number of OSDs was not performed. Instead, we created a sample of OSDs that symmetrically covers the trade space presented by the two design hypotheses, and was bracketed by time and computational limitations. Within these limitations, we believe the sample of OSDs provides a useful case study for how much airborne flight leg choices influence our ability to address CHEESEHEAD19 science objectives. We endorse the suggestion of a value-engineering preceding the creation of OSDs, and with more lead time and computational resources it is also possible to realize additional OSDs.
>
> Proposed changes to the manuscript: Add the suggestion of a value-engineering analysis preceding the creation of OSDs to Sect. 4.3 Remaining challenges and future work.

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).*

Author reply: Agreed, thanks for the pointer!

Proposed changes to the manuscript: Add cross-reference to Sect. 4.1 where the issue is discussed in detail.

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?

Author reply: Sect. 2.5 Environmental Response Functions explains "***The source area weights provided a linkage*** *between the sensible and latent heat flux responses in the atmosphere and their spatially resolved drivers at the LES surface (available energy as a proxy for net radiation) and in the* ***first vertical LES layer (land surface temperature*** *and moisture as a* ***proxy for remote-sensing observations****).*" So while our LST proxy from the first vertical LES layer (10 m resolution) retains the horizontal spatial patterning (25 m resolution), the amplitude of this near-surface air temperature can be expected to be reduced compared to actual surface temperature. Actual surface temperature was not available from the LES run because fluxes were prescribed directly as lower boundary condition in lieu of a land surface model. Furthermore, we constructed ERF response surfaces from source-area-averaged land surface temperature and moisture. This further reduced their amplitude and also explains some of the white gaps in Fig. 3c.

Proposed changes to the manuscript:

Following the suggestion from referee 2, change "land surface temperature (LST)" to "near-surface temperature (NST)", and "land surface moisture (LSM)" to "near-surface moisture (NSM)". Clarify figure caption: "*Example ERF response surfaces. (a) Sensible heat flux as a function of* ***source-area-averaged*** *energy input and near-surface temperature (NST* ***from the first vertical LES layer, Sect. 2.5****). (b) Latent heat flux as a function of* ***source-area-averaged*** *energy input and near-surface moisture (NSM* ***from the first vertical LES layer, Sect. 2.5****). For this visualization, all other drivers are kept at their median value.*" Add a paragraph to Sect. 4.1 Optimizing the CHEESEHEAD19 observing system design: discuss the amplitude ramifications of using first vertical LES level LST and LSM, as well as source area averaging and potential solutions for future applications.

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.*

Author reply:

- We thank the reviewer for this observation and agree that – in particular when taken out of context – the term "scientific return" in itself could be interpreted as grandiose, which was not our intent.
- We feel there might be two common root causes for this and a similar concern by referee #2. First, the current manuscript structure lacks separation between (i) the novel, general NS-ERF approach to OSSEs and (ii) its maiden, example application to the CHEESEHEAD19 airborne case study in the form of the specific LES-ERF realization. Second, we introduce NS-ERF-specific terminology in the text (e.g., optimality criterion, hypothesis, objective, observing system design, **scientific return** etc.). However, we do not currently provide a clear conceptual overview of the relationships among these terms, thus creating a lack of context for interpreting the CHEESEHEAD19 case study results.

Proposed changes to the manuscript:

- Replace "scientific return" with "information gain" throughout the manuscript.
- Re-organize manuscript to more clearly distinguish between (i) the novel NS-ERF approach to OSSEs and (ii) its maiden, example application to the CHEESEHEAD19 airborne case study in the specific form of LES-ERF.
- Reflect this shift by rewording the title to "Novel approach to observing system simulation experiments improves information gain of surface-atmosphere field measurements". Specifically, shift title emphasis from case study results (replace "*…double scientific return…*" with "improves information gain of surface-atmosphere field measurements") to creating an extensible OSSE framework for designing surface-atmosphere observing systems (add "novel approach").
- Reword Sect. 2.2 header to "NS-ERF observing system simulation experiments". Add a flowchart/"visual glossary" to the beginning of the section for inter-relating generalized NS-ERF components/terminology including the aim to maximize information gain. At the end of the section (following "*We then used the arithmetic mean and standard deviation to aggregate CR1–CR3 across flight patterns, flight sequences, and ultimately among themselves into a single score (Sect. 3.3).*"): define information gain specifically for the CHEESEHEAD19 case study as the arithmetic mean of three optimality criteria that directly correspond to CHEESEHEAD19 science objectives.
- In above Sect. 2.2 and discussion Sect. 4.1 Optimizing the CHEESEHEAD19 observing system design: clearly distinguish the potential to improve the information gain of CHEESEHEAD19 airborne measurements as evidenced by the 2-hour case study from the actual scientific return of the measurements taken during the CHEESEHEAD field campaign.

Line 678. “und” should be “and”.

Author reply: Thanks for catching this.

Proposed changes to the manuscript: Correct spelling.

References

Margairaz, F., Pardyjak, E. R., and Calaf, M.: Surface thermal heterogeneities and the atmospheric boundary layer: The relevance of dispersive fluxes, Boundary Layer Meteorol., 175, 369-395, doi:10.1007/s10546-020-00509-w, 2020.

Raupach, M. R., and Shaw, R. H.: Averaging procedures for flow within vegetation canopies, Boundary Layer Meteorol., 22, 79-90, doi:10.1007/BF00128057, 1982.

Sühring, M., Metzger, S., Xu, K., Durden, D., and Desai, A.: Trade-offs in flux disaggregation: a large-eddy simulation study, Boundary Layer Meteorol., 170, 69-93, doi:10.1007/s10546-018-0387-x, 2018.

Xu, K., Sühring, M., Metzger, S., Durden, D., and Desai, A. R.: Can data mining help eddy covariance see the landscape? A large-eddy simulation study, Boundary Layer Meteorol., Online First, doi:10.1007/s10546-020-00513-0, 2020.