Author reply to Anonymous Referee #2 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 #2 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

Metzger et al present an article where they describe the importance of using observing system designs (OSD) to maximize scientific insights from surface air exchange measurements. The authors propose that one particular observation system design that is useful for the study of biosphere-atmosphere interactions is the Observation System Simulation Experiment (OSSE), which is obtained through the use of Large Eddy Simulations (LES) in combination with Environmental Response Functions as previously described by Metzger (2013). Thus, the OSSE used in this study is referred as LES_ERF..

The authors argue that most researchers in the field set up the instrumentation first to gain "knowledge from data", however, an alternative is to use OSDs such as this OSSE to gain "data from knowledge" and thus optimize the scientific insight gained from the observations. For example, this type of optimization can be used to determine an optimal location for an eddy-covariance sensor both vertically, and horizontally, thus guaranteeing the adequate fetch but also guaranteeing the correct spatial heterogeneity that may account for mesoscale circulations.

The goal of the paper is to show that making informed Observing System Designs (OSDs) for surface-atmosphere field measurements can improve the amount of useful information, or as the authors say, "double the scientific return". And that creating OSSE can be one of the best ways to attain an optimal OSD. The novel approach is to provide design information prior to testing OSDs in the field.

The authors make use of the CHEESSEHEAD19 dataset, which was originally designed to test the hypothesis that mesoscale features, driven by surface heterogeneity can explain the lack of closure in energy balance. The campaign was also used to evaluate the use of Environmental Response Functions (ERFs) for estimating "fluxes in a box" an approach previously created by the authors. In this paper, the authors use LES_ERF to find the optimal flight strategy to *maximize the amount of useful information to evaluate the energy balance closure at the CHESSEHEAD19 site. The OSDs are built around 2 hypotheses, one is that it is critical for airborne EC to measure perpendicular to the prevailing wind, and 2, that it is more informative to fly a finely spaced pattern. 13 different OSD were created: the first based on tower-only data, and the other 12 based on the combination of four track angles (0,45,60,90 °), and 3 flight patterns (Alternating, Outbound, Return). The results show that flight patterns with a track angle of 45-90 double the percent improvement based on three optimality criteria when compared to parallel flight patterns.*

Author reply: Many thanks for this excellent summary.

I think the paper addresses relevant scientific questions within the scope of Atmospheric Measurements Techniques, and that it presents novel concepts, ideas, and tools. However, I think there are some issues with the clarity of the methodology and with the application of this technique to other studies.

Author reply: Many thanks for sharing these concerns, which we address individually below.

The methodology can be very confusing. This is what I understood from section 2.2. The methodology consists of first combining information from different sources into ERFs. This would be, for example, combining information from existing towers and aircraft data to create a space-and-time aligned dataset. The second step is to then create multiple OSDs by using LES, and finally benchmark the candidate OSDs by evaluating if the candidate OSD can recreate the environment from the original ERF. One of the 3 ways to evaluate this approach, is the Energy Balance Ratio, defined as the sum of the sensible and latent heat fluxes produced by the ERF, to the sum of latent and sensible heat fluxes produced by the LES. It is my understanding that the numerator in equation 1 does not vary and that the denominator should vary according to different OSDs. But I may be wrong. This section could use some synthesis to make it clearer to the reader. I think there are some conflicts between what is said in the caption and what is said in section 2.2.

Author reply:

- Many thanks for pointing out apparent inconsistencies in Sect. 2.2. We believe that these arise primarily from two ambiguities:
- First, Figure 3 initially introduces ERF (Figure 3a) and then LES (Figure 3b), rather than following the actual sequence of LES-ERF data processing (first LES, then ERF).
- Second, we directly prescribed time-dependent, heterogeneous sensible and latent heat flux grids as LES surface forcings. For this we utilized the pre-existing sensible and latent heat flux grids of Metzger et al. (2013). Metzger et al. (2013) coincidentally used ERF and observational data to derive these grids in their independent study, though any other appropriate dataset or methodology could have been used to provide the LES surface forcings. The

Metzger et al. (2013) grids were merely a suitable dataset accessible to us at the time, irrespective of methodology.

Proposed changes to the manuscript:

- Reverse the sequence of Figure 3a and Figure 3b to follow the actual sequence of LES-ERF data processing (first LES, then ERF).
- Remove any notion of ERF when citing the use of the Metzger et al. (2013) sensible and latent heat flux grids, to forego conflation with LES-ERF.
- In Sect. 2 Materials and methods add a paragraph that links the introduction to a "big picture" overview of the subsequent methods subsections. Distinguish between (i) the novel NS-ERF approach to OSSEs and (ii) its maiden, example application to the CHEESEHEAD19 case study in the form of LES-ERF.
- In the beginning of Sect. 2.2, add a flowchart of the generalized NS-ERF approach that acts as a "visual glossary" by inter-relating individual components/terminology (e.g., optimality criterion, hypothesis, objective, observing system design, scientific return etc.). Explain how these generalized components are then specifically realized in the CHEESEHEAD19 case study in the form of LES-ERF.
- Explain that the numerator in Eq. 1 does not vary and that the denominator varies according to the different OSDs.

After reading the results (L. 440) I now see that there is a baseline OSD for the tower-only dataset and 12 other OSDs for the combined tower-aircraft. The tower-only dataset has a given spatial coverage that can be improved by aircraft sampling, and the goal is to maximize this coverage by deciding on certain flight tracks. By applying different OSDs with different flight angles, it was found that the spatial coverage is maximized in a perpendicular flight track (25% improvement). In table 1, similar analyses are given for energy balance ratio, and spatial patterning, the other two optimality criteria.

Author reply: Many thanks for this excellent summary! These relationships are also introduced in Sect. 2.3 l. 269ff.: "*Based on this super-sample we evaluated 13 candidate OSDs. Applying LES-ERF to 44 site-hours of data from the virtual EC tower network alone provided the* ***baseline OSD****. Combining data from the virtual EC tower network with one of the 3 flight sequences × 4 flight patterns = 12 airborne EC combinations provided 12* ***alternative OSDs****. Each of the alternative OSDs consisted of 44 site-hours virtual tower EC data and 11 flight tracks × 25 km = 275 km virtual airborne EC data. This configuration allows us to* ***evaluate the change in the optimality criteria*** *(Sect. 2.2) for each of the 12 joint tower and aircraft* ***alternative OSDs relative to the tower-only baseline OSD****.*".

Proposed changes to the manuscript: Add the referee's "plain language summary" to above paragraph in Sect. 2.3.

Reading section 2.4, I see that the LES are created with surface fields of H and LE created from the ERF. So, isn't this some circular reasoning? You are creating OSDs from ERFs that are benchmarked against LES created using ERF… Please explain

Author reply: Please see our responses to above referee comment "*The methodology can be very confusing…*". We believe that the present comment originates from the same two ambiguities detailed in our response above.

Proposed changes to the manuscript: Please see our proposed changes in response to above referee comment "*The methodology can be very confusing…*".

I think the results are clear and easy to follow but as mentioned previously the methodology can be a little confusing and I think needs more synthesis and needs to be clearer.

Author reply: Please see our responses to above referee comment "*The methodology can be very confusing…*". We believe that the present comment originates from the same two ambiguities detailed in our response above.

Proposed changes to the manuscript: Please see our proposed changes in response to above referee comment "*The methodology can be very confusing…*".

It is my understanding that the LES_ERF approach is designed specifically for the combination of airborne EC measurements with tower EC measurements. If this is the case this needs to be stated clearly in the abstract and perhaps even in the title.

Author reply: Many thanks for sharing this observation! The presented approach to OSSEs is indeed widely applicable beyond combining airborne EC measurements with tower EC measurements. We feel that the impression of limited applicability likely originates from the current manuscript structure lacking separation between (i) the novel, general Numerical Simulation - Environmental Response Function (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 of NS-ERF. Furthermore, the manuscript does not currently specify which alternative form NS-ERF components could take for natural climate solutions, emission inventory validation, urban air quality, industry leak detection, and multi-species applications.

Proposed changes to 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.
- 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.
- Substantiate extensibility claim by adding concrete examples for case study substitution where use cases are mentioned throughout the manuscript. In the discussion Sect. 4.2, spell out how to extend to any sort of study where

measurements of X across space and time need to be combined to estimate some surface or atmospheric property Y using input covariates Z in ERF, as simulated in NS.

To me, it is not clear what the "scientific return" means. How can you double the scientific return? Isn't this all subjective? Whatever information you gained by using one flight path instead of the other depends on how you interpret it. Doesn't it? What the results show is that flight patterns with a track angle of 45-90 double the percent improvement based on three optimality criteria, when compared to parallel flight patterns. I'm also not sure how you can "order-of-magnitude improve flight operation and crew safety". The last two statements are part of the main conclusion but the way they are quantified seems subjective.

Author reply:

- We agree with the referee's sentiment that "*Whatever information you gained by using one flight path instead of the other depends on how you interpret it*".
- OSSEs in general aim to provide an objective interpretation of "information gain" based on quantifying the ability to address the science objectives of a given application. To systematically realize this aim for surface-atmosphere field measurements we created the NS-ERF approach to OSSEs, and provide a case study example (CHEESEHEAD19 airborne design in the form of LES-ERF).
- In Sects. 2.2 and 2.3 we introduce the relevant NS-ERF components and their 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 components, thus creating a lack of context for objectively interpreting the CHEESEHEAD19 case study results.
- Sect. 4.1 specifies: "*On the other hand, the parsimonious number of only **6 flight sequences** and an even smaller number of 3 sets of waypoints simplify flight planning and navigation. In combination with entirely avoiding the town and airfield of Park Falls this **promotes flight crew safety by an order of magnitude compared to the originally envisioned 48 flight sequences**. Specifically, it frees up the flight crew from arduous navigation details, thus reducing fatigue, increasing awareness during the 100 m low-level flight maneuvers, and ultimately reducing the margin for human error.*"

Proposed changes to the manuscript:

- Replace "scientific return" with "information gain" throughout the manuscript.
- Reword the title to "Novel approach to observing system simulation experiments improves information gain of surface-atmosphere field measurements".
- Reword Sect. 2.2 header to "NS-ERF observing system simulation experiments". Add a flowchart/"visual glossary" to the beginning of Sect. 2.2 for inter-relating 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 Sect. 5 Conclusion, replace *"order-of-magnitude improve flight operation and crew safety"* with "improve flight operation and crew safety by reducing the originally envisioned 48 flight sequences to a parsimonious number of only 6 flight sequences".

Lastly, a question of applicability. How many other researchers are in the capacity to apply airborne EC measurements with such large-scale deployment of towers and the capacity to run computationally expensive LES at this scale? The authors present a good analysis of other large-scale field campaigns in section 4.2 but still, I'm not sure about the use of this approach to support the last conclusion of the abstract that "the approach lends itself to optimize observing system designs also for natural climate solutions, emission inventory validation, urban air quality, industry leak detection, and multi-species applications" What would be the cost-benefit analysis of implementing a large-scale field campaign like this for every natural climate solution project?

Author reply:

- Many thanks for raising this point. The presented NS-ERF approach to OSSEs is applicable to field measurements in general, and large-scale deployments or even an airborne component are by no means a requirement. For additional detail please see our responses to the above referee comment "*It is my understanding that the LES_ERF approach is designed specifically for the combination of airborne EC measurements with tower EC measurements…*"
- So when designing a natural climate solution (or other) project, NS-ERF could be applied at that project scale, e.g. much smaller compared to CHEESEHEAD19. Furthermore, an alternate NS-ERF configuration could be considered such as replacing LES (from LES-ERF) with RANS (to RANS-ERF) as discussed in Sect. 4.3. Such modular adjustments and scalability reduce computational expense and makes the general NS-ERF approach to OSSEs accessible to a wide range of applications.
- The list of large-scale field campaigns in the manuscript are meant to provide context for the CHEESEHEAD19 airborne case study.

Proposed changes to the manuscript:

- Please see our proposed changes in response to above referee comment "*It is my understanding that the LES_ERF approach is designed specifically for the combination of airborne EC measurements with tower EC measurements…*"
- Specify that the NS-ERF approach to OSSEs is applicable to field measurements in general, and large-scale deployments or even an airborne component are by no means a requirement.

- Expand concrete examples for case study substitution with scalability considerations incl. project size, NS-ERF configuration options, etc.

Specific comments

31 This approach doubled the ability to explore energy balance closure? This is a very subjective statement. How do you double the ability to explore?

Author reply: Thanks for this pointer. Please see our responses to above referee comment "*To me, it is not clear what the "scientific return" means…*". We believe that the present comment originates from the same ambiguity detailed in our response above.

Proposed changes to the manuscript: Please see our proposed changes in response to above referee comment "*To me, it is not clear what the "scientific return" means…*". Furthermore, change "*…doubled CHEESEHEAD19's ability…*" to "doubled CHEESEHEAD19's potential" throughout the manuscript.

371. Please specify the subsection in Section 4 where we can learn why the PBL was so low

Author reply: Thanks for the pointer!

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

409 Should this be "near-surface moisture"? "Land surface moisture" sounds like soil moisture and is redundant.

Author reply: Thanks for this pointer.

Proposed changes to the manuscript: Change "land surface moisture (LSM)" to "near-surface moisture (NSM)" and "land surface temperature (LST)" to "near-surface temperature (NST)".

L.465 Indicate in the table that these are percent values

Author reply: Many thanks for your suggestion.

Proposed changes to the manuscript: Update Table 1 – Table 3 accordingly.

I find that in the introduction and methods there are multiple statements that are hard to follow, such as:

192 "Here, we propose the extensible LES-ERF approach that explicitly simulates the joint scientific return in response to different candidate OSDs for addressing user-defined design hypotheses"

I think part of the problem is the use of the term "scientific return", which should be reevaluated.

Author reply: Thanks for this pointer. Please see our responses to above referee comment "*To me, it is not clear what the "scientific return" means…*". We believe that the present comment originates from the same ambiguity detailed in our response above.

Proposed changes to the manuscript: Replace "scientific return" with "information gain" throughout the manuscript. For additional detail please also see our proposed changes in response to above referee comment "*To me, it is not clear what the "scientific return" means…*".

I also think that an effort to synthesize the introduction and the methods would help the readability of the paper.

Author reply: Agreed, many thanks for this observation!

Proposed changes to the manuscript: In Sect. 2 Materials and methods add a paragraph that links the introduction to a "big picture" overview of the subsequent methods subsections. Distinguish between (i) the novel NS-ERF approach to OSSEs and (ii) its maiden, example application to the CHEESEHEAD19 case study in the form of LES-ERF.

L.286 What is the LES time step?

Author reply: The LES time step was set at 0.4 s for the analysis period. Mentioned in line 303.

Technical comments

Should the verb be "creating" rather than "observing". Aren't you creating these OSSE?

Author reply: We were not able to locate this comment in the manuscript.

References

Metzger, S., Xu, K., Desai, A. R., Taylor, J. R., Kljun, N., Schneider, D., Kampe, T., and Fox, A.: Spatio-temporal rectification of tower-based eddy-covariance flux measurements for consistently informing process-based models, 46th AGU annual Fall Meeting, San Francisco, U.S.A., 9 - 13 December, 2013.