<u>Author's response to the comments from Editor:</u>

Dear Ad Stoffelen,

thank you for your comments regarding our revised manuscript. Below, please find our responses and changes that have been implemented into its latest version.

1) Comment from referee: The methodology appears to rely on conditional sampling, e.g., if x and y measure truth t, then the mean of y-x for given x is evaluated and interpreted as the bias (or SD) of y versus t. This is incorrect of x has an error. The error of x will broaden the distribution of x w.r.t. t and therefore extreme x are larger than extreme t, such that y should be biased against x, even though x and y have only random error w.r.t. t (See, e.g., Stoffelen, 1998); See, inter alia, Figure 2.
Author's response: Thank you for pointing this out. The occurrence of these artifical biases was already picked up by Referee III during the first iteration of the manuscript (P. 9, L.29-32). To assess their level of influence on the uncertainty analysis, the biases (for q_a shown in Fig. 2, but also for U and q_s) were illustrated as a function of the *in situ* source, exemplarily for 2001. Results indicate that in 80% of all match ups, relative difference between HOAPS and the in situ mean 5-percentiles (indicated by black squares, respectively) range between ± 6-10%, which we consider as negligible. Differences in the tail regimes (below/above the $10^{th}/90^{th}$ percentile) exceed 10%. A minimization of these pseudo biases can be achieved by considering the average of both approaches (that is, HOAPS and *in situ* as abscissa variables), which is envisaged for future HOAPS uncertainty characterizations. Moreover, defining bins on the basis of percentiles (as has been done, rather than equidistant bins) contributes to constraining the pseudo biases.
Despite the occurrence of pseudo biases, we would like to add:
i) the influence of the pseudo biases, specifically in the tail regimes, becomes smaller when investigating bin-wise biases in four-dimensional space, which is fundamental to our uncertainty characterization. Speaking graphically, pair-wise biases may fall into a *neighboring* or *nearby* bin (of 20^4 = 160000 bins in total) within the four-dimensional look up tables (LUTs). Not only do neighboring biases highly correlate; this “displacement“ can even average out, once millions of match ups have been assigned to unique bins within the LUTs.
ii) as stated in Sect. 2 (P.8, L.13-18), our uncertainty estimates should be interpreted as *upper-boundary* estimates. In the tail regions, pseudo biases therefore provoke artificial increases in the HOAPS uncertainty estimates. We believe that upper-limit uncertainty estimates are more confidential for the user community compared to lower-limit estimates concealing retrieval issues.
Changes in the manuscript: We now point out that an investigation of biases w.r.t. the *in situ* sources is envisaged for future HOAPS versions (P.10 , L.1-2). Additionally, we cite Stoffelen (1998) when mentioning that errors in buoys can lead to pseudo biases and therefore to an increase of the HOAPS uncertainty estimates (P.9 , L.32-33).

2) Comment from referee: It is assumed that errors in q, U and T are independent, which appears rather odd? In areas with high SST variability, high wind variability will occur, as well as variability in humidity and air temperature. Also, in moist convection, all atmospheric parameters tend to vary simultaneously and in correlated ways. Besides correlations expected from physical processes, correlations are also expected in the simultaneous retrievals. A retrieval is an algorithm where radiance measurements are compromised in order to obtain a geophysical retrieval. Ergo, an error in U in the retrieval is likely associated with compensating errors in the other retrieved geophysical variables. Furthermore, if the same QC, cal/val and retrieval algorithms are used for two different instruments, error correlation of those multi-variable retrievals is likely too. This should be made clear.
Author's response: It is not clear to us, to which part of the manuscript the comments regarding the independency of errors belong to. We believe they point at the random uncertainty decomposition using triple collocation analysis. In this case, the reader is referred to Kinzel et al. (2016). Their Sect. 2c (P. 1460) throughly describes why we assume the individual uncertainty components contributing to the error models to be independent of the satellite platform. Moreover, an error correlation term explicitly contributes to the variances of differences (their Eq. 1), which is non-negligible when the individual

error terms are not independent. This error correlation term is not neglected in Kinzel et al. (2016), but is explicitly accounted for. We certainly agree that an error in e.g. U in the retrieval is likely associated with compensating errors in other geophysical retrievals. For example, parts of the random uncertainties shown in Fig. 2a (error bars) receive a systematic component in Fig. 2b (squares) (P.10, L.8-10). In fact, this motivated us to characterize the LHF-related uncertainties using a *multi-dimensional* approach, where we explicitly separate systematic from random uncertainties.
Changes in the manuscript: Kinzel et al. (2016) is cited in context of the error models (P.11 , L.28). Additionally, we mention that error correlation terms are explicitly accounted for when decomposing the random uncertainties (P.11, L.30).

3) Comment from referee: The error model is not clear formally. What is assumed to be truth and what instrumental and geophysical variations are exactly captured by the error variables?
Author's response: We assume that the comment targets the random uncertainty decomposition approach summarized in Sect. 3.3. As noted in the manuscript (P.11, L.18-19), technical details regarding the uncertainty decomposition are provided in Kinzel et al. (2016). Specifically regarding the error models, their Eqs. 2a-2b provide more insights. In case of *in situ* data, the only random uncertainty source is related to instrument noise (E_{ins}). Regarding the satellite, both random model uncertainties (E_M) and sensor noise (E_N) contribute to the random uncertainty component. When collocating, random collocation uncertainties (E_C) come into play. The first three listed uncertainty sources are related to the instruments, while E_C is a function of the geophysical parameter and the ambient conditions. The variances of differences are then derived bin-wise, once a bias correction w.r.t. SWA-ICOADS has been performed. We believe it is reasonable to consider the *in situ* data as our bias free ground reference (P.8, L.13-18), once sensor height corrections (in case of U) and cool skin effects (in case SST) have been carried out. Again, this assumption implicates that our LHF-related uncertainty estimates should be treated as upper-limit estimates, which is already picked up towards the end of Sect. 2 (P. 8, L.15-16).
Changes in the manuscript: We cite Kinzel et al. (2016), specifically when mentioning the error model (P.11, L.28) (see comment 2) above).

4) Comment from referee: The manuscript does not appear to clearly separate statistical and geophysical effects. Of course, statistical results may be linked to physical processes, but also to coincidental occurrence of high humidity and low winds, for example of conditional binning artifacts.
Author's response: We are unsure as to what section the comment refers to in the manuscript and therefore provide a general answer. We agree that no clear separation as to the type of effects has been done. Our multi-dimensional bias analyses result in bin-dependent biases (systematic uncertainties) and their spread (random uncertainties). The assignment of a single bias (e.g. q_a (HOAPS) minus q_a (*in situ*)) to one of the 20^4 bins depends on physical parameter dependencies. Therefore, the spanning of the four axes is geophysically motivated. The actual binning however, which results in bin-dependent biases and their spreads, is a statistical approach, as we do not explicitly solve for the parameter functionalities. We would like to point out that the scope of this manuscript lies on a general uncertainty characterization of HOAPS LHF-related parameters, irrespective of which effects cause them.
Changes in the manuscript: A note has been added when introducing the multi-dimensional bias approach that our motivation is geophysical, whereas the actual implementation is statistical (P.11, L.2).

<u>**Cited studies:**</u>

Kinzel, J., Fennig, K., Schröder, M., Andersson, A., Bumke, K., and Hollmann, R.: Decomposition of Random Errors Inherent to HOAPS-3.2 Near-Surface Humidity Estimates Using Multiple Triple Collocation Analysis, J. Atmos. Oceanic Technol., 33, 1455–1471, doi: 10.1175/JTECH-D-15-0122.1, 2016.

Stoffelen, A.: Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. J. Geophys. Res., 103 (C4), 7755–7766, doi: https://doi.org/10.1029/97JC03180, 1998.