Author's response to the general comments from referee I:

Thank you for revealing your valuable criticism regarding the manuscript. Below, please find our responses to your specific comments, along with the implemented changes to our manuscript. All page and line numbers as well as figure numbering refer to the *revised* manuscript. Note specifically that the figure numbering has changed during the review process.

MAIN COMMENTS:

i) Main comment from referee: *However, we are interested in whether the estimated uncertainty is common in (satellite) products or inherent in HOAPS-3.3. If the present results are inherent in HOAPS-3.3, the results are useful for only people to use HOAPS-3.3. However, if the results are common in most satellite products, the value of this article is considerably larger. If possible, we would like to know uncertainties about other products in order to judge whether the estimated uncertainty for HOAPS-3.3 in this study is common or not. I guess it is not so easy for the authors to estimate uncertainties for other products. If so, I would like the authors to investigate the relation between the uncertainties of HOAPS-3.3 obtained by this study and the differences between HOAPS and other products, pointed out by previous paper (Iwasaki et al. (2014)).*

Author's reponse: We chose to publish an AMT paper, as our manuscript describes a *technique* for assigning uncertainties to latent heat flux (LHF)-related satellite data. We do not aim at performing an uncertainty assessment of all available data records. Instead, as mentioned in the title, uncertainties are given for HOAPS, which has more than 200 users. We therefore agree that some of our findings cannot be generalized. As is discussed, our displayed uncertainties are in parts related to retrieval uncertainties and sensor noises, which are unique to every data set and satellite instrument, respectively, and are therefore not applicable to other satellite climatologies. We are not aware of any air-sea flux related remotely sensed data set to date that is equipped with instantaneous uncertainty estimates. HOAPS-3.3 therefore leads the way towards a more transparent satellite data analysis, as the user may individually decide how to treat the data, given the available retrieval uncertainties.

More important, we want to highlight the fact that our approach can easily be applied to other satellite data sets, as long as a sufficiently large amount of collocations can be achieved. Choosing a similar in situ data basis and identical collocation criteria compared to our manuscript, random in situ (here: E_{ins}) and collocation uncertainties (here: E_c) are thought to be comparable to our results, independent of the investigated satellite climatology. As you state, this considerably increases the value of this article.

The uncertainty estimates cannot be set into relation with other satellite climatologies, as no further uncertainty values exist for comparison. However, we agree that the research community would benefit from investigations answering the questions „Do other LHF-related data sets lie within the uncertainty range specified by HOAPS-3.3? If not, how can we explain this discrepancy?". As noted in Sect. 4.7 of the present manuscript, we are currently preparing a follow-up publication regarding this aspect. It will present our findings in a larger perspective and thus increase the importance of our uncertainty analysis.

To increase the value of the present manuscript, we have established links to E-P intercomparisons illustrated by Iwasaki et al. (2014) whenever it fits the context (see „changes" below). One must keep in mind, however, that HOAPS-3 (as is used in Iwasaki et al. (2014)) differs from HOAPS-3.3 used within the present work. Apart from a temporal extension by seven years, this includes changes to the calibration model of SSM/I brigthness temperatures (Fennig et al. (2013)), an updated version of the AVHRR Pathfinder Data Set (SST), and the inclusion of SSMIS data (Fennig et al. (2015)).

Iwasaki et al. (2014) is valuable when it comes to intercomparing various freshwater flux products over the global ice-free oceans. It identifies individual parameter contributions to the overall observed differences and allows for assessing which parameters contribute most to the positive

trend in E. Yet, such an intercomparison does not allow for drawing conclusions regarding the uncertainty of the individual data sets. Observed differences between two data sets could either diminish or amplify when applied to the respective climate data set. In this regard, the present manuscript is very progressive, as it sets a basis for assigning uncertainty measures to climate data records. For example, our uncertainty estimates allow for concluding whether the illustrated differences were to be expected or not. Large differences, coupled to small HOAPS-3.3 uncertainty estimates, would point at retrieval issues related to the data set compared to.
Changes in the manuscript: Iwasaki et al. (2014) is cited for the first time in Sect. 1 (P.3, L.27). In the following places of Sect. 4, it is picked up again, where is relates HOAPS to other LHF climatologies: P.17, L.5/25; P.20, L.2; P.21, L.24/26.
Sect. 1 (P.4, L.29) now includes a sentence, which emphasize the fact that the methodology may easily be transferred to other retrievals, which increases the value of our manuscript. This is revisited in Sect. 5 (P.20, L.27) and also implemented in the abstract (P.1, L.17).

ii) Main comment from referee: *This article is based on Kinzel et al. (2016). However, the paper is not referred in the present introduction. It is curious. The purpose of this article is not so clear for me. I think the purpose of this study is comprehensive estimation of uncertainty characterization of HOAPS-3.3 latent heat flux (LHF) related parameters in addition to specific humidity examined in Kinzel et al. (2016).*
Author's reponse: We agree that the introduction benefits from citing Kinzel et al. (2016), as it introduces the concept of random uncertainty decomposition, which is perfomed within the present study. The approach presented in Kinzel et al. (2016) should be understood as one of several *prerequisites* for our work, as it a) (only) focuses on random uncertainties and b) does not cover wind speed (U), LHF, and evaporation (E). We will provide a citation in an appropriate place und put Kinzel et al. (2016) into a larger context.
Changes in the manuscript: Kinzel et al. (2016) is now referenced in Sect. 1 (P.4, L.10/21), where it is also put into a larger perspective.

iii) Main comment from referee: *For example, the authors attribute the global minimum during boreal summer 1991 to the Mount Pinatubo eruption. However, we cannot find the minimum in 1991 in other products except HOAPS (Iwasaki et al. (2014, their Fig.6a). Therefore, the minimum may be due to the HOAPS retrieval error related to the Mount Pinatubo eruption.*
Author's reponse: Regarding the 1991 minimum related to the Mount Pinatubo eruption: we agree on this. Please refer to the specific comment #20 further below for more details on this. The explanation for the SST feature seen in HOAPS LHF during 1991 was already implemented in the submitted version (see P.18 ,L.25f of revised manuscript).

iv) Main comment from referee: *Also, since all HOAPS parameters are derived from SSM/I and SSMIS microwave radiometers, the sampling errors are expected to be large compared with other products using many kinds of microwave radiometers.*
Author's reponse: We agree that differences in sampling between different instruments exist, which may cause sampling biases. However, it should be kept in mind that the manuscript demonstrates an application of the introduced methodologies and does not focus on an assessment or intercomparison of sampling uncertainties. The SSM/I and SSMIS sampling uncertainties are accounted for, which play a marginal role on climatological time scales. This is mirrored in the small magnitudes of monthly mean E_{smp} in Table 2 of the present manuscript.

v) Main comment from referee: *Although the second paragraph in the section 5 introduces HOAPS 4.0, I feel the paragraph is not necessary in this section.*
Author's reponse: We believe it is important to note that the newest version of HOAPS, that is HOAPS 4.0 (released in October 2017), includes an update of the uncertainty estimates. Apart from this, we outline new features and improvements with respect to HOAPS-3.3 in two sentences. We

agree that it is somewhat out of place in the submitted manuscript. This short paragraph is therefore placed towards the end of Sect. 5.
Changes in the manuscript: The short paragraph related to HOAPS 4.0 has been moved to the end of Sect. 5 (P.22, L.6ff).

vi) Main comment from referee: *Moreover, the authors discuss about precipitation in this section, but I think this issue may exceed the scope of this study because they do not carry out uncertainty estimates of HOAPS precipitation here.*
Author's reponse: We generally agree with this comment and will therefore remove parts of the provided literature review on issues with satellite precipitation (P) estimates. However, we want to continue emphasizing the importance of quantifying P uncertainties, because it ultimately allows for assessing uncertainties in freshwater budgets (E-P). In this context, the mentioned study by Burdanowitz et al. (2016) is valuable, as it lays the basis for this purpose.
Changes in the manuscript: The paragraph related to uncertainties in P has been shortened (P.21, L.30ff).

SPECIFIC COMMENTS:

1)Comment from referee: P.1, L.1: "of LHF" → "of in situ LHF"
Author's reponse: We agree that 'in situ' should be added in this context
Changes in the manuscript: 'in situ' has been added to the revised manuscript (P.1, L.1).

2) Comment from referee: *P.3, L.21-27: In this paragraph, we need clear description about characteristics related to uncertainties, of HOAPS LHF product compared with other products obtained by numerous intercomparison studies*
Author's reponse: The mentioned paragraph serves to merely introduce the HOAPS climatology. Apart from listing included parameters, the brief literature review on HOAPS intends to demonstrate its usefulness in climate research and highlight its performance in context of intercomparison studies. For further information, the reader is referred to the quoted references. We believe that a thorough description related to uncertainty characteristics exceeds the scope of this introductory paragraph. However, we agree that highlighting some distinct differences among the data sets (without a focus on uncertainty estimates) would improve our introduction.
Changes in the manuscript: A paragraph has been added to the revised manuscript (P.3, L.25ff), which points at substantial differences between LHF data sets (including HOAPS) on a local scale. A second paragraph deals with performed uncertainty characterizations related to LHF (P.4, L.4ff). It shows what has been done to date and points at the shortcoming that, apart from NOCS v2.0, no uncertainty estimates are available to the users.

3) Comment from referee: *P.5,L.16-21: Large El Nino and La Nina occurred in 1997-1998. Therefore, 1997-1998 is a special period. Why did the authors use the data in this period?*
Author's reponse: We agree that 1997-1998 were "special" years, in a climatological sense. We argue that for training purposes, it is not essential whether the contributing data was obtained during climatologically anomalous years or not. What counts is that a) the network is trained with match ups, which are physically connected and b) the whole possible range of atmospheric conditions (i.e., in this case wind speeds) is covered by a representative amount of data. It that sense, match-ups from 1997-1998 are beneficial, as they guarantee a full coverage of all conditions. Thus, potential extremes are covered in our training data base.

4) Comment from referee: *P.5,L.33-34: The assumption of a constant relative humidity of 80 % and air-sea temperature difference of 1 K is considerably artificial. To what extent does have the assumption impact on estimation of uncertainty?*

Author's reponse: Thank you for bringing this up. We did not investigate the uncertainty introduced by these two widely used assumptions, as it may be neglected for two reasons. First, air temperature only has a secondary effect on LHF (in contrast to SHF) through the stability of the atmospheric column. At the same time, the assumption of 1 K temperature difference with respect to SST is a good approximation for vast regions over the global oceans. However, we agree that over upwelling regimes, which are very confined compared to the global oceanic area, this approximation is violated. Compare conclusion section of Wells and King-Hele (1990). Second, our uncertainty estimation procedure described in Sect. 3 is exclusively based on high-quality match-ups of HOAPS and in situ measurements. The data density of both ship and buoy records is comparably low in the upwelling regimes, which further reduces the impact of our two assumptions. Due to the comparatively small amount of reference data, we presumably underestimate resulting uncertainties in these regions. Using for example ancillary reanalysis-based data would violate our ambition to create a completely remotely-sensed data record, which is a key feature of HOAPS.

5) Comment from referee: *P.6,L.25: (2003) ---(2013)*
Author's reponse: We agree that Bentamy et al. (2013) is worth citing here.
Changes in the manuscript: A citation of Bentamy et al. (2013) has been added to revised manuscript (P.7, L.3).

6) Comment from referee: *P.8,L.15: In what ways are these features similar?*
Author's reponse: The term 'similarity' refers to the similarity of the bias distributions as a function of the x-axis parameters. That is, lowest SST (i.e., high latitudinal SST) are underestimated in HOAPS (likewise, q_a is underestimated for (high latitudinal) q_a below 5 g kg^{-1}). The HOAPS underestimation also accounts for subtropical SST in the range of 25°-29°C (likewise, q_a is underesimated for q_a between 15-19 g kg^{-1}). By contrast, HOAPS SST are slightly overestimated for SST ranging between approximately 15°-24°C and the inner tropics (30°C). Likewise, q_a is overestimated for q_a between 7-12 g kg^{-1} and for inner tropical 20 g kg^{-1}).
Changes in the manuscript: The wording has been modified in the revised manuscript (P.9, L.14f).

7) Comment from referee: *P.8,L.22: " off the Arabian Peninsula". We cannot recognize the data off the Arabian Peninsula in Fig. 1. We need the distribution of average q_a for this.*
Author's reponse: Indeed, thank you for pointing this out. This paragraph is meant to exemplarily present the benefit of multi dimensionality, whereas the illustration of q_a patterns referred to is not the primary focus. We therefore omitted an additional map showing the distribution of q_a and U over the Arabian Sea. However, we have included a global map showing the average difference between HOAPS and in situ q_a
Changes in the manuscript: 'not shown' has been added twice to the revised manscript (P.9, L.24f).

8) Comment from referee: *P.9 L.11: Is the bin width equal or not? How did you determine the bin width?*
Author's reponse: The bin width is not equidistant. It is rather determined by fixed percentiles of data, where 5% of all contributing match-ups are assigned to a single bin. In consequence, 20 bins result, which are narrow for large data densities and become wide close to the tails of the distribution. This is also picked up it the caption of Fig. 2.
Changes in the manuscript: A note on the bin configuration has been added to the revised manscript (P.9, L.3ff) and is again picked up in context of Sect. 3.2 (P.10, L.16) and the caption of Fig. 2.

9) Comment from referee: *P.9, L.17: Why did you choose the different data period between (dq_a, dU) and (dq_s)?*

Author's reponse: For dq_a and dU, the vast amount of in situ data justified the restriction to collocations between 2000 and 2008. For d_{qs}, the time period from 2002-2005 was left out, as corresponding local equatorial overpasses of the operating NOAA-17 were disadvantageous for our double collocation analysis. Recall that only *night-time* SST were collocated to in situ measurements to avoid the warm layer effect (see Sect. 2.2 of this manuscript). Fulfilling the requirement of local night time, the overpass times of NOAA-17 were inappropriate for gathering a large number of collocations. Instead, collocation during 2006-2008 were used. Additionally, the period 1998-2001 was taken as reference to allow for a sufficiently large collocation data basis. In consequence, dq_s match ups are based on collocations from 7 years only. This does not pose a problem, as in situ SST measurements were available more frequently compared to U and q_a.

10) Comment from referee: *P. 10, L.4: The average of daily coefficients is applied for estimation of instantaneous LHF uncertainties here. Why are not instantaneous values but daily values applied? Also, is the difference between daily and instantaneous coefficients small or large?*
Author's reponse: Thank you for bringing this up. We had similar thoughts regarding the representativity of daily versus instantaneous correlations. Deriving instantaneous correlation coefficients, however, has a key disadvantage. Most of the global ocean is scanned only 1-2 times per day by a single SSM/I or SSMIS instrument, some regions over the subtropics not at all. This implies that the amount of instantaneously derived geophysical parameters is locally very limited. Resulting correlation coefficients would therefore not be representative. We therefore decided to apply global averaged coefficients, which are remarkably stable throughout the year on a day-to-day basis (not shown). Due to this decision we are not capable of comparing our coefficients to instantaneous correlation coefficients. We are aware that differences may occur.
However, we furthermore investigated, how much the sum of all correlation terms in Eq. (2) contributes to instantaneous $\sigma_{LHF,sys}$. On average, omitting these correlation terms modifies the resulting instantaneous $\sigma_{LHF,sys}$ by merely 0.5 ± 5 W m^{-2}. Thus, even if global mean correlation coefficients were not always the most accurate choice, they do not represent a key contribution to resulting LHF uncertainty estimates.
Changes in the manuscript: The two reasons for why we apply the average of daily mean global correlation coefficients have been included into the revised manuscript (P.11, L.31ff).

11) Comment from referee: *P.10,L.8: Could you explain about the definition of "gridded uncertainty products"?*
Author's reponse: Sorry for not being precise here. By „gridded products", we general mean satellite data that has been spatially and temporally averaged and that is available for fixed grid cells (dx,dy) and time periods (dt), like 'HOAPS-C' and 'HOAPS-G'. This stands in contrast to instantaneous, level-2 data (points in time and space, like 'HOAPS-S'), which form the basis of our uncertainty analysis. To avoid confusion, we will not mention this in the revised manuscript and rather rephrase this sentence.
Changes in the manuscript: The wording has been modified in the revised manuscript (P.12, L.8f)

12) Comment from referee: *P.10,L.17: What is a true value for E_c?*
Author's reponse: We do not understand the question. Please see Table 1 of our manuscript for magnitudes of HOAPS LHF-related E_c resulting from the random uncertainty decomposition.

13) Comment from referee: *P.11,L.19-22: Here, all daily sampling uncertainties are derived as a function of the number. However, sampling error for a daily-mean value depends on not only the number but also observation times.*
Author's reponse: This is absolutely correct. Assuming a specific number of daily overpasses was a prerequisite for showing the sampling uncertainties as a function of operating satellites (Table 2 of our manuscript). P.11, L.18 of the submitted manuscript indicates that the daily sampling uncertainties are estimated using "simultated satellite records", which are derived using the two

buoy records closest in time to local satellite overpasses. The assumption of having two overpasses per day is reasonable, as this applies to vast regions of the global oceans. We assume that sampling uncertainties are inverse proportional to the amount of daily overpasses, but do not investigate this dependency further. As the number of daily overpasses increases with an increasing number of satellites, we rather resolve the resulting sampling uncertainties as a function of orbiting platforms. This is in line with conclusions by Tomita and Kubota (2011), who found that multi-satellite simulations for e.g. q_a considerably reduced the sampling uncertainty, compared to single satellite simulations.
Changes in the manuscript: The wording has been modified in the revised manuscript (P.12, L.18f) to point out that our estimates are based on the assumption of having two overpasses per day.

14) Comment from referee: *P.12,L.1-3: We find several geographical words such as " Arctic", " polar" and " inner tropics". However, it is difficult for us to obtain the relation between the ranges of the random satellite retrieval uncertainty and the geographical location from Fig. 1 and Table 1. Also are the values shown in this paragraph consistent with those in Table 1? For example, " 0.3 and 1.8g kg^{-1}" is " 0.7 and 1.8g kg^{-1}" in line 1?*
Author's reponse: Thank you for pointing this out. We agree that this is confusing and will clarify this in the revised manuscript, as Table 1 does not show distributions of the random retrieval uncertainty as a function latitude and longitude. Regarding the consistency of values shown in Fig. 2 and Table 1: Note that directly comparing results of Table 1 to Fig. 2 (and expecting equality) is not correct. Fig. 2 shows bin-wise biases and their spread in *one-dimensional* space. The values in Table 1, however, result from the *multi-dimensional* bias analysis, multiple triple collocation (MTC) analysis, and subsequent random uncertainty decomposition. This implies that random retrieval uncertainties of q_a presented in Table 1 are compatible with the global distributions shown in Fig. 3a. Regarding 0.3 g kg^{-1} vs. 0.7 g kg^{-1}: we apologize for this mistake, '0.3' is a typo and has been corrected to 0.7 g/g kg^{-1} in the revised manuscript.
Changes in the manuscript: The geographical terms have been removed and have been replaced with q_a magnitudes (P.13, L.10ff). The typo has been corrected (P.13,L.10). The captions of Table 1 and Fig. 3 have been modified to point at the similarity of both representations (i.e., showing $E_{retr,ran}$).

15) Comment from referee: *P.12, L.1-28: Accuracy of in situ data is considerably different depending on used sensors. For example, the accuracy of wind speeds is 1.0m/s or 10% for usual NDBC buoys, while that is 0.3 m/s for TOA buoys. Are these differences between them negligible for the present analysis?*
Author's reponse: Thank you for providing this differentiation regarding accuracies of buoy measurements. Sect. 4.1 deals with the random uncertainty component (that is, precision) and does not target accuracies. However, we generally agree that different instruments are associated with a variety of (random) measurement uncertainties. Sect. 4.1 (and thus Table 1) results from a random uncertainty decomposition procedure (compare Kinzel et al. (2016)), which crucially depends on the amount of contributing triple collocations and thus in situ measurements. Our collocation data basis is very large, including a variety of exclusively high-quality in situ measurements. The results of the decomposition itself should be interpreted in a way, such that *average* random insitu measurement errors can be separated from *average* random retrieval and collocation uncertainties, depending on the magnitude of q_a, U, and q_s. See for example the orange, red, and black squares as a funtion of q_a in Fig. 2 of Kinzel et al. (2016) for an illustration of this decomposition. Each of these orange squares can be understood as a *bin-averaged* random in situ uncertainty contribution. Thousands of in situ data records contribute to each of these squares/bins. One needs to therefore consider our random in situ uncertainties as an *average* over all in situ data sources for a specific parameter regime, i.e., bin. Therefore, individual in situ accuracies do not receive much weight.
Changes in the manuscript: Sect. 3.3 has been extended by two sentences (P.11, L.15ff), which emphasize that the random uncertainty magnitudes illustrated in Table 1 are derived bin-wise and

result from thousands of triple collocated match ups (and thus in situ records).

16) Comment from referee: *P.14, L.13: Could you tell me the definition of the climatological total uncertainties (E_{clim})? Are the climatological total uncertainties (E_{clim}) different from the systematic uncertainty?*
Author's reponse: Sorry for not being precise enough here. For each grid box of Fig. 3, we define the climatological uncertainty (E_{clim}) as the mean root mean squared sum of $E_{sys,}$ $E_{retr,ran,}$ and E_{smp} (1988-2012). As $E_{retr,ran}$ scales with 1/N, with N being the amount of observations per grid box, it becomes virtually zero for the temporal averages shown in Fig. 4 of our manuscript. Likewise, monthly mean E_{smp} are small (see Table 2 of our manuscript). Thus, on climatological timescales, E_{clim} and E_{sys} do not differ. We will emphasize the definition of E_{clim} more clearly in the revised manuscript.
Changes in the manuscript: A sentence has been added to Sect. 4.3 (P.15,L.12ff), which explains our definition of E_{clim}. It has also been added to the caption of Fig. 4.

17) Comment from referee: *P.16,L.28: What is the meaning of "isolated time periods"?*
Author's reponse: *We apologize for not being precise here. This was to state that during individual months ('isolated time periods'), the global mean uncertainty (one value) deviates from the respective average of 1988-2012.*
Changes in the manuscript: The wording has been modified in the revised manuscript (P.18, L.9f).

18) Comment from referee I: *P.17, L.3-19: E_{clim} is considered to be only one value from the meaning of a climatological value. Is it right? If so, I cannot understand the meaning of " respective E_{clim} over the Pacific upwelling regimes reaches 25 W m^{-2} specifically during boreal spring 1998" found in line 6-7.*
Author's reponse: E_{clim} is defined separately for each grid box (see comment #16 above on this), which is why we are explicitly able to e.g. point at climatological uncertainties over the Pacific upwelling regime.
Changes in the manuscript: See comment #16 above.

19) Comment from referee: *P.17, L.28: "climatological regional wind speeds range between 4.5.-11 m s^{-1} (fig.4b). As for q_a" --> "climatological regional uncertainties in wind speeds range between 4.5.-11 m s^{-1}(fig.4b). As for U"*
Author's reponse: We are not sure whether we understand this comment correctly. As formulated, the range of 4.5-11 m s^{-1} consideres the regional wind speed *itself*, not its related uncertainties. Fig. 4b shows regional and global mean HOAPS U, along with systematic and random retrieval uncertainties. The individual medians range between 4.5-11 m s^{-1}. Seasonality is most pronounced over the Indian monsoon region, WBC, and the North Atlantic (see JJA and DJM in Fig. 4b for this). Similar conclusions can be drawn for q_a (Fig. 4a), regarding maxima in seasonality for those three regions.

20) Comment from referee: *P.18, L.10: The global minimum during boreal summer 1991 is linked to the Mount Pinatubo eruptions. However, the remarkable minimum can be found in only HOAPS product and cannot be found in other products as shown in Fig. 6(a) of Iwasaki et al. (2014). Therefore, the minimum would be related to retrieval model uncertainty. The present analysis can investigate this issue and present its effectiveness by the investigation.*
Author's reponse: Thank you very much for pointing at the valuable study by Iwasaki et al. (2014), which we missed to cite so far. Indeed, the global minimum is linked to the Mount Pinatubo eruption and is not observed in the remaining satellite and reanalysis products. Similar to our work, the authors point at the cause of this low bias, which is attributed to AVHRR aerosol issues. In consequence, this created low-biased SST (i.e., low-biased q_s), which in turm resulted in

unrealistically low near-surface humidity gradients and thus low-biased E. This has already been picked up in e.g. Andersson et al. (2010) and is therefore a known issue related to the retrieval model. The recently released HOAPS 4.0 climatology (Andersson et al., 2017) does not include this feature anymore, as the SST reference has changed to the NOAA 0.25° daily Optimum Interpolation Sea Surface Temperature (OISST, Reynolds et al. (2007)), which corrects for this effect (see Reynolds, 1993). We are not aware of further systematic retrieval issues and the overall good performance of HOAPS in relation to other satellite and reanalysis data sets is mirrored in e.g. Iwasaki et al. (2014) (e.g. their Fig. 3). Regarding the classification of the low-biased LHF during 1991 (see Fig. 5 of our manuscript) with respect to the given uncertainty ranges: the low-biased LHF lies within the average HOAPS LHF retrieval uncertainty range (gray shading) between 1988-1998.
Changes in the manuscript: The explanation for the SST feature seen in HOAPS LHF during 1991 was already implemented in the submitted manuscript (P.18, L.25ff of revised manuscript). Furthermore, Iwasaki et al. (2014) has been included to the reference list and is cited where appropriate throughout the revised manuscript (see general comment #1 at the top of this document for more details).

21) Comment from referee: *P.18,L.15: As mentioned before, could you please explain about definition of climatological uncertainty? I cannot catch the meaning of " the 12-month running mean climatological uncertainty". Is a climatological uncertainty defined each month?*
Author's reponse: See comment #16 and #18 regarding the definition of E_{clim}. From these grid point wise E_{clim}, a global mean climatological uncertainty is derived for each month. This implies that twelve values result for each year. For smoothing purposes, an annual (that is , 12-month) running mean is performed over these 25x12 = 300 global monthly mean values.
Changes in the manuscript: The wording which describes Fig. 6 has been modified in the revised manuscript (P.18, L.29ff). Keeping the definition of E_{clim} in mind (see comments #16 and #18), it becomes clear that a global mean value of E_{clim} can be calculated for each month, to which running means can be performed. Furthermore, the caption of Fig. 6 has been slightly adjusted.

22) Comment from referee: *P.18, L.21-P.19,L. 5: In this paragraph, the results by many previous studies are introduced. However, the relation between the results and what Fig.5 shows is not so clear. I wonder this paragraph is necessary.*
Author's reponse: We agree that the focus of our manuscript lies on the uncertainty characterization, rather than on the positive trend seen in LHF.
Changes in the manuscript: The respective paragraph has been shortened (P.20, L.1-6).

23) Comment from referee: *Fig 2. (c) and Fig. 3. (c) It is difficult to know the distribution pattern in these figures. How about the change of a color bar?*
Author's reponse: These colorbars were chosen in order to be identical to the colorbars of Fig. 3a and 4a, respectively. Doing this, one can directly see the comparatively small uncertainty contributions of q_s in relation to q_a. Specifically regarding Fig. 3c, the distribution may not always be distinct. However, the most important feature in Fig. 3c, that is the maximum over the Indo-Pacific warm pool region, is well resolved. Pattern descriptions are additionally given for Fig. 3c (P.14, L.30-34) and Fig. 4c (P.16, L.20-25)
Changes in the manuscript: A comment regarding the same color bar range of q_a and q_s has been included in the figure captions of Fig. 3 and 4.

Cited studies:

Andersson, **A.**, Graw, K., Schröder, M., Fennig, K., Liman, J., Bakan, S., Hollmann, R., Klepp, C.: Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data - HOAPS 4.0, Satellite Application Facility on Climate Monitoring, doi: 10.5676/EUM_SAF_CM/HOAPS/V002, 2017

Fennig, K., Andersson, A., and Schröder, M.: Fundamental climate data record of SSM/I brightness temperatures, doi:10.5676/EUM_SAF_CM/FCDR_SSMI/V001, 2013.

Fennig, K., Andersson, A., and Schröder, M.: Fundamenal Climate Data Record of SSM/I / SSMIS Brightness Temperatures, CM SAF, doi: 10.5676/EUM_SAF_CM/FCDR_MWI/V002, 2015

Iwasaki, S., Kubota, M., and Watabe, T.: Assessment of various global freshwater flux products for the global ice-free oceans, Remote Sensing of Environment, 140, 549-561, doi: 10.1016/j.rse.2013.09.026, 2014.

Reynolds, **R. W.**: Impact of Mount Pinatubo Aerosols on Satellite-derived Sea Surface Temperatures, Journal of Climate, 6, 768-774, doi: 10.1175/1520 0442(1993)006<0768:IOMPAO>2.0.CO;2, 1993.

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K., and Schlax, M. G.: Daily High-Resolution-Blended Analyses for Sea Surface Temperature, J. Climate, 20, 5473–5496, doi: 10.1175/2007JCLI1824.1, 2007.

Tomita, H. and Kubota, M. K.: Sampling error of daily mean surface wind speed and air specific humidity due to sun-snychronous satellite sampling and its reduction by multi-satellite sampling, International Journal of Remote Sensing, 32, 3389–3404, doi: 10.1080/01431161003749428, 2011.

Wells, N. And King-Hele, S.: Parameterization of tropical ocean heat flux, Quart. J. Roy. Meteor. Soc., 116, 1213–1224, doi: 10.1002/qj.49711649511, 1990