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

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

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.

SPECIFIC COMMENTS:

1) Comment from referee: *Page 5, Line 10, "... the pixel-level HOAPS-3.3 data in sensor resolution is used...". What are the spatial and temporal resolutions of the pixel-level HOAPS-3.3 data? Which nine sensors are used in the pixel-level HOAPS-3.3 climatology?*
Author's reponse: The spatial resolution of both data sources is channel-dependent. For SSM/I data (DMSP F08-F15), it varies from 69 km by 43 km (19 GHz channel) to 37 km by 28 km (37 GHz channel). Sampling frequencies take on a value of 25 km, corresponding to scan lines every few seconds. Regarding SSMIS (DMSP F16-F18): The spatial resolution varies from 74 km by 47 km (19 GHz channel) to 41 km by 31 km (37 GHz channel). As for SSM/I, sampling frequencies are given by 25 km. Overall, 9 different DMSP sensors contribute to HOAPS-3.3: F8, F10-F11, and F13-F18.
Changes in the manuscript: The DMSP satellite platforms have been included into the revised manuscript (P.5, L.10f). Furthermore, the spatial resolution has been implemented (P.5, L.11ff).

2) Comment from referee: *Page 5, Line 15: what is the temporal resolution of q_a retrievals? And at what height?*
Author's reponse: Unfortunately, no information is provided by Bentamy et al. (2003) as to the sensor heights of the (in situ) q_a retrievals. What is known is that their updated regression coefficients are derived using 1000 collocations between globally distributed ship data and validated DMSP satellite data (F10-F14) during 1996-1997. As the retrieval is based on these match ups, we believe that the expression of "temporal resolution" is somewhat misleading. The globally distributed match ups do not have a temporal resolution and are rather point measurements in time and space.

3) Comment from referee: *Page 5, Line 32: which surface pressure data are used in computing LHF?*
Author's reponse: The COARE-3.0 algorithm assumes a standard sea level pressure (SLP) of 1013.25 hPa when iteratively calculating LHF, which is also used for deriving HOAPS LHF. Brodeau et al. (2017) investigated the effects of this SLP approximation in bulk parameterizations of tubulent air-sea fluxes, amongst others. The authors conclude that errors of such an approximation remain well below discrepancies related to the computation of the transfer coefficients themselves. Their sensitivity experiments show that q_s- and ρ-induced errors range between merely ±5% (given an SLP range from 950 hPa to 1040 hPa) with an opposite and therefore potentially compensating effect on LHF. Apart from this, both SSM/I and SSMIS are not capable of deriving SLP. Making use of auxiliary (e.g. reanalysis) data to implement SLP would violate HOAPS' unique feature of relying completely on satellite input.
Changes in the manuscript: A note has been added to the revised manuscript (P.6, L.9) that a constant SLP is presumed.

4) Comment from referee: *Page 5, Line 33: "...surface air temperature, which is estimated by assuming a constant relative humidity of 80 % (Liu et al., 1994) and air-sea temperature difference of 1K". How accurate is this assumption? During winter cold air outbreaks over the western boundary current regions, the air-sea temperature differences can exceed 10 K. In this case, the*

assumption will lead to a bias in air temperature. How is surface air temperature compared to the in situ dataset?
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 (for our purposes).
First, air temperature only has a secondary effect on LHF (in contrast to SHF) through the stability of the atmospheric column. 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 during cold air outbreaks over the WBCs or in 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 regions addressed above, 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 removely-sensed data record, which is a key feature of HOAPS.
No SSM/I or SSMIS retrievals exist that are capable of accurately retrieving oceanic surface air temperature (SAT) from space. This implies that SAT is not available as an official HOAPS product and has thus not been compared to the in situ reference. Future efforts will take on this challenge.

5) Comment from referee: *Page 6, Line 1: Provide a map showing the spatial distribution of in situ (ship and buoy) reference data density over the global domain.*
Author's reponse: We agree that providing such a map is useful to the reader. We therefore implemented a map showing the spatial distribution of match ups (ship/buoy vs. satellite) over the global oceans, exemplarily for q_a. It shows all collocated match ups between 2001-2008 that contribute to Fig. 2 (≈ 13.8 million match ups per subplot in total). Match ups for U and q_s occur even more frequently, but are not shown in the revised manuscript.
Changes in the manuscript: A map showing the distribution of q_a collocations between 2001-2008 has been implemented into the revised manuscript (Fig 1, left panel). It is briefly described in terms of density distributions (P.8, L.29ff).

6) Comment from referee: *Page 6, Lines 4-5: Does the reference dataset include the 1996-97 period that is used in training q_a algorithm?*
Author's reponse: We are not able to answer this question, as Bentamy et al. (2003) does not provide any information as to which ship records were used to train their q_a retrieval. Yet, the multi-dimensional bias analyses are restricted to match ups between 1998 and 2008 (depending on the parameter, see P.10, L.25f). This implies that no temporal overlap between the reference data archive and the ship records used for training purposes exists.

7) Comment from referee: *Page 7, Lines 28-29: The "instantaneous and climatological uncertainties" are not explained. How are they related to systematic, random, and sampling uncertainties?*
Author's reponse: Sorry for not being precise enough here; we agree that this needs clarification. "Instantaneous" uncertainties are pixel-level uncertainties. These uncertainties can either be systematic (compare Fig. 4 over revised manuscript) or random (see Fig. 3 of revised manuscript). On an instantaneous basis, sampling uncertainties do not exist.
By contrast, we define "climatological" uncertainties as *total* uncertainties averaged over the time period 1988-2012 (as illustrated in Figs. 4 and 5 of revised manuscript). That is, E_{clim} is formally the mean root mean squared sum of E_{sys}, $E_{retr,ran}$, and E_{smp} averaged over 1988-2012. As $E_{retr,ran}$ scales with 1/N, with N being the amount of observations per grid box (see Eq. 3), it becomes virtually

zero when averaging over long time periods. Likewise, monthly mean E_{smp}, which applies even more so to multi-annual averages. On climatological time scales, E_{clim} and E_{sys} therefore hardly differ. This is why Fig. 4 of the revised manuscript can be treated as both „systematic“ and „climatological“ uncertainty.
Changes in the manuscript: The explanation of the methodology has been extended (P.8, L.12ff). This includes a link from instantaneous and climatological uncertainties to systematic, random, and sampling uncertainties. A mathematical description of E_{clim} is furthermore provided (P.15, L.12f).

8) Comment from referee: *Page 8, Line 10: Definition of water vapour path?*
Author's reponse: The water vapour path (“wvpa“) refers to the vertically integrated water vapour and is therefore a measure of humidity contents in the atmospheric column. It is thus suitable to use as an indicator of the ambient atmospheric conditions. For more information regarding the HOAPS-3.3 wvpa retrieval, please refer to Schlüssel and Emery (1990).
Changes in the manuscript: The term „“water vapour path“ has been replaced by “vertically integrated water vapour“ (P.9, L.2).

9) Comment from referee: *Page 8, Lines 11-14: It seems that HOAPS q_a is wet biased in the tropical wet zone and dry biased in the subtropical dry zone. The bias pattern seems to be similar to GSSFT v3 q_a product (Prytherch et al. 2014, Int. J. Climatol.; Jin et al. 2015, J. Atmos. Ocean. Technol.).*
Author's reponse: Thank you for pointing this out. Indeed, Figure 4c in Prytherch et al. (2014) shows a strong resemblance between HOAPS-3.2 and GSSTF3. Both data records are based on the same algorithm and follow an inter-satellite calibration procedure. The minor differences in the tropics are thought to be related to either different quality control standards or differing Earth incidence angles. Given the close resemblance of GSSTF3 and HOAPS-3.2 shown in Prytherch et al. (2014), the difference pattern (GSSTF minus buoys and OAFlux) shown in Jin et al. (2015) was to be expected. The distribution is closely related to the q_a-dependent bias pattern shown in our manuscript (Fig. 2a).
Changes in the manuscript: Prytherch et al. (2014) is cited in this context (P. 9, L.9f).

10) Comment from referee: *Page 8, Lines 21-22: Indeed, the 1-D bias analysis is not sufficient. Please provide a figure showing the global pattern of the mean differences between HOAPS and the reference data. Need to discuss the uncertainty pattern in terms of humidity regimes.*
Author's reponse: Thank you for your suggestion. Originally we thought the reader would be distracted by such a difference map, as we would like to emphasize the importance of considering *multiple* atmospheric state parameters, i.e., the multi-dimensional bias analysis. However, we agree that the manuscript improves when including such a difference map (HOAPS minus in situ q_a).
Changes in the manuscript: The difference map has been included into the revised manuscript (Fig 1, right panel). It is briefly described in Section 3.1 (P.9, L.10f,L.25f), where a connection to Fig. 2a (of revised manuscript) is established.

11) Comment from referee: *Page 9, Line 24: “Recall that the aim is to characterize uncertainty and not bias patterns”. The sentence is confusing. Bias is one kind of uncertainties.*
Author's reponse: We disagree with this statement. According to the International Vocabulary of Metrology (VIM, 2012), the (measurement) uncertainty is a *non-negative* parameter characterizing the *dispersion* of the quantity values being attributed to a measurand, based on the information used (VIM, 2.26). By contrast, a (measurement) bias (VIM, 2.18), which corresponds to an estimate of a systematic measurement error (VIM, 2.17), may be either positive or negative and, if known, can be corrected for. Keeping these two definitions in mind, a bias, which is a signed value, is strictly speaking not a kind of uncertainty. In order to turn the bias into an uncertainty estimate, we use the absolute systematic difference as an upper boundary of the (more simple) bias distribution.
Changes in the manuscript: The wording in the revised manuscript has been modified and moved

further up in Sect. 3.2 (P.10, L.17-22).

12) Comment from referee: *Page 10, Eqs (2)-(3): Which figures are produced from Eqs.(2)-(3)?*
Author's reponse: Figs. 3-6 are based on Eqs. 2 and 3. Details are provided in the following. Whereas Eq. 3 merely expresses that the total instantaneous LHF uncertainty consists of a systematic and a random component, Eq. 2 forms the basis of LHF pixel-level uncertainties using uncertainty propagation. That is, applying Eq. 2 equips each LHF pixel with a *total*, that is systematic plus random uncertainty contribution. In consequence, Figure 4d directly results from Eq. 2, that is the systematic uncertainty contribution (the random component convergences to zero, due to averaging over long time period). Likewise, the systematic uncertainty contributions by U, q_s, and q_a , which contribute to Eq. 2, are illustrated in Figs. 4a-c.
Note that the random uncertainty measures resulting from Eq. 2 still incorporates random uncertainty contributions of the collocated in situ data (E_{ins}) as well as the collocation procedure itself (E_c). Each random uncertainty contribution resulting from Eq. 2 needs to therefore be corrected to isolate the random *retrieval* uncertainty. This random retrieval uncertainty is what we would like to characterize in the HOAPS climatology. The random LHF uncertainty resulting from Eq. 2 is therefore corrected pixelwise, using the results of the random uncertainty decomposition (see Sect. 3.4 and e.g. Figure 2 in Kinzel et al. (2016) for q_a). The average field of these instantaneous, corrected random retrieval uncertainties is shown in Fig. 3d. Respective random retrieval uncertainty components contributed by U, q_s, and q_a, are shown in Figs. 3a-c, respectively. As noted in the manuscript, Fig. 3 shows the *instantaneous* point of view, that is N=1.
Likewise, Fig. 5 shows both systematic (rectangles) and instantaneous random retrieval (bars) uncertainties. It therefore shows the maximum uncertainty one can expect for a single pixel for different geographical regimes. Figure 5 is therefore based on both Eqs. 2 and 3. The same accounts for Fig. 6. The technical aspects are described in Sect. 3.4-3.5 .

13) Comment from referee: *Page 10, Line 10: Why only random satellite retrieval component, not the total random uncertainty, is computed?*
Author's reponse: The purpose of our uncertainty characterization is to assign systematic, random, and sampling uncertainties to all *satellite-related* LHF parameters. This approach is unique and important, as simply assigning total random uncertainties does not allow the user to understand to what extent they are associated with the retrieval itself or other uncertainty sources. This implies that contributions by collocation (E_C) and in-situ data (E_{ins}) need to be corrected for (i.e., removed) by applying the random uncertainty decomposition (Sect. 3.3). What remains is the random retrieval uncertainty, which consists of both random model uncertainty (E_M) and sensor noise (E_N) (see Kinzel et al. (2016), their Eq. 5).
Immler et al. (2010) formulate an implication of such an approach for consistencies like this: „Roughly speaking, consistency is achieved when the independent measurements agree within their individual uncertainties“ (their Sect. 2.5, Eq. 6). In other words, the decomposition of uncertainties allows for comparing two independent measurements with *own* (that is, independent) uncertainties, which makes conclusions regarding consistency more meaningful. The decomposition and contributing random uncertainties are thoroughly explained in Kinzel et al. (2016), their Sect. 2c.
Changes in the manuscript: Immler et al. (2010) has been added to Sect. 1 for a clearer motivation of our uncertainty decomposition approach (P.4, L.1f).

14) Comment from referee: *Pages 10-11, sections 3.4-3.5: The two sections are not directly related to any figures. Suggest to revise and combine.*
Author's reponse: We disagree that these two sections are not directly related to any figures/tables in the manuscript. For transparency, we believe a clear separation of all HOAPS-related uncertainties, that is systematic and random retrieval uncertainty (Sect. 3.3-3.4) and sampling uncertainty (Sect. 3.5), is appreciated. Sect. 3.3 is a main prerequisite for what is shown in Figs. 3 and 5, respectively. Sect. 4.3 (and Table 2 therein) is dedicated to only E_{smp}, which is first picked up

in Sect. 3.5.

15) Comment from referee: *Page 13, Line 10: Fig.2 is regarded as a 2-D representation of the error bar magnitude of Fig.1a. A figure showing the global pattern of HOAPS3.3 - minus - in situ needs to be provided to help interpret Fig.2.*
Author's reponse: The differences map points at biases, which are not linked to the random retrieval uncertainties shown in Fig 3a. Yet, the differences map (HOAPS minus in situ) has been added to the revised manuscript, where it is also commented on (P.9, L.10f,L.24f). This is already picked up in a different context (see comment #10 on this). As noted in the manuscript, the quoted passage is meant to qualitatively link the error bars in Fig. 2a to the four-dimensional (Fig. 3a) *random retrieval uncertainty* representation. Differences in their magnitudes were to be expected, as the bars in Fig. 2a include both E_C and E_{ins}, which have been corrected for in Fig. 3a. However, the q_a-dependent distribution of error bar magnitudes (Fig. 2a) are very closely related to the $E_{retr,ran}$ pattern (Fig. 3a) . That is, random retrieval uncertainties are largest for subtropical ranges of q_a (11-17 g kg^{-1}, Fig. 3a), which is mirrored in largest uncertainty bars in Fig. 2a. Likewise, these magnitudes reduce for tropical q_a ranges of roughly 20 g kg^{-1}. Smallest magnitudes are generally found in high latitudes, where q_a is smallest (below 7 g kg^{-1}, see Fig. 2a). The intention was to show the spatial distribution of random uncertainty in HOAPS-3.3 q_a. As mentioned later on, this random uncertainty can be neglected if monthly to multi-annual averages are considered, while systematic components become the dominating source of uncertainty. Spatial maps of these long-term means of systematic uncertainties are provided in Fig. 4.
Changes in the manuscript: See comment #10.

16) Comment from referee: *Page 13, Fig. 2: The instantaneous random uncertainty map of q_a (Fig.2a) has a pattern similar to the uncertainty map of q_a produced by OAFlux (Yu et al. 2008, OAFlux technical report), though HOAPS3.3 has a much larger magnitude.*
Author's reponse: Thank you for bringing up this comparison. We agree that the error distribution shown in Yu et al. (2008) resembles our instantaneous random uncertainty distribution. Regarding uncertainty magnitudes: Yu et al. (2008) declare “mean errors“ as monthly mean standard deviations (std) (time period: 1958-2006). This definition considerably differs from our approach. Furthermore, it remains unclear as to how this std is derived. Apparently, several data sets contribute to its estimation (NCEP1, NCEP2, ERA40, satellites), which may be the cause for lower magnitudes shown in their Fig. 21. Whereas our uncertainty estimates are exclusively HOAPS-related (that is, related to only one data record), the error estimation presented in Yu et al. (2008) does not clarify as to how the global error distribution includes contributions by the individual data sets.

17) Comment from referee: *Page 14, Line 3: In addition to Table 2, please add a zonal-mean average of the monthly mean sampling uncertainties to show the latitudinal distribution of the uncertainties.*
Author's reponse: We investigated the latitudinal dependency of all sampling uncertainties. Due to the large averaging time period (monthly means), there is hardly any zonal dependency evident in any of the parameters (not shown). This was to be expected, as a differentiation between tropical and extratropical buoys for quantifying monthly mean sampling uncertainies did not reveal differences in uncertainty magnitudes (see end of Sect. 3.5). As indicated in Table 2, sampling uncertainties averaged over such long time scales only show a dependency on the amount of orbiting platforms. However, this effect is not seen in the zonal means, as at least three instruments were in operational mode between 1995-2008.
Changes in the manuscript: A comment has been included into the revised manuscript (P.12, L.26f) that no latitudinal dependency of the sampling uncertainties exists on the monthly mean basis.

18) Comment from referee: *Page 14, Line 13: How is E_{clim} defined? Please provide a mathematical expression of E_{clim}.*
Author's reponse: Please refer to comment #7 on this.
Changes in the manuscript: Please refer to comment #7 on this.

19) Comment from referee: *Page 14, Line 15: "Figures 3a-e can also be treated as the systematic uncertainty distribution". What is the relation between Figures 3a-e and the mean difference map of HOAPS-3.3 minus in situ? See comment Page 13, Line 10. The maps shown in Figures 3a-e are not bias patterns, as bias has both positive and negative signs. What is the meaning of the systematic uncertainty?*
Author's reponse: We apologize that the current formulation may be confusing. Regarding the phrase you quoted: When averaging over 25 years (1988-2012), random and sampling uncertainties become virtually zero. This implies, given our definition of E_{clim} (see comment #7 on this), that E_{clim} is pratically equal to the systematic uncertainty (E_{sys}), which in turn is the absolute representation of the bias (see Sect. 3.2). Throughout our manuscript, we do not speak of „bias patterns", as we are characterizing *uncertainties*, which are per definition non-negative. The average of an array of biases with respect to a reference can be zero, while none of the individual match ups are actually equal. This automatically points at a non-zero uncertainty. In this regard, we agree that Fig. 4 (of revised manuscript) does not show bias patterns (unlike Fig. 1 (right) in the revised manuscript), but rather patterns of E_{sys}. E_{sys} is therefore the *upper boundary* of the (more simple) bias distribution (see Sect. 3.2).
Changes in the manuscript: The q_a difference map (HOAPS minus in situ) has been included into the revised manuscript (Fig. 1 (right), see comment #10 and #15 on this). It is briefly described and related to Figs. 2a (P.9, L.25). Also, the composition of Sect. 3.2 has been changed.

20) Comment from referee: *Page 18, Line 5: "On average, it increases by roughly 4.5 W m^{-2} (4.7%) per decade…". Which term gives rise to this large increase, $q_a - q_s$ or U? The continuing increase in LHF during the "hiatus" period in the 2000s does not seem realistic from the perspective of the global water budget balance (see Robertson et al. 2014, J.Clim).*
Author's reponse: Thank you for bringing this up. As mentioned in Sect. 4.7, this linear LHF increase over time is picked up by numerous studies and is resolved in several climatologies. Yu et al. (2007), for example, point at an OA Flux LHF increase of 9 W m^{-2} over a time period of 22 years (1981-2002), which closely resembles our linear trend estimate. Our trend analysis includes a strong negative offset in HOAPS LHF during 1991. As pointed out in the manuscript (P.18, L.25ff), this is associated with retrieval issues related to the Mount Pinatubo eruption and is therefore an artificial signal. If this is solved, as has been done for the latest HOAPS version, HOAPS 4.0 (Andersson et al., 2017), the offset is smaller, which ultimately reduces the linear trend. Also, possibly related to the hiatus, global mean HOAPS LHF slightly decrease after 2008. GSSTF3 also exhibits an LHF increase up to 2007/8 (which is even stronger than that of HOAPS) and a subsequent decrease (see Robertson et al. (2014), their Fig.2b and Fig.8). Regarding the increase up to 2008, the same conclusion may be drawn for SeaFlux (Robertson et al. (2014), their Fig. 2c). As to the cause of the LHF increase: Q-term analysis indicates that linear trends of both U and q_s are positive, whereas that of q_a is negative. In consequence, both U and (q_s-q_a) give rise to the observed LHF increase. For the time period of 1988-2005, this also becomes evident in Iwasaki et al. (2014), their Fig. 9.

21) Comment from referee: *Page 19, Line 14: Remove the sentence. Aren't the uncertainty estimates supposed to be a common practice for all gridded products?*
Author's reponse: We think that it is appropriate to include this sentence in our manuscript, as we are not aware of any other satellite climate data set with such an (extensive) uncertainty characterization. We certainly agree that this should be a common practice in the future. It seems, however, that HOAPS-3.3 (and HOAPS 4.0, by now) leads the way.

Cited studies:

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

Bentamy, A., Katsaros, K. B., Mestas-Nuñez, A. M., Drennan, W. M., Forde, E. B., and Roquet, H.: Satellite Estimates of Wind Speed and Latent Heat Flux over the Global Oceans, J. Climate, 16, 637–656, doi:10.1175/1520-0442(2003)016<0637:SEOWSA>2.0.CO;2, 2003.

Brodeau, L., Barnier, B., Gulev, S. K., and Woods, C.: Climatologically Significant Effects of Some Approximations in the Bulk Paramererizations of Turbulent Air-Sea Fluxes, Journal of Physical Oceanography, 47, 5–28, doi:10.1175/JPO-D-16-0169.1, 2017.

Immler, F. J., Dykema, J., Gardiner, T., Whiteman, D. N., Thorne, P. W., and Vömel, H.: Reference Quality Upper-Air Measurements: guidance for developing GRUAN data products, Atmospheric Measurement Techniques, 3, 1217-1231, doi: 10.5194/amt-3-1217-2010, 2010

International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM 3rd edition), JCGM, 200, 2012

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.

Robertson, F. R., Bosilovich, M. G., Roberts, J. B., Reichle, R. H., Adler, R., Ricciardulli, L., Berg, W., and Huffman, G. J: Consistency of Estimates Global Water Cycle Variations over the Satellite Era, Journal of Climate, 27, 6135-6154, doi: 10.1175/JCLI-D-13-00384.1, 2014

Schlüssel, P.: Satellite Remote Sensing of Evaporation over Sea, in: Radiation andWater in the Climate System: Remote measurements, Vol. 45, NATO ASI Series, pp. 431–461, Springer-Verlag, Berlin, Germany, 1996.

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

Yu, L., Xiangze, J., and Weller, R.A.: Multideace Global Flux Datasets from the Objectively Analyzed Air-sea Fluxes (OAFlux) Project: Latent and Sensible Heat Fluxes, Ocean Evaporeation, and Related Surface Meteorological Variables, OAFlux Project Technical Report, OA-2008-01, WHOI, 2008.

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

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:

1) Main comment from referee:

a) The paper is hard to read. Often it requires re-reading a paragraph a number of times, to understand. It also has to do with the structure of the paper. It would help to define the main methodology of the data analysis and to have this as the main thread throughout the paper.

I think I understand the methodology, but I am still not sure.
Let me explain my interpretation of the method:
(i) Four dimensional look-up tables (LUT) are created of co-located data, so differences between data sets are stratified according to q_a, U, SST, and wvpa.
(ii) The mean difference between HOAPS and insitu data are interpreted as biases.
(iii) The variances of the difference are used for the triple co-location method, resulting in error estimates.
(iv) This results in LUT's of biases and random error estimates.
(v) In applications (e.g. global maps of mean and random error of q_a) the observations of q_a, U, SST and wvpa point to the table and provide errors of each observation. These can be averaged to obtain the desired map.

b) I feel that it would be helpful to describe upfront that this is the general methodology and follow it throughout the paper. So this would lead to 3 main sections in the paper: (i) Description of the methodology, (ii) Results of the methodology, i.e statistics on the LUT data, and (iii) Application to HOAPS evaporation. In case I am completely wrong on the interpretation of the paper, there is even more reason to be clear about the methodology.

c) Another question is: what is the main result of the paper? If my interpretation is correct, then the 4-dimensional table of error estimates is the main result, because it would allow a user to make estimates of anything he/she is interested in (e.g. monthly averages, daily averages, or El-Nino years). So it is worth thinking about communicating this 4D table to the users. Most of the current paper is about applying the methodology, but these are in fact just examples.

Author's reponse:

Related to a) We have restructured the paper to be more clear about the methodology. The introduction has been rearranged and shortened (see also specific comment #1 below) for a much clearer understanding of the motivation, the benefit, and the structure of the paper. We have furthermore appended a flowchart to this document, which guides you through the individual steps of data processing, intermediate data products, and resulting HOAPS uncertainty measures.

Regarding your methodology interpretation (the flowchart assists):
(i) Correct.
(ii) Yes, differences of paired collocations are considered as biases. Depending on U, q_a, SST, and wvpa, these single biases are assigned to one of the 20^4 bins. Once all collocations have been

assigned to a bin, bin-averaged systematic uncertainties are computed based on the *absolute* differences of all assigned biases. That is, we consider the upper end of bias considerations (see comment #10 on this).
(iii) Exactly. Although the variances of differences, applied to the triplets, only help to decompose the *random* uncertainty estimates to end up with HOAPS-related (that is, retrieva-related) random uncertainty estimates.
(iv) The LUTs of systematic and (uncorrected) random uncertainties already result from i). iii) helps to decompose the random uncertainty components, isolate the retrieval-related part, and finanly correct the random uncertainty LUTs.
(v) Yes, this is correct.

Related to b) Regarding your proposed three main sections: We believe that this has already been done in a similar manner and is reflected in the numbering of the Section: Methodology (Sect. 3) and Results/Applications (Sect. 4). We do not want to dedicate „Results of the methodology“ an own section, as we think that it belongs to the methodological part of the manuscript (Sects. 3.3, 3.4).
However, we agree that the submitted manuscript was not structured clearly enough. This shortcoming has been improved (see comment related to a) above).

Related to c) We agree that *one* of the main outcomes of the manuscript is the benefit of the multi-dimensional bias analyses. Particularly because the approach can be easily transferred to other satellite retrievals and potentially also other remotely sensed parameters. The approach itself should be stressed more clearly in the conclusion. Communicating our specific LUTs to the users is not helpful, though, as they are tailored to HOAPS-3.3 (due to the double collocations described in Sect. 3.1). The results of applying an updated version of the LUTs to instantaneous HOAPS data are implemented in the most recent HOAPS 4.0 Version (Andersson et al. (2017)) in form of systematic and random uncertainties.
We believe that the application of the uncertainty characterization approach is equally important, as it leads to uncertainty estimates for a widely used data record. On the one hand, none of the remaining LHF-related satellite climatologies are equipped with such estimates. On the other hand, Sect. 4 demonstrates a variety of different approaches for illustrating the uncertainties and allows for identifying regions where uncertainties in the satellite retrieval are an issue and need to be accounted for.
The focus of this paper is therefore twofold: 1) describing the method and 2) applying the method to arrive at HOAPS-3.3 uncertainty estimates.

Changes in the manuscript:

Related to a and b) The whole manuscript has been revised for a clearer reading experience. This specifically targets Sects. 1 and 3. The last paragraph of Sect. 1 now guides the reader through the manuscript step by step. Section 3.3. and 3.4 have been swapped to be consistent with the sequence of analyses.

Related to c) The benefit of the multi-dimensional bias analyses for uncertainty characterizations has been highlighted more clearly in Sect. 5 (P.20, L.21ff).

---

2) Main comment from referee:

Estimation of biases is non-trivial. In fact this is very important because, as the authors point out, for long term averages the systematic errors dominate.

My concern is two-fold:
a) I have the feeling that it is assumed that DWD-ICOADS data is bias-free? If this is the basis for the bias estimates, then it deserves more discussion also in view of what has been published in literature.
b) Fig. 1 is used as an example to illustrate the estimation of biases. However, it is likely that artificial biases occur in binned scatter plots of noisy data if correlated variables are used on abscissa and ordinate. This applies to Fig. 1a where hair(HOAPS) is used on both vertical and horizontal axes. It also applies to hair versus wind because these variables are correlated due to the physics of the mixing (more wind brings hair closer to the surface value). To check, one could e.g. bin the differences of Fig. 1a in classes of hair(insitu). Also hair(insitu) is noisy because it has large representativeness errors (point observation, whereas HOAPS has a large footprint).

c) Finally, if one can be confident about the bias estimation, then it should also be trivial to apply a bias correction to HOAPS. This would just leave the uncertainty in C_E which is a parametrization constant used for satellite as well as in-situ data. Please discuss.

Author's reponse:

Related to a) It is correct that we assume the DWD-ICOADS data base to be bias free (see last paragraph in Sect. 2.2 of submitted manuscript). Our filtering procedure ensures that only high-quality in situ data is used for collocation analysis. Systematic effects of known origin are thought to have been removed or at least minimized within the quality checking procedure at the Marine Climate Data Center of DWD. Other systematic uncertainties like differing sensor heights and cool skin effects have been eliminated prior to our analysis due to sensor height corrections using in situ platform meta data (U) and cool skin corrections (q_s). We are aware of the fact that no ground "truth" exists, but are confident that our extensive data base is the best ground "reference" available. Freeman et al. (2016) present a great overview of the variety of ICOADS applications, which also include the calibration and validation of satellite data (e.g. Bentamy et al. (2003), Bentamy et al. (2013), Jackson et al. (2009), Jackson and Wick (2010)).
It should be kept in mind that our systematic uncertainty estimates represent the upper limit of a more simple bias estimation. Assuming a bias free ground reference therefore does not violate our conclusions, although a small contribution to the systematic uncertainties may be caused by the in situ reference.
One could argue that our uncertainty estimates in regions of poor in situ data coverage are questionable. However, as picked up in Sect. 3.2, we overcome the regional dependency by characterizing uncertainties as a function of ambient atmospheric conditions. Poor in situ data densities are therefore of secondary importance, as their ambient atmospheric conditions may be similar in regions with considerably more match ups.

Related to b) Thank you for the suggestion to investigate the one-dimensional patterns of dq as a function of the in situ source. We exemplarily performed this analysis for 2001 with approximately 1.8 million match ups. We compared the magnitudes of the mean 5-percentiles, which (in case of HOAPS) are illustrated as black squares in Fig. 2. For U, q_s, and q_a, our results indicate that in 80% of all match ups (i.e., excluding the margins), relative differences between HOAPS and in situ mean 5-percentiles range between ± 6-10%, which we consider as negligible. We presume that a two- instead of one-sided regression approach would lead to even more robust 5-percentile means. Towards the margins of the distributions, relative differences become larger. We believe that this does not have a noteworthy impact on the four-dimensional analyses, as the biases in one-dimensional space may become smaller or even cancel out when the remaining three atmospheric state parameters are considered concurrently.
Independent of this, biases as a function of in situ LHF-related parameters cannot be investigated in four-dimensional space, as vertically integrated water vapour („wvpa"), an important indicator for

the prevalent atmospheric condition, is not available from in situ measurements. This would lead to an undesirable simplification of our uncertainty analysis approach. Additionally, our match up data base only lasts until 2008. In consequence, no uncertainties could be assigned to pixel level HOAPS data from 2008 onwards, if the multi-dimensional bias approach was based on in situ data.

Related to c) Regarding a bias correction of HOAPS data: Our approach aims at characterizing uncertainties inherent to HOAPS. This allows users to implement this information into their analyses and arrive at appropriate conclusions. We have further emphasized the benefit of such estimates in the revised version of Sect. 1. The focus is therefore not put on bias correction with respect to DWD-ICOADS. A sustainable consequence of large uncertainties should in fact point at the need of modifying the retrieval algorithm instead of bias correcting the data. It is our impression that a bias correction is feasible, if a constant bias (in terms of dependent variables, region, or time) is present relative to a fiducial reference. Such a reference is not available at present.

Changes in the manuscript:

regarding a) Freeman et al. (2016) is picked up in context of describing the in situ data base (P.6, L.18f). Furthermore, some further references are given regarding our assumption of bias-free ICOADS measurements (P.8, L.3ff).

regarding b) We briefly mention the artificial biases due to correlating variables and conclude that two-sided regression analyses could reduce these spurious biases (P.9, L.17ff).

SPECIFIC COMMENTS:

1) Comment from referee:
Section 1: Although well written, the introduction is rather long and contains sometimes fluffy language. No reference is made to an earlier study by Kinzel et al. (2016). What is new compared to earlier work? Reference is made to other data sets and to studies that provide error estimates. However, nothing is said about published error estimation methods.
Author's reponse: We agree that the introduction of the submitted manuscript is too long. We have restructured Sect. 1 following your suggestions and believe that this essentially improved the manuscript (see also “main comment 1“ on this). Furthermore, Kinzel et al. (2016) has been included in the revised manuscript to point at the random uncertainty decomposition approach. We agree it is important to distinguish between earlier work and new aspects of this manuscript. This also includes a statement regarding earlier error estimation methods and those present in our manuscript.
Changes in the manuscript: The first 24 lines have been considerably shortened.
The whole introduction has been restructured to be clearer about the motivation and benefit of our study. It now clearly differentiates between earlier approaches (that is, mostly intercomparison studies, P.3, L.25ff) and the novelty of our uncertainty characterization (that is, uncertainty estimates that are exclusively related to a specific data set, in particular HOAPS, P.4, L.1f.; P.4, L.14ff). At the same time, we highlight the new aspects of our approach (e.g. four-dimensional LUTs, P.4, L.20f). In this regard, Kinzel et al. (2016) has been included into the revised manuscript and is put into context (P.4, L.10f, L.21f). The advantage of multi-dimensional LUTs has been included into the abstract (P.1, L.7f).

2) Comment from referee: Page 5, Line 32: The sentence with "The latter depends" suggests that it refers to q_a in the sentence before, but what it intends to say is that the COARE algorithm needs stability and that specific assumptions are made. Please rephrase.
Author's reponse: The wording was chosen on purpose, as we wanted to point out that the

saturation vapour pressure and hence q_a depends on the surface air temperature. However, we agree that this may be confusing and the focus should be put on the stability calculation.
Changes in the manuscript: We changed the wording to “It includes atmospheric stability calculations, which necessitate surface air temperatures as input. These are estimated by assuming...“ (P.6, L.7f).

3) Comment from referee: Page 6, Line 20-24: The non-correction of q_a for measuring height is confusing. Why not using the real measuring height in the bulk formula? Perhaps it is possible to say in one sentence what the results are of the height difference effects as estimated by Kent et al. (2014).
Author's reponse: Prytherch et al. (2014) and Kinzel et al. (2016) point at the disadvantages related to q_a height corrections. We agree that a statement regarding the height correction effect is useful. Kent et al. (2014) quantify the height correction effect to be 0.11 g kg^{-1} for the time period 1971-2006, owing to the continuously increasing measurement platform heights. However, this effect is masked by bias corrections associated with measurement techniques, which are thought to be 2-3 times larger.
Changes in the manuscript: Results by Kent et al. (2014) regarding the height difference effects are briefly mentioned (P.7, L.3ff).

4) Comment from referee: Page 7, Line 13: Cool skin corrections are applied to in situ observation but not to HOAPS-3.3 SST (AVHRR based). This makes sense in priciple because AVHRR measures the skin temperature. However, there must be a calibration procedure of AVHRR, which is probably against bulk SST data. So, what does calibrated AVHRR data represent, bulk or skin SST?
Author's reponse: Thank you for bringing this up. Indeed, AVHRR was calibrated against bulk SST. Formally, this would necessitate a cold skin correction. However, compared to OISST (Reynolds et al. (2007)), AVHRR has a cold bias of unknown origin, which is in the order of the skin correction. We therefore refrained from performing the correction and consider the AVHRR SST as a skin SST. Note that this cold bias problem is overcome in HOAPS 4.0 (Andersson et al. (2017), which is based on OISST. For the HOAPS 4.0 retrieval (Andersson et al. (2017)), OISST is corrected for the cold skin effect.

5) Comment from referee: Pages 4-5 section 2.1: It would be informative to mention pixel size of the microwave sensors.
Author's reponse: Yes, we agree.
Changes in the manuscript: Pixel sizes have been included into Sect. 2.1 of the revised manuscript (P.5, L.11-13).

6) Comment from referee: Page 8, Line 11: The sentence "Figure 1a overestimates ...." is confusing. Formally it is correct, but, after reading the first time it suggests that the biases range from 7-12 g/kg and that the plot is for the inner tropics.
Author's reponse: Indeed, this may be misunderstood.
Changes in the manuscript: The wording has been changed in the revised manuscript to: “For q_a values between 7-12g kg^{-1} , HOAPS-3.3 overestimates near-surface specific humidities (see Figure 2a). Overestimations are also observed in the inner tropics, where q_a is in the order of 20 g kg^{-1}“ (P.9, L6f).

7) Comment from referee: Page 8, Line 17: The expression "over-(under-)estimated" is perhaps better than "over-(under-)represented"
Author's reponse: Thank you for this suggestion.
Changes in the manuscript: "over-(under-)represented has been replaced by "over-(under-)estimated" (P.9, L.12).

8) Comment from referee: Scatter plots in Fig.1: In all the plots except (c) the variables on the vertical axis are correlated with the variable of the horizontal axis. This is most obvious for Fig. (a) where hair-HOAPS is used in both abscissa and ordinate. In such cases the binning according to one axis can show biases that are not necessarily real. Whether this is really the case can be easily demonstrated by making the same plot but now with hair-insitu on the horizontal axis. Similarly unrealistic bias may be seen in (b) and (d) because wind and wvpa are derived with from the same satellite channels and therefore correlate with hair-HOAPS. Please discuss.
Author's reponse: We assume that “hair-insitu“ means “q_a(in situ)“ and not the mathematical difference between HOAPS and in situ q_a? We are aware of the correlation between the individual variables. The aspect of correlating variables is an important remark, which we thoroughly discuss in context of the “main comment 2“ (part b), see further above). In fact, this is fundamental for our multi-dimensional approach: characterizing systematic and random uncertainty estimates of U, q_s, and q_a as a function of atmospheric state parameters, which (as we believe) have an impact on the parameters themselves. Specifically regarding Figure 2d): wvpa is not available from in situ measurements, which is why a bias dependency on in situ wvpa cannot be investigated.
Changes in the manuscript: See “main comment 2“ (part b) further above.

9) Comment from referee: Page 9, Line 21: Please specify what "even stronger winds" are.
Author's reponse: “stronger wind“ mean wind speeds exceeding 20 m s^{-1}.
Changes in the manuscript: The wording has been changed in the revised manuscript (P.10, L.9).

10) Comment from referee: Page 9, Lines 24-26: This paragraph is hard to read. After reading, a number of times times, I think I understand. Is it not better to say: "Our goal is to document the upper bound of the bias and therefore we take the absolute value of the possible systematic error in CE"?
Author's reponse: We agree that this paragraph is somewhat confusing and out of place. It has been moved further up into the appropriate context.
Changes in the manuscript: The wording has been modified and has been moved further up into the appropriate context (P.10, L.17ff).

11) Comment from referee: Page 10, line 15 and page 11, line 7: I suggest to replace "Next to" by "In addition to"
Author's reponse: Thank you for this suggestion.
Changes in the manuscript: The wording has been changed in the revised manuscript (P.12, L.12).

12) Comment from referee: Page 11, section 3.5: This section is hard to read. If I understand correctly, it addresses the question: Does it matter for the averages that the satellites sample the ocean at particular times of the day only, given that a diurnal cycle may be present? The authors investigate by looking at buoy data and by comparing averages that cover the full diurnal cycle with samples at satellite overpass times only. Part of the confusion is because it mentions spatial sampling, but I don’t think this section covers that? Please simplify for clarity.
Author's reponse: Exactly. For the monthly mean HOAPS product (HOAPS-G), sampling uncertainties need to be quantified because of the diurnal cycle of the geopyhsical parameters. Due to the sun-synchronous satellite overflights, diurnal cycles or frontal passages are likely to be missed. This will affect the monthly mean averages. We agree that the the term “spatial sampling“ is misleading, as we only cover the temporal sampling issue.
Changes in the manuscript: The aspect of “spatial sampling uncertainties“ has been removed from the revised manuscript to avoid confusion.

13) Comment from referee: page 13, Lines 9-11: I am not sure that it is helpful here to refer to Fig. 1a, because it is showing the combination of E_ins(q_a_a) and E_retr(q_a), which is different from

Fig. 2a. The authors point this out but instead of clarifying something it confuses.
Author's reponse: We agree that this may be confusing.
Changes in the manuscript: This section has been shortened to become more clear. (P. 14, L.18ff)

14) Comment from referee: Page 13, 23: Suggestion: replace "merely" by "only"
Author's reponse: Thank you.
Changes in the manuscript: “merely“ has been replaced by “only“. (P.14, L.31).

15) Comment from referee: Page 13, Line 24: What is meant by "local minimum in that region for q_a"? E_retr(q_a) has a maximum over the warm pool.
Author's reponse: This is a mistake in our manuscript, thank you for pointing this out. We wanted to point at the q_a random retrieval uncertainty, not q_a itself.
Changes in the manuscript: the wording has been changed (P.14, L.31f).

16) Comment from referee: Page 13, Line 29: In the sentence "Respective values partly exceed 50 W/m^2", what is meant by "respective" and "partly"? Do the authors mean: "In these areas, values are found in excess of 50 W/m^2"?
Author's reponse: Yes, this is correct.
Changes in the manuscript: The wording has been changed in the revised manuscript. (P.15, L.3).

17) Comment from referee: Page 14, Line 33: "direct eddy covariance" is not wind speed.
Author's reponse: Sorry for not being correct here. The wind stresses are based on inertial-dissipation methods. Together with eddy covariance based LHF, the turbulent fluxes of a variety of satellite, reanalysis, and combined products are evaluated.
Changes in the manuscript: The wording has been changed. (P.15, L.31ff).

18) Comment from referee: Page 16, Lines 1-2: This is an interesting example, where it is explained that q_a retrievals may be in error because of dry air advection. However, it is not clear how the systematic error analysis picks up the area of the Agulhas current. The systematic error estimation is entirely driven by U, q_a and SST and wvpa (if I understand correctly).
Author's reponse: It is correct that the systematic uncertainty estimation is entirely driven by combinations of ambient U, q_a, SST, and wvpa. Our multi-dimensional bias approach does not point at specific regions. This implies that we cannot be 100% certain that the observed uncertainties over the Agulhas region are exclusively associated with local retrieval issues. In general, match ups over a region contribute to the look up tables (LUTs), which implies these regions are somewhat mirrored in the LUTs. However, they are not explicitly resolved.
The following serves to explain how the LUTs pick up the Agulhas region: Figure 1 (left) indicates that numerous collocations between buoys/ships and satellite exist in this area, which is characterized by a unique combination of ambient U, q_a, SST, and wvpa. In case of the mentioned dry cold air outbreaks from the South, q_a will be anomalously low and hence q_s-q_a and LHF anomalously large. According to Santorelli et al. (2011), satellite retrievals seem to encounter difficulties with these dry cold air outbreaks, which implies that they will not capture q_a correctly. This would for example be seen when investigating dq_a. That is, differences between satellite and in situ q_a would be negative, which directly impacts our four-dimensional uncertainty analysis. In conclusion, repetitive retrieval issues over a specific regions will be manifested in the LUTs and will eventually be seen in systematic uncertainty maps. At the same time, underestimated q_a along the Agulhas Current contribute to an increase in the random uncertainty component of the LUTs.

19) Comment from referee: Section 4.6 and Fig. 4: Here both systematic and random errors are discussed region by region and climatologically versus January/July. Earlier in the paper it was concluded that the random errors were small compared to the systematic errors. However in Fig. 4 the random errors are larger than the systematic errors. Furthermore I would expect that the

climatological data (I assume averaged over the entire period) has much more data than the January or July data and therefore much smaller random errors.
Author's reponse: As mentioned in Sect. 4.6, the error bars in Fig. 5 point at *instantaneous* random uncertainties (such as those shown in Fig. 3). The idea is to show the maximum uncertainty to be expected for a specific region and season on an *instantaneous* basis. This approach allows for illustrating random uncertainties, as they often even exceed the systematic counterpart for pixel-level data, as is seen when comparing Fig. 3 to Fig. 4. Fig. 3 shows averaged *instantaneous* random uncertainties as a function of region and time. If properly scaled according to the considered period of time, they decease with increasing time period and become insignificant at monthly or multi-annual (that is, climatological) time scale. Keeping this in mind, this also answers the question as to the smaller random uncertainties for multi-annual mean (1988-2012) compared to seasonal means (1988-2012): The difference in error bar magnitudes is not related to averaging periods, as these are averaged *instantaneous* random uncertainties as a function of region and time. We agree, however, that this is not clearly stated in the manuscript.
Changes in the manuscript: The wording as been modified in the revised manuscript to clarify what is shown in Fig. 5. (P.18, L.29-33) This also targets the caption of Fig. 5.

20) Comment from referee: Page 18, Line 31: Please replace "outperforms" by "exceeds"
Author's reponse: Thank you for this suggestion.
Changes in the manuscript: This section has been considerably shortened. The phrase is no longer included in the revised manuscript.

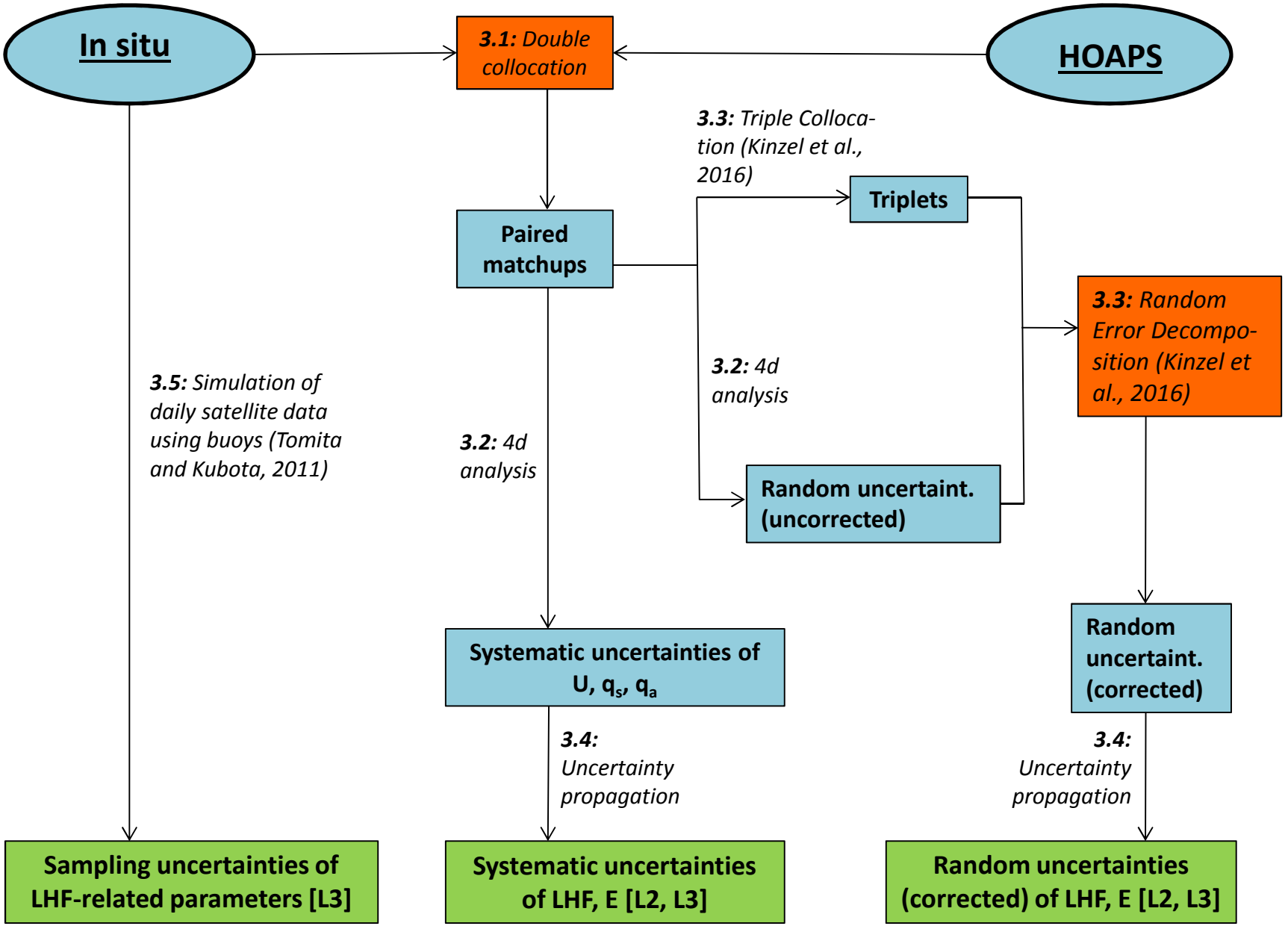
Cited:

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

Freeman, E., Woodruff, S. D., Worley, S. J., Lubker, S. J., Kent, E. C., Angel, W. E., Berry, D. I., Brohan, P., Eastman, R., Gates, L., Gloeden, W., Ji, Z., Lawrimore, J., Rayner, N. A., Rosenhagen, G., and Smith, S. R.: ICOADS Release 3.0: a major update to the historical marine climate record, Int. J. Climatol., p. in press, doi:10.1002/joc.4775, 2016.

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.

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.


In situ
HOAPS
3.1: Double collocation
3.3: Triple Collocation (Kinzel et al., 2016)
Triplets
Paired matchups
3.3: Random Error Decomposition (Kinzel et al., 2016)
3.2: 4d analysis
3.5: Simulation of daily satellite data using buoys (Tomita and Kubota, 2011)
3.2: 4d analysis
Random uncertaint. (uncorrected)
Systematic uncertainties of U, q_s, q_a
Random uncertaint. (corrected)
3.4: Uncertainty propagation
3.4: Uncertainty propagation
Sampling uncertainties of LHF-related parameters [L3]
Systematic uncertainties of LHF, E [L2, L3]
Random uncertainties (corrected) of LHF, E [L2, L3]

Uncertainty Characterization of HOAPS-3.3 Latent Heat Flux Related Parameters

Julian Liman[1], Marc Schröder[1], Karsten Fennig[1], Axel Andersson[2], and Rainer Hollmann[1]

[1]Satellite-Based Climate Monitoring, Deutscher Wetterdienst, Frankfurter Strasse 135, 63067 Offenbach, Germany
[2]Marine Data Center, Deutscher Wetterdienst, Bernhard-Nocht-Straße 76, 20359 Hamburg, Germany

Correspondence to: Julian Liman, Email: julian.liman@dwd.de

Abstract. Latent heat fluxes (LHF) are one of the main contributors to the global energy budget. As the density of in situ LHF measurements over the global oceans is generally poor, the potential of remotely sensed LHF for meteorological applications is enormous. However, to date none of the available satellite products include estimates of systematic, random~~retrieval~~, and sampling uncertainties, all of which are essential for assessing their quality. Here, this challenge is taken on by applying regionally independent multi-dimensional bias analyses to LHF-related parameters (wind speed U, near-surface specific humidity q_a, and sea surface saturation specific humidity q_s) of the Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite (HOAPS) climatology. The multi-dimensional approach overcomes the issue of sparse in situ data densities over large oceanic areas, which makes it very promising. In connection with multiple triple collocation analyses, it is demonstrated how both instantaneous ~~(gridded)~~ uncertainty measures may be assigned to each pixel~~(grid box)~~. A high-quality *in situ* data archive including buoys and selected ships serves as the ground reference. Results show that systematic LHF uncertainties range between 15-50 W m^{-2} with a global mean of 25 W m^{-2}. Local maxima are mainly found over the subtropical ocean basins as well as along the western boundary currents. Investigations indicate that contributions by q_a (U) to the overall LHF uncertainty are in the order of 60 % (25 %). From an instantaneous point of view, random retrieval uncertainties are specifically large over the subtropics with a global average of 37 W m^{-2}. In a climatological sense, their magnitudes become negligible, as do respective sampling uncertainties. ~~Time series analyses show footprints of climate events, such as the strong El Nio during 1997/98.~~ Regional and seasonal analyses suggest that largest total ~~(i.e., systematic + instantaneous random)~~ LHF uncertainties are seen over the Gulf Stream and the Indian monsoon region during boreal winter. In light of the uncertainty measures, the observed continuous global mean LHF increase up to 2009 needs to be treated with caution. ~~First intercomparisons to other LHF climatologies (*in situ*, satellite) reveal overall resemblance with few, yet distinct exceptions~~The demonstrated approach can easily be transferred to other satellite retrievals, which increases the significance of the present work.

1 Introduction

Exchanges of energy and moisture at the atmosphere–ocean interface represent a critical coupling mechanism within the climate system. ~~Roughly 20 of the total radiation absorbed by the Earth's surface is transferred back to the atmosphere by means of turbulent heat fluxes (Trenberth et al., 2009) .~~ Specifically, latent heat fluxes (LHF) significantly control the surface energy budget and are, ~~next~~ in addition to radiative fluxes, one of the main contributors to heating and cooling of the oceans. ~~Approximately 86 of the global evaporation occurs over the ocean basin (Baumgartner and Reichel, 1975) , demonstrating that this water and oceanic energy transfer is a key component of the overall Earth's energy budget.~~ The fifth assessment report of the Intergovernmental Panel on Climate Change (IPCC) emphasizes the role of heat transfer between ocean and atmosphere in driving the oceanic circulation. ~~It stresses that flux anomalies can impact water mass formation rates and alter oceanic and atmospheric circulation (IPCC, 2013) due to its influence on sea surface salinity and thus the ocean surface's density (Grodsky et al., 2009) .~~Additionally, LHF modifies the atmospheric stability distribution and triggers convection, which in turn strongly impacts cloud formation and precipitation. ~~Next to its impact on oceanic processes, this highlights the important role of LHF in modulating the atmospheric circulation on a variety of scales.~~.

To improve our understanding of the global energy and water cycle variability as well as model simulations of climate variations, it is of great importance to accurately measure LHF over the global oceans at the highest possible resolution (e.g. Chou et al., 2004). The need for accurate surface fluxes has, for example, been picked up by the World Climate Research Programme (WCRP), the ~~WCRP~~ Global Energy and Water Cycle Experiment (GEWEX), and the Climate Variations (CLIVAR) Science Steering Group (e.g. Curry et al., 2004). ~~This is ideally achieved through accurate observations and correct implementations of parameterizations in coupled models.~~ Liu and Curry (2006), for example, stress that accurate LHF are essential for a correct forcing of ocean models and for evaluating numerical weather prediction. Additionally, reliable long-term global LHF data ~~sets~~ records represent a substantial input to assimilation experiments, for instance the oceanic synthesis performed by the German contribution to Estimating the Circulation and Climate of the Ocean (GECCO, GECCO2, e.g. Köhl and Stammer, 2008; Köhl, 2015). ~~Such syntheses allow for capturing variability and trends in the turbulent exchange processes, which may exert changes to the entire climate system.~~

Several LHF data records exist, which differ in ~~satellite~~ instrumentation, creation process, data density, as well as spatial and temporal extent. These are either based on *in situ* measurements, reanalysis~~or~~ , remotely sensed data, or a merged version of these. Apart from isolated direct *in situ* measurements using e.g. sonic anemometers, all data ~~sources have in common that~~ methods share a need of bulk flux algorithms ~~are applied~~ such as COARE 3.0a (Fairall et al., 2003) to derive LHF. The near-surface wind speed (U), the saturation specific humidity at the sea surface (q_s), and the near-surface specific humidity (q_a) serve as input bulk parameters, on which the parameterized LHF primarily depend.

~~However, global LHF time series are often subject to uncertainties of unknown magnitudes, which for example hampers the conclusion whether there is a significant multi-decadal trend in global~~ In particular, satellite climatologies have a vast potential for climate research applications, as they incorporate data with high spatial resolution, cover time periods up to several decades, and provide a complete oceanic coverage over ice-free regions.Of these, the Japanese Ocean Flux data sets with Use of Remote

Sensing Observations (J-OFURO) satellite climatology (Kubota et al., 2002), the Goddard Satellite-based Surface Turbulent Heat Flux (GSSTF) Version 3 product (Shie et al., 2012), the updated version of the French Research Institute for Exploitation of the Sea (IFREMER) turbulent flux estimates (Bentamy et al., 2013), the SeaFlux Version 1 and 2 data sets (Clayson et al., 2015), and the Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite (HOAPS) climatology (Andersson et al., 2010; Fennig et al., 2012), amongst others, include LHF-related parameters.The HOAPS data set is a completely satellite-based, single-source climatology of precipitation, evaporation, related turbulent heat fluxes, and atmospheric state variables over the global ice-free oceans. The usefulness of HOAPS for climatological applications has been demonstrated in numerous intercomparison studies and promising results have been published by Bentamy et al. (2003), Bourras (2006), Klepp et al. (2008), Winterfeldt et al. (2010), Andersson et al. (2011), and Stendardo et al. (2016).

In the framework of assessing sea surface freshwater fluxes, Romanova et al. (2010) conclude that HOAPS-3 is well suited for global applications and serves as an important and independent data set that should be included in future ocean syntheses.

Independent of the data source, all global LHF time series are subject to uncertainties, often of unknown magnitudes. On the one hand, *in situ* LHF climatologies, which ~~often~~ include data from buoys and ships, are known to contain biases (e.g. Wang and McPhaden, 2001), to be of variable quality, and to be unevenly sampled. Although research vessel measurements of e.g. q_a are expected to be of good quality (e.g. Roberts et al., 2010), they are regionally limited, which also accounts for data from moored buoys (Weller et al., 2008). Issues related to poor data densities over the Southern Ocean, amongst others, are for example stressed in Yu and Weller (2007), Bourassa et al. (2013), and Prytherch et al. (2014). In consequence, this impedes a meaningful discussion regarding the quality of LHF in this climatologically important region (Josey, 2011). ~~Despite the above-addressed issues, the research community has put effort into uncertainty characterizations regarding *in situ* LHF-related measurements. Whereas random uncertainties of ship-based LHF-related parameters are for example discussed in Gleckler and Weare (1997) , Kent and Berry (2005) , and Kent and Taylor (2006) , systematic uncertainties are assessed in e.g. Kent et al. (1993) and Kent and Taylor (1996) . An example of an *in situ* LHF climatology incorporating uncertainty estimates is given by NOCS v2.0 (Berry and Kent, 2009) .~~

On the other hand, long global reanalysis products such as ERA-Interim (Dee et al., 2011) and NCEP-NCAR (Saha et al., 2010) have a high temporal resolution~~and extent of time series~~, but are not capable of resolving local-scale processes due to a lack of spatial detail (Winterfeldt et al., 2010). Specifically over data-sparse regions, more weight is given to the atmospheric model, which is also prone to uncertainties (e.g. Gulev et al., 2007). ~~At some level~~Thus, atmospheric reanalysis ~~thus~~ suffer from problems in their freshwater budgets (e.g. Schlosser and Houser, 2006; Trenberth et al., 2007).

~~Several remote sensing data records incorporate LHF-related parameters, e.g. the Japanese Ocean Flux data sets with Use of Remote Sensing Observations (J-OFURO) satellite climatology (Kubota et al., 2002) , the Goddard Satellite-based Surface Turbulent Heat Flux (GSSTF) Version 3 product (Shie et al., 2012) , the updated version of the French Research Institute for Exploitation of the Sea (IFREMER) turbulent flux estimates (Bentamy et al., 2013) , the SeaFlux Version 1 and 2 data sets (Clayson et al., 2015) , and the Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite (HOAPS) climatology (Andersson et al., 2010; Fennig et al., 2012) . As all incorporate data with high spatial resolution and cover up to several decades, they have a vast potential for climate research applications.~~

~~The HOAPS data set is a completely satellite-based, single-source climatology of precipitation, evaporation, related turbulent heat fluxes, and atmospheric state variables over the global ice-free oceans. The usefulness of the HOAPS climatology for climatological applications has been tested among numerous intercomparison studies and promising results have been published within Bentamy et al. (2003) , Bourras (2006) , Klepp et al. (2008) , Winterfeldt et al. (2010) , Andersson et al. (2011) , and Stendardo et al. (2016) . In the framework of assessing sea surface freshwater fluxes, Romanova et al. (2010) for example conclude that HOAPS-3 is well suited for global applications and serves as an important and independent data set that should be included in future ocean synthesis.~~

~~As *in situ* and reanalysis data records,~~ Similarly, remotely sensed LHF climatologies are also prone to uncertainties. ~~Next~~ In addition to calibration uncertainties and aliasing problems (Bentamy et al., 2003) , uncertainty sources either originate from uncertainties in the parameterization ~~(Zeng et al., 1998; Brunke et al., 2002, 2003)~~ (Brunke et al., 2002, 2003) or may be linked to the inaccuracy of the input bulk variables (Bourassa et al., 2013). In the framework of an oceanic LHF assessment, Brunke et al. (2011) for example conclude that the uncertainty of HOAPS-3 LHF is ~~largely composed of bulk variable-caused issues~~ to a great extent caused by the bulk variables due to inaccuracies of their individual retrievals. Liu and Curry (2006) reason similarly, while assessing discrepancies of remotely sensed and reanalysis LHF during the 1990s. Romanova et al. (2010) recall that specifically early satellite-based products contain large uncertainties, as also shown by investigations regarding the hydrological cycle by Mehta et al. (2005). ~~The knowledge of both accuracy and precision of the remotely sensed bulk parameters is critical for assessing the quality of satellite-based LHF, as the uncertainties propagate through the applied LHF bulk formula.~~ Finally, irregular sampling from space introduces sampling uncertainties, which may locally become substantial (e.g. Gulev et al., 2007). A current overview study by Loew et al. (2017) highlights the necessity of a thorough satellite-based data validation and pools different approaches across communities.

~~To~~

To date, disagreements and/or weaknesses in data sets are often revealed by performing intercomparison studies, such as those presented by Chou et al. (2004) and Yu et al. (2011) . Another example including HOAPS-3 LHF is presented in Andersson et al. (2011) , who show considerable differences on a local scale. Similar findings are published in Iwasaki et al. (2014) , who compare HOAPS-3 and other data sets to a reference climatology. Results indicate that differences are largest close to 15° N/S, which mostly arise from differing q_a.

Generally, such intercomparison studies are valuable for the research community. By this, however, the source of observed differences remains unknown and can therefore not be attributed to a specific data set. To better quantify the quality of satellite-based data sets, Prytherch et al. (2014) recently emphasized ~~the value grid box based,~~ that comprehensive uncertainty estimates ~~(in their case of q_a) would have~~ are valuable for climate research purposes. To date, none of the above-listed, satellite-based data ~~sets~~ records are accompanied by LHF-related uncertainty estimates, which hampers a quality assessment of the air–sea fluxes and related parameters.

Such uncertainty assessments ~~would~~ go beyond conventional LHF intercomparison studies~~(as e.g. presented Chou et al., 2004; Yu et al. they would~~ , as they allow for quantifying the data's accuracy (systematic uncertainty)and precision (random uncertainty). Consistency among two data sets would for example be achieved when independent measurements agree within their individual

uncertainties, as Immler et al. (2010) formulates the benefit of such an approach. Assimilation schemes like GECCO require such uncertainty information prior to assimilating respective fields in ocean models.

Few studies have taken on the challenge of uncertainty assessments in context of LHF-related climatologies. Whereas random uncertainties of ship-based LHF-related parameters are for example discussed in Gleckler and Weare (1997) , Kent and Berry (200 and Kent and Taylor (2006) , systematic uncertainties are assessed in e.g. Kent et al. (1993) and Kent and Taylor (1996) . An example of an *in situ* LHF climatology incorporating uncertainty estimates (based on optimal interpolation) is given by NOCS v2.0 (Berry and Kent, 2009) . A satellite-related uncertainty assessment is published by Brunke et al. (2011), who decomposed overall biases with respect to direct *in situ* records into a bulk variable and a residual component, the latter which also includes the measurement uncertainty. ~~A current overview study by Loew et al. (2017) highlights the necessity of earth observation data validation and pools different approaches across communities. Finally, assimilation schemes like GECCO require such uncertainty information prior to assimilating respective fields in ocean models~~Recently, Kinzel et al. (2016) presented an elegant approach for decomposing random uncertainties inherent to independent data sets using triple collocation. Apart from NOCS v2.0, none of the remaining LHF-related climatologies, irrespective of their data source, include comprehensive uncertainty information appended to the data.

In the framework of the German ~~research~~ Research Foundation (DFG) initiatives 'FOR1740' and 'FOR21740' ('Atlantic Freshwater Cycle', *http://for1740.zmaw.de/*), the lack of uncertainty information inherent to satellite data is overcome by ~~taking on the challenge of quantifying~~ specifying systematic, random, and sampling uncertainties ~~inherent to~~ exclusively associated with HOAPS-3.3 LHF-related parameters. This manuscript not only introduces the methodology, but also demonstrates its application to arrive at HOAPS-3.3 LHF-related uncertainty estimates.

Once the applied data sources have been described in more detail (Sect. 2), double collocation analysis is performed (Sect. 3.1). Respective matchups serve as input to multi-dimensional bias analyses (Sect. 3.2), which result in estimates of instantaneous systematic and total random uncertainty. Finally, random uncertainty decomposition (Kinzel et al., 2016) isolates the required retrieval-related contribution from collocation and *in situ* measurement contributions (Sect. 3.3). Rigorous error propagation to the instantaneous LHF-related data is performed subsequently, which accounts for how uncertainties in the bulk parameters propagate into uncertainties of LHF themselves ~~, while accounting for covariances of the contributing parameters.~~

~~Section 2 presents the applied datasources in more detail~~(Sect. 3.4). The described sequence allows for assigning HOAPS-3.3 related systematic and random uncertainty estimates to the pixel-level data, which is not available for any other satellite data record to date. As to ~~systematic uncertainty patterns~~monthly mean sampling uncertainties (Sect. 3.5), the approach ~~of double collocation and multi-dimensional bias analyses is introduced in Sect. 3. This is complemented by the strategy of decomposing random uncertainties via multiple triple collocation to separate the eligible random satellite retrieval uncertainty from collocation and *in situ* measurement contributions.~~ by Tomita and Kubota (2011) is employed. All uncertainty components are presented in Sect. 4, ~~where seasonal and regional differentiations allow for assessing the uncertainty spectrum. This is supplemented by trend analysis in light of~~ which includes regional and seasonal differentiations. Section 4 also comprises a trend analysis applying the derived uncertainty estimates. ~~Section 5 provides a~~ A summary and a brief outlook regarding

ongoing work . are provided in Sect. 5. The introduced methods can easily be transferred to other retrievals, highlighting the value of this study.

2 Data

2.1 HOAPS-3.3 Pixel-Level Data Records

Apart from the sea surface temperature (SST), all HOAPS parameters are derived from intercalibrated Special Sensor Microwave/Imager (SSM/I) and Special Sensor Microwave Imager/Sounder (SSMIS) passive microwave radiometers, which are installed aboard the polar orbiting satellites of the United States Air Force Defense Meteorological Satellite Program (DMSP). HOAPS provides consistently derived global fields of freshwater flux related parameters. Regarding sensor specifications and orbital paths, the reader is referred to e.g. Andersson et al. (2010).

Here, the focus lies on the HOAPS-3.3pixel-level, which has been produced as an extension to the HOAPS-3.2 data set (Andersson et al., 2010; Fennig et al., 2012) in the framework of the ongoing DFG research activity. It Its extensive documentation is available online (Fennig et al., 2013) . HOAPS-3.3 covers the time period from 1987 to 2015, during which a total number of nine satellite instruments were in operational mode . (F8-F18). The spatial resolution of the pixel-level data is channel-dependent. For SSM/I, it varies from 69 km by 43 km (19 GHz channel) to 37 km by 28 km (37 GHz). Likewise, it ranges from 74 km by 47 km (19 GHz channel) to 41 km by 31 km (37 GHz) for SSMIS sensors. Compared to HOAPS-3.2, HOAPS-3.3 has been temporally extended up to 2015 and is based on a pre-release of the CM SAF SSM/I and SSMIS FCDR. This reprocessing included a homogenization of the radiance time series by means of an improved inter-sensor calibration with respect to the DMSP F11 instrument. Earth incidence angle normalization corrections were applied, following a method described by Fuhrhop and Simmer (1996). Its extensive documentation is available online (Fennig et al., 2013) . Since the HOAPS-3.1 release, HOAPS is hosted by the EUMETSAT Satellite Application Facility on Climate Monitoring (CM SAF), whereupon its further development is shared with the University of Hamburg and the Max Planck Institute for Meteorology (Hamburg). In this study, the pixel-level HOAPS-3.3 data in sensor resolution is used, which implies that no aggregation for gridding purposes has been applied.

HOAPS-3.3 q_a relies on a direct, four-channel retrieval algorithm by Bentamy et al. (2003), which is based on a modified version of the two-step multi-channel regression model by Schulz et al. (1993) and its refinement by Schlüssel (1996). 1000 globally collocated pairs of SSM/I TBs brightness temperatures (TBs) and ship data between 1996-1997 1996-97 were used to estimate the new values for the coefficients in the Schulz model.

To account for the non-linearity of the problem, the HOAPS-3.3 U algorithm uses a neural network approach with three layers after Krasnopolsky et al. (1995) to derive the wind speed at 10 m above sea level (a.s.l.). The network was trained with a composite data set of buoy measurements, which was compiled using matchups of SSM/I F11 brightness temperatures (TBs) TBs and near-surface wind speed measurements from the National Oceanographic and Atmospheric Administration (NOAA)

National Data Buoy Center (NDBC) and the Tropical Atmosphere Ocean (TAO) array between 1997-98. Radiative transfer simulations based on radiosonde profiles served as input for the training data set (Andersson et al., 2010).

HOAPS-3.3 SST is based on the AVHRR Pathfinder Version 5.2 and is obtained from the US National Oceanographic Data Center and the Group for High Resolution Sea Surface Temperature (http://pathfinder.nodc.noaa.gov). The data ~~are~~ is an updated version of the Pathfinder Version 5.0 and 5.1 collection described in Casey et al. (2010). A static bias correction of +0.17 K has been applied to HOAPS-3.3 SST data in order to revert the Pathfinder Version 5.2 skin correction and thus achieve consistency with Version 5.0 used in HOAPS-3.2.

HOAPS-3.3 sea surface saturation specific humidity q_s is derived by applying the Magnus formula (Murray, 1967) to SST, while accounting for a constant salinity correction factor of 0.98. ~~Zeng et al. (1998) , e.g., showed that omitting the factor under strong wind conditions has a significant impact on resulting LHF.~~

HOAPS-3.3 LHF is based on the Coupled Ocean–Atmosphere Response Experiment (COARE) 2.6a bulk flux algorithm. With minor modifications of physics and parameterizations, the algorithm is published as ~~COARE 3.0~~COARE-3.0a by Fairall et al. (2003). ~~U, q_s, and q_a are required~~ It includes atmospheric stability calculations, which necessitate surface air temperatures as input. ~~The latter depends on the surface air temperature, which is~~ These are estimated by assuming a constant relative humidity of 80 % (Liu et al., 1994) and air-sea temperature difference of 1 K (Wells and King-Hele, 1990). A constant sea surface pressure of 1013.25 hPa is prescribed within the bulk flux algorithm. COARE-3.0 is widely accepted within the scientific community; its benefits are for example presented in the framework of an intercomparison study by Brunke et al. (2003).

2.2 DWD-ICOADS Data Archive

Hourly *in situ* measurements of U, q_s, and q_a (bulk parameters, as of now) have been provided by the Marine Climate Data Center of the German Meteorological Service (DWD), supervised by the Marine Meteorological Office (Seewetteramt, SWA). While data prior to 1995 is excluded due to a comparatively poor *in situ* data coverage, the data set used here includes measurements up to 2008. It comprises global high-quality shipborne measurements as well as data provided by drifting and moored buoys. In case of data gaps within the SWA archive, the *in situ* data basis was extended at SWA by available International Comprehensive Ocean–Atmosphere Data Set (ICOADS) measurements (Version 2.5, Woodruff et al., 2011). ~~These~~ A comprehensive literature overview on research applications involving ICOADS data is given by Freeman et al. (2016) . Both SWA and ICOADS records contain hourly global measurements obtained from ships, moored and drifting buoys as well as near-surface measurements of oceanographic profiles. Several quality checks were performed at SWA prior to using the merged DWD-ICOADS data, which resulted in quality index assignments to each observation. Details regarding the flagging procedures carried out at SWA are given in Kinzel et al. (2016).

In preparation for the uncertainty analyses, further filtering and correcting procedures to both ship and buoy data were carried out. Regarding ship records, annual lists of Voluntary Observing Ships (VOS) metadata (Kent et al., 2007) were employed. Most of the supplementary buoy metadata was extracted from the Data Buoy Cooperation Panel, which particularly includes a fleet of moored buoy arrays operated by NDBC. Metadata of the Global Tropical Moored Buoy Array, such as TAO-TRITON

(Pacific-), PIRATA (Atlantic-), and RAMA (Indian Ocean) were obtained from the Pacific Marine Environment Laboratory (PMEL).

ICOADS VOS estimates of q_a are based on wet bulb temperature measurements, typically using mercury thermometers, which are often exposed in either (ventilated) screens or sling psychrometers (Kent et al., 2007). q_a is eventually derived by applying the psychrometric formula. By contrast, q_a estimates of buoys originate from measurements of air temperature and relative humidity. For this study, q_a of both VOS and buoys were not corrected to the HOAPS-3.3 reference of 10 m a.s.l., assuming neutral stratification. A discussion related to this approach is published in Kinzel et al. (2016). It is in line with Prytherch et al. (2014), who conclude that a conversion to 10 a.s.l. (neutral stability) substantially adds to the noise in the resulting *in situ* q_a. The aspect of correcting q_a with respect to height and stratification is also elucidated in Bentamy et al. (2003) and Bentamy et al. (2013) , whereas correction effects are presented in Kent et al. (2014). The authors for example quantify the height correction effect due to continuously increasing measurement platform heights between 1971-2006 to be 0.11 g kg^{-1}. However, this effect is masked by bias corrections associated with measurement techniques, which are thought to be 2-3 times larger.

DWD-ICOADS VOS U are either measured using anemometers (likewise for buoys) or are estimated from the sea state, depending on the preference of the country recruiting the VOS (Kent et al., 2007). By means of the measured wind speed and direction, the true wind speeds are derived considering the ship's speed and direction. If a specific anemometer height was not given, it was estimated from the annual global mean height difference with respect to the thermometer platform. For each year, this single height difference value is based on all contributing ship records with complete metadata information. Prior to 2002, no thermometer heights were available; consequently, the height difference was set to 6 m (average between 2002-2008). In case both sensor heights were unknown, the linear fits shown in Table 4 of Kent et al. (2007) were used to derive anemometer heights based on available ship length metadata. It was assumed that these ship type dependent linear fits (Kent et al., 2007, their Fig. 11) introduce negligible uncertainties to the sensor height derivation. Given the anemometer heights of both VOS and buoys, *in situ* wind speeds were corrected to the HOAPS-3.3 standard height of 10 m a.s.l. to remove inhomogeneities, using the iterative equivalent neutral stability approach of Fairall et al. (2003). With the exception of e.g. (stable stratified) upwelling regimes or local instabilities, the equivalent neutral stability assumption is valid over vast regions of the open oceans. The correction using a neutral wind equivalent profile has been suggested by e.g. Shearman and Zelenko (1989). It is argued that in case of VOS, the omission of a correction would lead to a positive wind speed bias, as the average wind sensor height is given by 18 m (Kent et al., 2014). By contrast, buoy U would be low-biased.

VOS SST measurement techniques differ in terms of platform, measurement depth, and extent of automation. Strictly speaking, *in situ* SST are sub-surface temperatures and thus differ from the HOAPS-3.3 Pathfinder SST, which are treated as a skin SST for the surface flux calculations. This necessitates an *in situ* cool-skin correction as a function of wind speed, following Donlon et al. (2002). Their Equation (2) was applied, omitting all records subject to wind speeds below 2 m s^{-1} (corrected to 10 m a.s.l.), as the exponential fit introduces additional uncertainty for very calm conditions. On average, the SST correction reduced the DWD-ICOADS SST by approximately 0.17 K. Moreover, the warm layer part of the COARE 3.0 algorithm is not implemented in HOAPS-3.3 due to the lack of a continuous diurnal cycle information on the surface radiation budget from the

SSM/I and SSMIS measurements. To be directly comparable to the *in situ* counterpart, all *in situ* measurements taken during local daytime were excluded. As only night-time *in situ* measurements during non-calm conditions were considered, the sea water temperature gradient within the uppermost meters of the water column is thought to be negligible. A SST correction with respect to the sensor depths was therefore omitted for both VOS and buoys, independent of the measurement platform.

All VOS data processing described above ~~were~~ was carried out for research vessels (so-called 'special ships') and merchant vessels only due to vast data amounts and in order to minimize *in situ* uncertainties. In case of multiple triple collocation analysis (Sect. 3.3), buoy records were excluded to ensure having a consistent, globally distributed data set as the ground reference for the random decomposition procedure. It is argued that the vast amount of remaining triplets authorizes this restriction.

Despite strict filtering and correcting procedures, *in situ* measurement uncertainties related to sensor types, measurement heights and positions, and solar radiation contamination may remain (e.g. Bourassa et al., 2013). Assessments regarding the quality of the reference data are beyond the scope of this article. The *in situ* data basis is therefore considered as the bias-free, ground reference. This assumption is in line with calibration and validation approaches of Bentamy et al. (2003) , Jackson et al. (2009) , and Bentamy et al. (2013) , amongst others. As will be shown in Sect. 3.2, the HOAPS systematic uncertainties presented in this work are interpreted as upper limit estimates. Therefore, the assumption of a bias free ground reference does not violate our main conclusions, although a small contribution to the systematic uncertainties may be caused by the in situ reference.

3 Methodology

This Section describes the technical background for deriving systematic, random, and sampling uncertainties inherent to HOAPS-3.3. By performing double collocation analysis (Sect. 3.1), multi-dimensional bias analysis (Sect. 3.2), and random uncertainty decomposition (Sect. 3.3), pixel-level data will be equipped with both systematic and random uncertainty estimates. When averaging in time, sampling uncertainties are also accounted for (Sect. 3.5). The uncertainties will be examined from an either instantaneous ~~or climatological~~ (i.e., pixel-level) or averaged (i.e., monthly or multi-annual mean) point of view, depending on the ~~scale of interest and thus the application . The random uncertainty decomposition presented in Kinzel et al. (2016) is therefore complemented, leading to~~ application (Sect. 4). The longer the averaging time period, the lesser the impact of both random and sampling uncertainties. This implies that on climatological scales, total uncertainties hardly differ from systematic uncertainties. The sequence of analyses allows for a complete HOAPS-3.3 uncertainty characterization of LHF-related parameters on all time scales, which complements the random uncertainty decomposition of q_a presented in Kinzel et al. (2016) .

3.1 Double Collocation Analysis

In preparation for uncertainty calculations, a double collocation analysis is performed for the time period of 2001–2008, resulting in paired matchups of LHF-related HOAPS-3.3 and *in situ* data. Although HOAPS-3.3 lasts until 2015, collocations between 2009–2015 were not performed, as the DWD-ICOADS data archive only lasts until 2008. The collocated pairs are

based on the so-called nearest neighbor approach; that is, HOAPS-3.3 pixels are assigned to respective *in situ* observations closest in time and space. Parameter-independent collocation criteria of Δx = 50 km and Δt = 60 min are chosen. These are more restrictive than those derived in e.g. Kinzel (2013). Due to the vast amount of available matchups this is justifiable and ensures that e.g. strong spatial and/or temporal gradients associated with fronts are discarded from further analysis.

Figure 1 (left) presents the resulting collocation density for 2001-2008, exemplarily for q_a. Matchups mainly occur in coastal regions (associated with buoys) and along major shipping lanes. By contrast, the Southern Ocean considerably lacks high-quality in situ measurements. The amount of U and q_s collocations exceeds those shown in Fig. 1 (left). For brevity, their distributions are not shown.

Figure 2a-d exemplarily shows scatter density plots of the q_a bias (2001-2008) as a function of the atmospheric state parameters q_a ("hair"), U ("wind"), SST ("asst"), and ~~water vapour path~~ vertically integrated water vapour ("wvpa"), resulting from the double collocation analyses. Overall, 13.8 million matchups contribute to each subplot. ~~Figure 2a indicates that HOAPS-3.3 overestimates near-surface specific humidities for~~ The illustrated bins are not equidistant; in fact, their width depends on the data density of the matchups. This implies that 5% of all collocated pairs are assigned to a single bin, respectively. Analogously to Fig. 2, one-dimensional bias analyses are performed for both dU and dq_s (not shown).

For q_a ~~between 7–12~~ values between 7-12 g kg^{-1} ~~and~~ , HOAPS-3.3 overestimates near-surface specific humidities (see Fig. 2a). Overestimations are also observed in the inner tropics~~(~~, where q_a ~~≈~~ is in the order of 20 g kg^{-1}~~)~~. In return, biases are negative ~~over Arctic~~ for polar (< 5 g kg^{-1}) and subtropical (12–17 g kg^{-1}) humidity regimes. The latter ~~regions~~ region is also subject to largest random uncertainties, which exceed 2 g kg^{-1}. See Kinzel et al. (2016) and Prytherch et al. (2014) for more details on the analysis of HOAPS-3.3 q_a and its resemblance to GSSTF3 q_a (Shie et al., 2012) . The spatial distribution of these q_a biases are shown in Fig. 1 (right). ~~The humidity bias and standard deviation dependency on SST (Fig. 2c)shows similar features regarding regimes of over- and underestimation.~~ Specifically the underestimations (overestimations) over subtropical (tropical) oceans are well resolved. Humidity biases as a function of wind speed are illustrated in Fig. 2b. The distribution is somewhat linear, where low (high) wind regimes are over-(under-)~~represented~~ estimated in HOAPS-3.3. In contrast to the remaining atmospheric state parameters, the random uncertainty decreases fairly linearly with increasing wind speeds. The q_a bias distribution as a function of SST (Fig. 2c) resembles that of the q_a-dependent distribution (Fig. 2a) regarding regimes of over- and underestimation. A dependency of dq_a on the total integrated water vapour (Fig. 2d) shows only few distinct features. Most matchups coincide with values below 20 kg m^{-2}. With the exception of smallest values, these result in positive biases with respect to HOAPS-3.3. As the abscissa and ordinate variables in Fig. 2 are correlated, we investigated the contribution of artificial biases by illustrating dq_a as a function of in situ q_a, U, und SST. Results indicate that the percental difference of the mean bin values (black squares) of HOAPS and DWD-ICOADS range between 6-10% (not shown). We are therefore confident that our approach is robust. Two-sided regression analyses could further reduce these spurious biases.

A comparison of e.g. Fig. 2a and b indicates that the simple one-dimensional bias analyses may be misleading when it comes to HOAPS-3.3 q_a-related uncertainty characterizations. Average q_a off the Arabian Peninsula, for example, are in the order of 14–15 g kg^{-1} (not shown). According to Fig. 2a, this is associated with a HOAPS-3.3 q_a underestimation~~.~~ , as is also seen in Fig. 1 (right). At the same time, climatological mean wind speeds are as low as 3–5 m s^{-1} (not shown), which goes along with a

HOAPS-3.3 q_a overestimation (Fig. 2b). This is no contradiction, but rather indicates that the HOAPS-3.3 q_a retrieval seems to encounter challenges for specific humidity and wind regimes. Furthermore, a constraint to one-dimensional analyses implies for example that parts of the random uncertainties illustrated in Fig. 2a (bars) receive a systematic component in Fig. 2b (squares). This conclusion motivates to proceed with ~~multi-dimensional~~ *multi-dimensional* bias analyses, where all possible atmospheric states, i.e. combinations of the four chosen atmospheric state parameters, are accounted for simultaneously. This approach finally allows for separating systematic from random uncertainties. Results illustrated in Fig. 2 can therefore be considered as a preliminary stage of the four-dimensional bias analyses introduced in Sect. 3.2, where each of the four atmospheric state variables (~~i.e.,~~ Fig. 2, x-axes) represent one dimension. ~~Analogously to Fig. 2, one-dimensional analyses are performed for both dU and dq_s (not shown).~~

3.2 Multi-Dimensional Bias Analyses

The bulk formula for LHF is given by

$$LHF = \rho_a L_V C_E U (q_s - q_a), \tag{1}$$

where ρ_a is the density of moist air and L_V the latent heat of vaporization. ρ_a is derived as a function of HOAPS-3.3 q_a and near-surface air temperature. Likewise, L_V is computed simultaneously as a function of HOAPS-3.3 SST.

Assuming uncertainties in ρ_a and L_V to be negligible and according to standard error propagation, the overall LHF uncertainty is a function of the systematic and random uncertainties introduced by the remaining parameters.

As to the Dalton Number C_E, the estimates of Fairall et al. (2003) are applied by assigning 5 % (10 %) of systematic uncertainty of C_E for wind speeds smaller (larger) than 10 m s^{-1}. For wind speeds exceeding 20 m s^{-1}, the estimate of Gleckler and Weare (1997) of 12 % is taken on. Independently of U, random uncertainties of 20 % are assigned, as proposed by Gleckler and Weare (1997) .

In case of U, q_s, and q_a ~~these~~ , the uncertainties are assumed to depend on the concurrent atmospheric state. The combination of q_a, U, SST, and ~~water vapour path~~ vertically integrated water vapour is thought to represent the concurrent atmospheric state best. Therefore, the one-dimensional consideration presented in Sect. 3.1 is expanded by creating four-dimensional look up tables (LUTs) including 20^4 entries, respectively. The dimension is reflected in the exponent, whereas its base represents the amount of bins per dimension. As described in Sect. 3.1, these bins are not equidistant. In case of dq_a, bin means of each of the four dimensions are indicated by the x-values of the black squares shown in Fig. 2a-d, respectively. The values of all four dimensional vectors are essential for assigning instantaneous, absolute differences (HOAPS-3.3 minus *in situ*) to the correct LUT~~bin and are predetermined by the respective x-values of the black squares shown in Fig. 2a-d~~. By averaging the content of each bin, systematic and *total* random uncertainties finally result as a function of the four atmospheric state parameters. The approach of processing absolute measures of the observed differences allows for moving from a simple bias analysis to an uncertainty characterization. The resulting systematic uncertainties, which are shown throughout Sect. 4, can therefore be treated as an upper boundary of a more simple bias distribution.

The ~~uncertainty dependency on specific ambient conditions~~ multi-dimensional uncertainty characterization approach overcomes the issues introduced by data-sparse regions, such as the Southern Ocean and the tropical oceans (e.g. Kent and Berry, 2005). Here, it is knowingly turned away from the dependency on matchup density, which implies that the LUTs are valid on a global scale. Due to the immense data availability, their pairwise input biases are confined to matchups from 2001–2008 (dq_a, dU) and 1998–2001/2006–2008 (dq_s). A thorough elucidation of the multi-dimensional bias analysis is presented in Kinzel et al. (2016), exemplarily for HOAPS-3.2 q_a (Sect. 2c and Fig. 5, left therein). Here, it is applied to all three bulk parameters, which results in both systematic and *total* random uncertainty LUTs.

~~As to the Dalton Number C_E, the estimates of Fairall et al. (2003) are applied by assigning 5 (10) of systematic uncertainty of C_E for wind speeds smaller (larger) than 10 m s^{-1}. For even stronger wind speeds, the estimate of Gleckler and Weare (1997) of 12 is taken on. Independently of U, random uncertainties of 20 are assigned, as proposed by Gleckler and Weare (1997) .~~

3.3 Random Uncertainty Decomposition

~~Recall that the aim is to characterize *uncertainty* and not bias patterns. This implies that *absolute* systematic uncertainty values are generally presented, i.e., magnitudes are invariably positive. Results presented in Sect. 4.4-4.6 can therefore be considered as illustrating the upper boundaries of systematic (that is, climatological) uncertainties.~~

3.4 ~~HOAPS-3.3 Uncertainty of LHF~~

~~The uncertainties in LHF are caused by uncertainties in all bulk input parameters contributing to Eq. . Assuming the underlying parameterizations to be correct, LHF uncertainties can thus be derived by carrying out standard error propagation. These uncertainty estimates are assigned at each point in time and space.~~

~~Total instantaneous LHF uncertainties, σ_{LHF}, are derived as follows:~~

$$\sigma_{LHF} = \sqrt{\left(\frac{\partial LHF}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial LHF}{\partial y}\right)^2 \sigma_y^2 + 2r_{xy}\left(\frac{\partial LHF}{\partial x}\frac{\partial LHF}{\partial y}\right)\sigma_x\sigma_y},$$

~~where x and y are place holders of U, q_s, q_a, and C_E. r_{xy} is the correlation coefficient between x and y. For each combination of x and y, the average of daily global mean correlation coefficients between 1995 and 2008 is applied.~~

~~σ_x and σ_y are~~ The *total* ~~uncertainties in x~~ random uncertainties introduced in Sect. 3.2 (and ~~y. These can be decomposed into systematic and random components:~~

$$\left(\frac{\partial LHF}{\partial x}\right)^2 \sigma_x^2 \widehat{=} \left(\frac{\partial LHF}{\partial x}\right)^2 \sigma_{x,sys}^2 + \left(\frac{\partial LHF}{\partial x}\right)^2 \sigma_{x,ran}^2 \left(N^{-1/2}\right)^2.$$

~~N is the number of HOAPS-3.3 satellite observations (N=1 for instantaneous LHF uncertainties). Note that in case of gridded uncertainty products, the random component becomes negligibly small, given long temporal and large spatial averages. Sampling uncertainties do not exist on an instantaneous basis and are therefore not considered in Eqs. -.~~

~~Recall that $\sigma_{x,ran}$ in Eq. represents the *overall* random uncertainty of x. To isolate its random satellite retrieval component (E_{retr}^{ran}), a random uncertainty decomposition is carried out, which is briefly reviewed in the following.~~

3.4 ~~HOAPS-3.3 Random Uncertainty Decomposition~~

~~This section briefly summarizes the concept of random uncertainty decomposition. For more mathematical and technical details, the reader is referred to Kinzel et al. (2016) .~~

~~Next to E_{retr}^{ran}, random uncertainty estimates resulting from collocations (e.g.~~ also those represented by the black error bars in Fig. 2) include random uncertainties associated with the collocation procedure (E_C) and *in situ* measurement noise (E_{ins}) (e.g. Bourras, 2006). To isolate the random retrieval uncertainty, E_{retr}^{ran}, which is exclusively HOAPS-related, multiple triple collocation (MTC) analysis is applied to matchups of U, q_s, and q_a for the time period 1995–2008. This section briefly summarizes the concept of random uncertainty decomposition. For more mathematical and technical details, the reader is referred to Kinzel et al. (2016) . MTC analysis includes a twofold triple collocation (TC, introduced by Stoffelen, 1998) , whereupon double collocated data described in Sect. 3.1 serves as input. Triplets incorporating two independent *in situ* measurements and one HOAPS-3.3 pixel represent the first arrangement, whereas a single *in situ* record and two HOAPS-3.3 pixels of independent satellite instruments form the second triplet structure (see Fig. 1 in Kinzel et al. (2016)). The collocation criteria applied in Sect. 3.1 are adopted ~~. Data~~ and data poleward of 60° N/S is excluded to avoid biases associated with sea ice effects.

Subsequent to a bias correction with respect to the *in situ* measurements, the variances of differences between two independent data sources X and Y, that is V_{XY}, are calculated following O'Carroll et al. (2008). Given three data sources and two types of TCs, this results in six combinations of V_{XY}. Next, error models for both ship and satellite records are defined. In case of ship records, these include E_{ins}, whereas for satellite records, they incorporate satellite sensor noise (E_N, synthetically derived) and retrieval model uncertainty (E_M). Applying these error models to the derived V_{XY} results in six equations incorporating E_{ins}, E_M, E_N, and E_C. These equations are successively solved for all random uncertainty sources as a function of ~~the respective bulk parameter.~~ U, q_s, and q_a, that is for 20 individual bins per parameter. Each of these bins include thousands of triple collocated matchups. Finally, $E_{retr}^{ran} = \sqrt{(E_M)^2 + (E_N)^2}$ is the ~~pursued~~ required random satellite retrieval uncertainty, which is derived for all 20 bins as a function of U, q_s, and q_a.

~~Thus,~~ MTC is a powerful tool to decompose *total* random uncertainties (i.e., E_{sum} = E_{retr}^{ran} + E_{ins} + E_C) inherent to LHF-related bulk parameters in order to isolate the random retrieval contribution E_{retr}^{ran}. Depending on the magnitude of the respective bulk parameter, the fractional contribution of E_{retr}^{ran} to E_{sum} is finally derived. That is, each entry of the total random uncertainty LUTs introduced in Sect. 3.2 is 'adjusted'. ~~Table 1~~ Section 4.1 presents a statistical summary of the instantaneous, decomposed random uncertainties inherent to U, q_s, and q_a.

3.4 Deriving HOAPS-3.3 LHF-Related Uncertainties

The uncertainties in LHF are caused by uncertainties in all bulk input parameters contributing to Eq. (1). Assuming the underlying parameterizations to be correct, LHF uncertainties can thus be derived by carrying out standard error propagation. These uncertainty estimates are assigned to each HOAPS pixel, depending on the four atmospheric state parameters.

Total instantaneous LHF uncertainties, σ_{LHF}, are derived as follows:

$$\sigma_{LHF} = \sqrt{\left(\frac{\partial LHF}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial LHF}{\partial y}\right)^2 \sigma_y^2 + 2 r_{xy} \left(\frac{\partial LHF}{\partial x}\frac{\partial LHF}{\partial y}\right) \sigma_x \sigma_y}, \quad (2)$$

where x and y are place holders of U, q_s, q_a, and C_E. r_{xy} is the correlation coefficient between x and y. For each combination of x and y, the average of daily global mean correlation coefficients between 1995 and 2008 is applied. Global mean coefficients are preferential compared to instantaneous r_{xy} for two reasons. First, the amount of instantaneous data for a specific region is limited, which may distort the results of the correlation analysis. Second, omitting all correlation-related terms in Eq. (2) modifies $\sigma_{LHF,sys}$ by merely 0.5 $\pm$5 W m^{-2} (not shown), which indicates that these terms do not receive much weight after all.

σ_x and σ_y are *total* uncertainties in x and y. These can be decomposed into systematic and random components. Note that the random component has been corrected for collocation and in situ uncertainty effects (see Sect. 3.3) and already represents the random retrieval uncertainty E_{retr}^{ran}.

$$\left(\frac{\partial LHF}{\partial x}\right)^2 \sigma_x^2 \hat{=} \left(\frac{\partial LHF}{\partial x}\right)^2 \sigma_{x,sys}^2 + \left(\frac{\partial LHF}{\partial x}\right)^2 \sigma_{x,retr,ran}^2 \left(N^{-1/2}\right)^2. \quad (3)$$

N is the number of HOAPS-3.3 satellite observations (N=1 for instantaneous LHF uncertainties). In case of temporal and spatial averaging over a sufficiently long time period, the random component becomes negligibly small. Sampling uncertainties do not exist on an instantaneous basis and are therefore not considered in Eqs. (2)-(3).

3.5 Sampling Uncertainty

~~Next~~ In addition to systematic and random uncertainties, inhomogeneous sampling may occur, specifically when temporal ~~and/or spatial~~ resolution in observations are coarse. As remotely sensed data is measured at selected times only, ~~spatial and~~ temporal sampling uncertainties therefore become an issue (Gulev et al., 2010), as the diurnal cycle may not be captured correctly.

~~In a first step, daily~~ Daily mean sampling uncertainties of HOAPS-3.3 LHF-related parameters are derived, using high-resolution buoy measurements. Overall, data of eight tropical (PMEL, hourly resolution) and 15 extratropical (NDBC, 10-minute resolution) moored buoys account for a possible climate regime dependency. All chosen buoy records comprise several years of data (1995-2008) and hardly show temporal data gaps. Here, the approach by Tomita and Kubota (2011) is followed to derive the sampling uncertainties by simulating two satellite data overpasses ~~based on the buoy records~~per day, using the buoy values. In case of U and SST, records are corrected for sensor heights and cool skin effects, respectively, as explained

in Sect. 2.2. *In situ* LHF are computed by means of the COARE-2.6a algorithm (Fairall et al., 2003). Daily means of 'true' buoy data are derived by averaging all daily buoy records, where only high-quality data (indicated by quality flags 1–2) is considered. The weighted average of the two closest (in time) 'true' buoy observations to local satellite overpasses corresponds to the so-called 'simulated' satellite data record (Tomita and Kubota, 2011, their Fig. 2). All daily sampling uncertainties are derived as a function of the number of simultaneously operating SSM/I instruments. These daily values form the basis for the monthly averages of selected parameters (E_{smp}), which are outlined in Table 2 (Sect. 4.4). The estimates are global means; an earlier, regime-dependent investigation resulted in negligible differences~~between the resulting sampling uncertainties~~. This implies that monthly mean systematic uncertainties do not exhibit a latitudinal dependency.

4 Results and Discussion

4.1 Magnitudes of HOAPS-3.3 Decomposed Random Uncertainties

Table 1 presents a statistical summary of the instantaneous random uncertainty decomposition for the bulk parameters U, q_s, and q_a, following the approaches described in Sect. ~~3.2 and~~ 3.1 to 3.3. Note that E_N is not included, as its synthetically derived value remains constant throughout the respective parameter range (for procedure, see Kinzel et al., 2016)~~remains constant throughout the respective parameter range~~. Asterisked values indicate global mean weighted averages and pooled variances of Kent and Berry (2005), resulting from a semivariogram approach. These are based on their Fig. 1, taking the illustrated grid averaged random uncertainties, the standard deviation as well as the number of observations into account. In the following, individual contributions to the overall random uncertainties are discussed, but not shown in terms of supplementary figures.

$E_{retr}^{ran}(q_a)$ ranges between ~~0.3~~ 0.7 and 1.8 g kg^{-1}, where minima (maxima) are found ~~in Arctic (subtropical~~below 5 g kg^{-1} (between 13-17 g kg^{-1}) q_a regimes. Whereas largest relative uncertainties are associated with polar q_a values (3-5 g kg^{-1}), lowest relative contributions below 10 % are confined to the inner tropics (20 g kg^{-1}). On average, both $E_c(q_a)$ and $E_{ins}(q_a)$ are approximately half the size of $E_{retr}^{ran}(q_a)$. The average of $E_{ins}(q_a)$ is 0.4 g kg^{-1} below the mean estimate of Kent and Berry (2005). It is hypothesized that the lower estimate of $E_{ins}(q_a)$ is a direct consequence of the rigorous *in situ* filtering procedure prior to MTC analysis. The difference may furthermore be triggered by the fact that Kent and Berry (2005) include data records dating back to the 1970s and 1980s, which may imply that ship records are included which do not fulfill the here applied quality control standards. In contrast to $E_{retr}^{ran}(q_a)$, $E_{ins}(q_a)$ increases rather linearly with q_a, which implies that smallest (largest) random *in situ* measurement uncertainties are found for lowest (highest) q_a. In contrast, $E_c(q_a)$ shows a similar distribution as $E_{retr}^{ran}(q_a)$, yet with considerably smaller amplitude. These random collocation uncertainties range between 0.4 and 0.7 g kg^{-1}, corresponding to 3–18 %. A graphical illustration of the q_a random uncertainty decomposition is shown in Kinzel et al. (2016) (their Fig. 2).

In case of U, all random uncertainties tend to be larger compared to q_a in a relative sense. In contrast to q_a, all three relative uncertainties exhibit a clear increase over large ranges of U, where minima and maxima in $E_{retr}^{ran}(U)$ ($E_{ins}(U)$, $E_c(U)$) range between 1.0–2.6 m s^{-1} (1.5–2.3 m s^{-1}, 0.8–2.0 m s^{-1}). Whereas $E_{retr}^{ran}(U)$ and $E_{ins}(U)$ are fairly constant for moderate wind speeds before continuously increasing, $E_c(U)$ seems to already saturate for mean wind speeds in the order of 10 m s^{-1} (not

shown). Similar to $E_{ins}(q_a)$, the $E_{ins}(U)$ estimate of Kent and Berry (2005) is roughly 40 % larger. Again, this difference is suspected to arise from the differences in the data set compositions. Kent and Berry (2005) furthermore elucidate that no corrections for height or adjustments to the Beaufort scale have been applied to their data, which would have caused a reduction in random uncertainty of 13 ± 1 %, according to the authors. Yet, $E_{ins}(U)$ almost exclusively represents the largest contribution to the random uncertainty budget of U. For all random uncertainty sources, strong wind regimes are linked to smallest relative uncertainties in the order of 12–15 %. In low-wind regimes, however, relative uncertainties exceed 50 % to even 100 %.

Both absolute and relative contributions of q_s-related random uncertainties remain well below those of q_a. Global mean values of all three random uncertainty sources are in the order of 0.5–0.6 g kg^{-1}. Regarding $E_{retr}^{ran}(q_s)$, this is comparable to the value published in e.g. McClain (1989), who estimated the global RMSE of AVHRR-derived SST to be in the order of 0.6–0.7 K ($\widehat{=}$ 0.4–0.5 g kg^{-1}). Similar to $E_{retr}^{ran}(U)$, $E_{retr}^{ran}(q_s)$ ($E_{ins}(q_s)$) shows a positive proportionality with largest values of 0.9 g kg^{-1} (1.5 g kg^{-1}). As for $E_{ins}(U)$, $E_{ins}(q_s)$ exceeds $E_{retr}^{ran}(q_s)$, specifically for q_s larger than 8 g kg^{-1}. In contrast to q_a, relative uncertainties are smallest in extratropical regimes with contributions of merely few percent. Largest relative uncertainties remain well below those of q_a and are in the order of 8–14 %.

4.2 Global Patterns of HOAPS-3.3 Random Retrieval Uncertainties

The results ~~shown~~ presented in Sect. ~~??~~ 4.1 are expanded by showing the global patterns of E_{retr}^{ran} in two-dimensional space.

Depending on the time period and thus on the number of SSM/I and SSMIS instruments in operation, the monthly global mean sum of instantaneous observations per 0.5°x0.5° grid cell ranges from approximately 90 (1988) to 650 (2006). In consequence, monthly means of E_{retr}^{ran} are considerably below the systematic counterpart (see scaling effect of N in Eq. (3)). Specifically from 1991 onwards, monthly globally averaged E_{retr}^{ran} of LHF-related parameters only reach 0.5–3 %. This reduction becomes even more striking when investigating multi-annual or even climatological means; LHF-related E_{retr}^{ran} virtually vanish on these scales. An increase (decrease) in these climatological random uncertainty values often directly results from a decrease (increase) in the number of pixel-level observations and thus not from a physical change due to shifts in the climate. This implies that results of trend analyses in random uncertainties, for example, may be misinterpreted. Therefore, the attention is drawn to the pixel-level (*instantaneous*) random uncertainty fields~~, which are subsequently related to the systematic counterpart in terms of distribution and magnitude~~. This instantaneous point of view causes their orders of magnitude to be similar to the results of E_{retr}^{ran} presented in Table 1. Note that the global averages shown in Fig. 3 in form of text strings are cosine-weighted, whereas the means illustrated in Table 1 do not take a regional dependency into account.

Figure 3 shows the instantaneous E_{retr}^{ran} patterns of HOAPS-3.3 LHF-related parameters between 1988 and 2012. ~~To a great extent, Fig. 3a can be interpreted as a two-dimensional representation of the error bar magnitudes~~ The magnitudes presented in Figure 3a are below those shown in Fig. 2a~~. Recall that~~ , as the random uncertainties ~~illustrated in Fig. 2a have not yet~~ have been corrected for the impact of $E_{ins}(q_a)$ and $E_c(q_a)$ (Sect. 3.3)~~, which is why their magnitudes exceed those shown in Fig. 3a.~~ . Maxima above 1.5 g kg^{-1} are located over all subtropical ocean basins, where q_a is in the order of 13–17 g kg^{-1}. A reduction within the inner tropics is clearly resolved, specifically over the warm pool region. $E_{retr}^{ran}(q_a)$ sharply decreases poleward to values of 0.6–0.9 g kg^{-1}. The global mean instantaneous $E_{retr}^{ran}(q_a)$ takes on a value of 1.2 g kg^{-1}.

The distribution of instantaneous $E_{retr}^{ran}(U)$ (Fig. 3b) shows a rather reversed pattern of q_a and closely resembles the climatological distribution of U itself. The global mean is given by 1.0 m s^{-1}. Global maxima cover large areas of the extratropical oceans, specifically over the Southern Ocean. Here, averages partly exceed 1.5 m s^{-1}. However, this results in less than 15 % retrieval uncertainty in a relative sense (not shown). In contrast, instantaneous $E_{retr}^{ran})(U)$ remain low (that is, below 0.8 m s^{-1}) over the (sub-) tropical ocean basins. This also applies to the warm pool area, which indicates a maximum in relative contribution close to 20 % due to climatological low wind speeds (not shown).

The pattern of instantaneous $E_{retr}^{ran}(q_s)$ (Fig. 3c) resembles that of q_a. However, the global mean magnitude of 0.3 g kg^{-1} represents ~~merely~~ only 25 % of the atmospheric counterpart. Absolute maxima in the order of 0.4 g kg^{-1} are located over the Indo-Pacific warm pool region, which stands in contrast to the local $E_{retr}^{ran}(q_a)$ minimum in that region~~for q_a~~. The comparatively small $E_{retr}^{ran}(q_s)$ also find expression in the low global mean relative uncertainty of 2 % (not shown). Values exceeding 4 % are confined to the extratropical ocean basins on both hemispheres.

Instantaneous $E_{retr}^{ran}(LHF)$ (Fig. 3d) show a strong proportionality to the climatological mean LHF pattern. In that respect, maxima are generally located over the subtropical central parts of all ocean basins (specifically the Indian Ocean) as well as along the western boundary currents. ~~Respective values partly exceed~~ In these areas, values are found in excess of 50 W m^{-2}. Apart from extratropical minima, low values in the tropics are confined to the eastern margins of the basins and the warm pool region.

Figure 3e shows the instantaneous random uncertainty of LHF relative to its natural variability. ~~This variability has been defined as the pixelwise~~ For each grid box, this variability is derived as the difference between the 5th and 95th percentile of instantaneous LHF observations between 2000–2008 ~~, based on the~~ (F13 platform only). Globally averaged, the relative random uncertainty equals to 17 %. Due to the large range of LHF along the western boundary currents (WBCs) and over the Central Indian Ocean, the absolute maxima seen in Fig. 3d are not resolved in Fig. 3e. Largest relative uncertainties exceeding 25 % are confined to the Southern Central Tropical Pacific and along the equatorial Atlantic.

4.3 Monthly Mean Sampling Uncertainties

~~Table 2 summarizes the monthly mean sampling uncertainties of several LHF-related~~ Global Patterns of HOAPS-3.3 ~~parameters as a function of concurrently operating SSM/I instruments. SST-related parameters show largest sampling uncertainties when three SSM/I instruments are simultaneously operating. This is not contradictory, as HOAPS-3.3 SST are AVHRR-based and thus not linked to the coverage of SSM/I instruments. From a climatological perspective, all magnitudes are negligibly small compared to respective systematic uncertainties. Regarding the main bulk parameters, orders of magnitude closely resemble those of monthly mean scaled E_{retr}^{ran}. It is concluded that their relative contribution to the monthly mean uncertainty budget is in the order of merely 1–2~~ However, one should keep in mind that sampling uncertainties become essential on considerably shorter time scales, i.e., in the framework of (sub-) daily analyses.

4.4 Global Patterns of HOAPS-3.3 Climatological Uncertainties

Figure 4 ~~a-e~~ shows the distribution of the climatological ~~*total*~~ uncertainties (E_{clim}) ~~between 1988 and 2012~~ for LHF and its related bulk parameters. E_{clim} is defined grid point wise as the mean root mean squared sum of instantaneous E_{sys}, E_{retr}^{ran} and E_{smp} between 1988-2012. As the contribution of E_{retr}^{ran} and ~~sampling uncertainties~~ E_{smp} converges towards 0% due to the vast number of observations, Figure 4a-e can also be treated as the systematic uncertainty distribution.

In an absolute sense, Fig. 4a mirrors the bias distribution shown in Fig. 2a. $E_{clim}(q_a)$ (Fig. 4a) generally range between 0.4–0.9 g kg^{-1}, where the global mean of 0.63 g kg^{-1} is approximately half the size of the instantaneous random counterpart shown in Fig. 3a. Maxima are found over the tropical central and western Pacific Ocean as well as the Caribbean and off the easternmost tip of South America. In the framework of a LHF intercomparison study, Smith et al. (2011) argue that satellite products have difficulties estimating q_a due to persistent stratus clouds, as observed west of Peru over the tropical eastern Pacific. This conclusion may be the cause for the elevated systematic uncertainties over the tropical eastern Pacific. In contrast, minima are located along both extratropical belts poleward of 50–60° N/S. ~~Secondly, isolated minima~~ Isolated minima also lie over the subtropical eastern margins of all ocean basins in the vicinity of 15–30° N/S, specifically over the Pacific basin. Interestingly, regions of comparatively low systematic uncertainties often coincide with regional maxima in random uncertainties (compare Fig. 3a). According to Fig. 2a, biases are smallest for climatological mean q_a of 4–5 g kg^{-1} and 13 g kg^{-1}, which fits well to the mentioned minima in Fig. 4a. Likewise, absolute bias maxima for q_a of 10 g kg^{-1} and 16–17 g kg^{-1} are resolved in both Fig. 2a and Fig. 4a.

The global mean of $E_{clim}(U)$ ~~is~~ shown in Fig. 4b ~~. Its global mean~~ equals to 0.81 m s^{-1}. On the one hand, maxima exceeding 1 m s^{-1} are located along the extratropical storm tracks, specifically over the northern hemisphere. On the other hand, local maxima are found along broad regions at 30° S and further equatorward over the Central Indian Ocean, off the Arabian Peninsula (both monsoon-related), and the central Northern Tropical Pacific. With the excepDIFDELCMDtion of the Southern Ocean, this is in line with Brunke et al. (2011), who conclude that reanalysis -, satellite -, and combined data sets tend to overestimate wind speeds ~~with respect to direct eddy covariance measurements~~compared to in situ records of inertial dissipation wind stresses, specifically over strong wind regimes. Monsoon-related characteristic features of Indian Ocean LHF variability, which also exhibit an impact on climatological uncertainties, are elucidated in e.g. Mohanty et al. (1996). Minima in the order of 0.5 m s^{-1} are mostly confined to the eastern margins of all ocean basins (Fig. 4b). The maxima over the northern hemispheric storm track are associated with climatological mean wind speeds of 9–11 m s^{-1}. This range also reveals largest positive biases in the one-dimensional bias consideration with respect to the *in situ* source (analogously to Fig. 2, but not shown for U). This also targets the maximum over the central Northern Tropical Pacific and all southern hemispheric maxima along 40–50° S. Although climatological mean wind speeds maximise over the Southern Ocean, respective systematic uncertainties rather show a slight poleward decrease. Again, this is in line with results from the one-dimensional dU analysis (not shown), which indicates that systematic uncertainties reduce for wind speeds above 12 m s^{-1}. Likewise, absolute bias minima are associated with low wind regimes in the order of 4–6 m s^{-1}. Climatologically lowest wind speeds of 2–4 m s^{-1} are for example found along

the Pacific coast of Central America (15° N), over the Arabian Sea, and over the Indo-Pacific warm pool region. HOAPS-3.3 tends to underestimate these wind speeds, as is mirrored in moderate $E_{clim}(U)$ (Fig. 4b).

The climatological uncertainty estimates illustrated in Fig. 4b exceed those found in e.g. scatterometer records in comparison to buoy measurements (e.g. Verhoef et al., 2017). On the one hand, this is linked to the fact that estimates in Fig. 4b should be treated as upper-boundary uncertainty estimates. On the other hand, scatterometers are specifically designed to derive near-surface wind speeds at highest accuracy. Passive microwave measurements, in return, allow for a much broader range of applications, which is a unique feature of HOAPS. An inclusion of scatterometer data into the HOAPS wind speed retrieval was not envisaged, due to differing overflight times and data coverage, i.e., additional uncertainties of unknown magnitude. Further potential uncertainty sources, which may contribute to the distribution shown in Fig. 4b, target currents, sea states, and the treatment of air mass density (i.e., the concept of stress-equivalent wind speeds, e.g. de Kloe et al., 2017).

$E_{clim}(q_s)$ covers the range of 0.1-0.6 g kg^{-1} and its global average is given by 0.23 g kg^{-1} (Fig. 4c). The pattern reflects a latitudinal dependency, which is equivalent to smallest (largest) biases towards the poles ((sub-) tropics). This observation is not generally valid, as is shown by the comparatively low values over large parts of the Eastern Tropical Pacific and Atlantic. Distinct maxima are found over the Arabian Sea and along northwestern Australia, the Caribbean, and west of Madagascar. Narrow bands of elevated systematic uncertainty are also resolved along the WBCs. With the exception of the WBCs, the regions of maxima are exposed to q_s in the range of 20–22 g kg^{-1}.

Figure 4d shows the resulting $E_{clim}(LHF)$. It closely resembles that of the global mean LHF pattern itself with values ranging between roughly 15–50 W m^{-2} and a global mean of ~~25.1~~ 25 W m^{-2}. Relating this pattern to Fig. 4a-c shows a substantial contribution of $E_{clim}(q_a)$ to the absolute maximum of $E_{clim}(LHF)$ in the Northern/Southern Tropical Central Pacific, the Caribbean, and the western tropical South Atlantic (compare Fig. 4a). However, due to the large climatological mean LHF, respective relative systematic uncertainties of q_a are merely in the order of 5–7 %. Correspondingly, imprints of $E_{clim}(U)$ are clearly seen along the WBCs, the Central Indian Ocean (10–15 % in a relative sense), and off the Arabian Peninsula (partly exceeding 15 %) (Fig. 4b). Likewise, the maxima in $E_{clim}(LHF)$ over the Arabian Sea, along the northwestern coast of Australia, and close to Madagascar show the footprint of $E_{clim}(q_s)$ (Fig. 4c). However, relative systematic uncertainties in q_s generally do not exceed 2.5 %. Locally, isolated $E_{clim}(LHF)$ maxima are resolved along 35° S. Specifically over the Agulhas Current, Santorelli et al. (2011) conclude that different satellite data sets show discrepancies, as they are not able to properly handle strong LHF associated with storm systems and potential LHF amplifications due to dry air advection northwards from the Antarctic (Grodsky et al., 2009). Furthermore, note that the maximum in the Arabian Sea is somewhat special, in as much as climatological mean LHF in this region are elevated, yet not extraordinarily large. This striking uncertainty maximum may be linked to occasionally occuring advection of hot, dry air masses from the deserts, which poses problems to the HOAPS-3.3 satellite retrieval. This hypothesis is strengthened by the fact that Iwasaki et al. (2014) show largest deviations in HOAPS-3 q_s with respect to their reference climatology, which are not seen in the remaining data sets.

Figure 4e relates $E_{clim}(LHF)$ to its natural variability (compare Sect. 4.2). The global average is in the order of 12 %. Apart from the WBC regimes and the Southern Ocean, largest relative uncertainties are in line with the $E_{clim}(LHF)$ maxima illustrated in Fig. 4d.

4.4 ~~Fractional contributions to total LHF uncertainty~~Monthly Mean HOAPS-3.3 Sampling Uncertainties

Table 2 summarizes the average of monthly mean sampling uncertainties of several LHF-related HOAPS-3.3 parameters as a function of concurrently operating SSM/I instruments. From a climatological perspective, all magnitudes are negligibly small compared to respective systematic uncertainties. SST-related parameters show largest sampling uncertainties when three SSM/I instruments are simultaneously operating. This is not contradictory, as HOAPS-3.3 SST are AVHRR-based and thus not linked to the coverage of SSM/I instruments. Regarding the main bulk parameters, orders of magnitude closely resemble those of monthly mean scaled E_{retr}^{ran} (not shown). It is concluded that their relative contribution to the monthly mean uncertainty budget is in the order of merely 1–2 %. However, one should keep in mind that sampling uncertainties become essential on considerably shorter time scales, i.e., in the framework of daily analyses.

4.5 Fractional contributions to total HOAPS-3.3 LHF uncertainty

Simply comparing Fig. 4a-c to Fig. 4d allows for qualitatively assessing which LHF-related parameter contributes most to $E_{clim}(LHF)$. However, this does not permit a quantitative conclusion. Following a modified version of the 'Q-term' approach demonstrated in Bourras (2006), $E_{clim}(LHF)$ is decomposed into fractions associated with U, q_s, q_a, and C_E. Results indicate that the global mean contribution of $E_{clim}(q_a)$ is largest (60 %). This specifically targets the Central Northern and Southern Tropical Pacific, the Caribbean, the regime off the eastern tip of South America, as well as the Central Indian Ocean. This finding is in line with that of Iwasaki et al. (2014) , who show that HOAPS-3 q_a contributes most to the observed deviation in E with respect to their reference climatology.

On average, the contribution by $E_{clim}(U)$ takes on a value of 25 %. Local hotspots are considerably larger, especially over the Arabian Sea, along the WBCs, and off Northwestern Australia. The fractional contributions due to both $E_{clim}(q_s)$ and $E_{clim}(C_E)$ equal to 7.5 %, respectively. $E_{clim}(q_s)$ is largest over the Arabian Sea (SST retrieval issues due to dust particles), whereas $E_{clim}(C_E)$ maximises over the Central Indian Ocean and along the North Atlantic WBC. The latter has also been shown by Bourassa et al. (2013), in as much as accuracy issues in C_E tend to occur over very low and very high wind speed regimes.

All findings are in line with Bourras (2006), Liu and Curry (2006), Grodsky et al. (2009), and Santorelli et al. (2011), who conclude that the main LHF uncertainty sources are related to the accuracy of q_a (and U). Similar conclusions are drawn by e.g. Tomita and Kubota (2006), who show that the main source of discrepancy between tropical satellite and buoy estimates may be attributed to the accuracy of q_a. ~~By comparison, HOAPS-3.3 uncertainty analyses are beneficial, as the~~ The findings of the above-quoted studies are restricted to either regional analyses, considerably shorter investigation periods, and/or comparatively thin reference data bases. Again, this points at the high value of the presented HOAPS-3.3 uncertainty analyses.

4.6 Regional and Seasonal ~~Analysis~~HOAPS-3.3 Uncertainty Analyses

Global mean ~~E_{clim} and~~ E_{retr}^{ran} ~~of all~~ and E_{clim} of LHF-related HOAPS-3.3 parameters are fairly constant in time throughout the whole climatology ~~. During isolated time periods, however, absolute~~ (Figs. 3-4). Absolute deviations from the global mean LHF (q_a, U) uncertainty become as large as 18 % (3 %, 8 %).

~~Next to~~ Apart from seasonal signals, these are footprints of distinct local anomalies. On the one hand, these anomalies seem to originate from events that temporarily modify the global climate. On the other hand, ~~Figures~~ Figs. 3-4 resolve considerable regional variability. Therefore, the aim is to (1) identify climate features that are manifested in both temporal and spatial uncertainty anomalies and discuss their origin (descriptive only). At the same time, (2) regional uncertainty differences shall be highlighted by focusing on climate hotspots (Fig. 5a-c).

Regarding (1): The imprints of moderate to strong El Niño events during boreal spring 1998 and 2010 are manifested in LHF-related E_{clim} and E_{retr}^{ran}. During these events, wind speeds over the Pacific upwelling regime are 1.5–2.0 m s^{-1} below the climatological average. As has been mentioned in Kinzel et al. (2016), this causes an increase in systematic uncertainties in U. Along with an enhanced $E_{clim}(q_s)$, the respective $E_{clim}(LHF)$ over the Pacific upwelling regime reaches 25 W m^{-2}, specifically during boreal spring ~~1998, which~~ 1998. This is approximately 10 W m^{-2} above the seasonal mean and more than 50 % of climatological mean LHF. As q_a are anomalously high with 20 g kg^{-1}, $E_{retr}^{ran}(q_a)$ is up to 0.2 g kg^{-1} below the seasonal mean (see Fig. 2 in Kinzel et al. (2016) for clarification).

By contrast, global minima in $E_{clim}(LHF)$ and $E_{retr}^{ran}(LHF)$ are confined to boreal autumn 1991, taking on a mean value of 20 W m^{-2} (33 W m^{-2}), respectively. These estimates are 20 % (11 %) below their climatological averages and are associated with absolute minima in HOAPS-3.3 LHF. The comparatively small systematic component is induced by $E_{clim}(U)$ ($E_{clim}(q_s)$) of -8 % (-14 %). The absolute minimum in LHF and its uncertainties during 1991 is a footprint of the Mount Pinatubo eruption, which caused low-biased SST due to AVHRR aerosol issues and thus unrealistically low near-surface humidity gradients (Romanova et al., 2010). Amongst others, this shortcoming in the HOAPS-3.3 climatology has already been picked up by Andersson et al. (2011).

Regarding (2): Figures 5a-c summarize the ranges of seasonal, regime-dependent uncertainty distributions. The color-coded boxes in Figures 5a-c represent the expected parameter ranges when considering ~~the~~ multi-annual (1988-2012) means of systematic uncertainty contributions~~(, that is E_{clim})~~. At the same time, the error bars indicate the ~~instantaneous~~ *instantaneous* random uncertainty components~~(, that is E_{retr}^{ran})~~. Both are shown separately, as they are independent of each other. With few exceptions, the random uncertainty contributions exceed the systematic counterpart, as is also mirrored in Figures 3 ~~e and 4e~~and 4.

Figure 5a indicates that the total ~~(i.e., $E_{clim} + E_{retr}^{ran}$)~~ uncertainty ranges in q_a are largest in (sub-) tropical regimes, concurrent to high q_a. In contrast to the Pacific upwelling region (red) and the Southern Ocean (cyan), the seasonal q_a variability over the Indian monsoon regime (green), the North Atlantic basin (dark blue), and specifically the North Atlantic western boundary current (brown) is striking. This also finds expression in differences in absolute uncertainties of up to $\pm$0.6 g kg^{-1} between

January and July. Largest uncertainties are in the order of $\pm$2.40 g kg^{-1} and are confined to the Indian summer monsoon season, whereas smallest uncertainties around $\pm$1 g kg^{-1} occur over the Southern Ocean.

Climatological regional wind speeds range between 4.5–11 m s^{-1} (Fig. 5b). As for q_a, the seasonality is most pronounced over the Indian monsoon region, WBC, and the North Atlantic. Largest total uncertainties exceeding $\pm$2 m s^{-1} throughout the year are observed over the Southern Ocean, which is primarily due to large $E^{ran}_{retr}(U)$ (compare Fig. 3b). The Indian monsoon region is somewhat special, in as much as summertime total uncertainties are largest on a global scale, while wintertime ranges are almost 50 % lower.

Figure 5c presents regionally dependent LHF and associated uncertainty ranges. As for Fig. 5a-b, seasonality is most distinct over the North Atlantic, WBC, and the Indian monsoon region. Largest $E_{clim}(LHF)$ exceeding $\pm$35 W m^{-2} are confined to the WBC regime (specifically during winter) and the monsoon region (climatological average, compare also Fig. 4d). Total uncertainty ranges maximise along the WBC, where $\pm$65–95 W m^{-2} are to be expected, which is 2–3 times larger compared to the ranges observed along the Pacific upwelling regime. Grodsky et al. (2009), for example, recall that an accurate representation of LHF along the Gulf Stream ~~region~~ is challenging due to strong surface currents and SST gradients as well as intraseasonal dependencies of how the stratified atmospheric boundary layer amplifies air-sea interactions. This reasoning may also apply to the Agulhas and Kuroshio region. The wintertime WBC uncertainty maximum is particularly caused by vast $E^{ran}_{retr}(LHF)$ of up to $\pm$60 W m^{-2} (see also signal in Fig. 3d). By contrast, regional $E_{clim}(LHF)$ become largest in the Indian monsoon region, where their climatological average is in the order of $\pm$40 W m^{-2} (compare also Fig. 4d).

4.7 Uncertainty Application: Trends in HOAPS-3.3 LHF

Figure 6 shows the HOAPS-3.3 global monthly mean LHF (thin black line) between 1988-2012 (70° S-70° N, cosine-weighted average). The global minimum below 80 W m^{-2} during boreal summer 1991 is linked to the Mount Pinatubo eruption. Overall maxima in the order of 110 W m^{-2} occur during 2008 and 2009.

The bold black line in Fig. 6 shows the annual running mean climatology of HOAPS-3.3 LHF. On average, it increases by roughly 4.5 W m^{-2} (4.7%) per decade (dark red line). If uncertainty ranges were discarded, this trend would be considered as significant at the 95 % level (p<0.00001, based on a two-tailed t-test). The addressed uncertainty estimates are illustrated as grey shadings and represent $\pm$1 standard deviation of the 12-month running mean ~~climatological uncertainty~~ E_{clim} (global average). They take on a mean value of $\pm$ 17 W m^{-2}.

A Bayesian approach to linear regression is applied including LHF uncertainty estimates following Kelly (2007), which yields a large range of linear trends (light red lines). Although the majority has a positive slope, some even indicate a climatological decrease in LHF. In light of the illustrated uncertainty range, the mean upward trend in HOAPS-3.3 LHF (dark red line) should therefore be treated with caution, as the magnitude of linear increase lies well within the grey shaded area.

The overall increase in LHF has been elucidated in several studies concerning various LHF data sets ~~. Amongst others,it was already detected by Liu and Curry (2006) for HOAPS2 (Fennig et al., 2006) ,GSSTF2 (Chou et al., 2004) ,and reanalysis data (NCEP-R2, ERA-40; Kanamitsu et al., 2002; Uppala et al., 2005) between 1989–2000,specifically over the (sub-) tropics~~ (e.g. Liu and Curry, 2006; Yu and Weller, 2007; Santorelli et al., 2011; Yu et al., 2011; Iwasaki et al., 2014) .

The authors attribute it to increases in both q_s (i.e., SST) and U, whereas the latter may be linked to stronger Hadley and Walker Circulations (Cess and Udelhofen, 2003). ~~Likewise, Gao et al. (2013) attribute largest contributions to observed positive trends in GSSTF2c LHF to q_s and U. Similar conclusions are drawn by Rahul and Gnanaseelan (2013) for the Indian Ocean, although local LHF decreases in the~~ The global mean increase of 9 W m^{-2} between 1981 and 2002, as is e.g. seen in Objectively Analyzed Air-Sea Heat Fluxes (OAFlux, Yu and Weller (2007)) ~~are in line with findings from a model study by Held and Soden (2006) .~~Yu and Weller (2007) present results from an OAFlux analysis and highlight the concurrent rapid warming of global SST (e.g. Levitus et al., 2005) and associated increasing q_s, especially over the North Atlantic. Concurrently, q_a decreas is in the order of 10 %, which is in line with ~~the findings~~ findings of Santorelli et al. (2011) and those illustrated in Fig. 6 ~~, yet one order of magnitude larger compared to the model study of Pierce et al. (2006) . Santorelli et al. (2011) confirm this global mean LHF increase in OAFlux and draw same conclusions for IFREMER LHF (Bentamy et al., 2008) , specifically for the North Atlantic. The increase in HOAPS-3 LHF is also seen over the Southern Ocean, as has been investigated by Yu et al. (2011) due to increases in both U and Δq. Locally, these increases between 1988–2000 are in the order of 30 W m^{-2}.~~ of the present work.

Figure 6 also shows that recent global means decrease again. Time series analyses for single satellite instruments suggest that this is a physical signal (i.e., associated with either multi-annual variability or a climate signal), rather than being associated with intercalibration issues among SSM/I and SSMIS instruments. ~~However, its decrease may~~ Additionally, the decrease may also be attributed to the slight negative SST bias from 2011 onwards. This bias is caused by anomalously high NOAA-19 sensor noises, which themselves may be traced back to erroneous flag assignments during cloud detection. This is thought to cause up to 5-10 % reduction in LHF. Closer investigations that involve other LHF climatologies exceed the scope of this study, but are needed to interpret this gradual decay.

First intercomparisons of HOAPS-3.3 LHF to *in situ* and further satellite climatologies have been carried out, where preliminary results indicate that nearly all compared data sets lie within the uncertainty range presented in Fig. 6 (not shown). A more detailed intercomparison study is envisaged; it will benefit from uncertainty estimates available in NOCSv2.0 and allow for concluding whether global mean deviations among the data sets lie within or outside of the HOAPS-3.3 prescribed uncertainty range.

5 Conclusions and Outlook

By means of multi-dimensional bias and MTC ~~analysis~~analyses, a universal approach for characterizing systematic, random retrieval, and sampling uncertainties inherent to HOAPS-3.3 LHF-related parameters has been presented. The multi-dimensional approach overcomes the issues of sparse data densities in remote regions, as it expresses the uncertainties as a function of the ambient atmospheric conditions. At the same time, MTC enables a decomposition of random uncertainty sources to isolate the contribution of the satellite retrieval. Both methods represent the main procedures to arrive at pixel-level uncertainty information, which essentially increases the value of HOAPS-3.3~~can therefore~~ . As to sampling uncertainties, monthly mean estimates have been calculated following the approach of Tomita and Kubota (2011) . To conclude, HOAPS-3.3 can be con-

sidered as the first LHF satellite-only climatology including instantaneous and gridded uncertainty estimates. As the method can be easily transferred to other retrievals, it lays the foundation for uncertainty characterizations of further LHF-related data sets, which increases the significance of this work.

It has been shown that maxima of systematic uncertainties (E_{clim}) reach up 50 W m^{-2}, specifically over the large regions of the subtropical oceans (mainly q_a-induced) and along the western boundary currents (mainly U-induced). Instantaneous random retrieval uncertainties (E_{retr}^{ran}) maximise along 20–30° N/S with values up to 60 W m^{-2}, clearly showing the footprint of random uncertainties of q_a. From a climatological perspective, all random retrieval uncertainty components contribute to the total uncertainty by merely 1–2 % on a monthly basis (and even less for longer periods), which also accounts for respective sampling uncertainties. Considerable regional and seasonal variability of LHF uncertainty ranges have been resolved from an instantaneous point of view, with maxima over the Gulf Stream and Indian monsoon region during boreal winter. Climate events, such as strong El Niño signals and the Mount Pinatubo eruption, are well manifested in both systematic and random LHF uncertainties, even on a global scale. In light of the available uncertainty estimates, it has been shown that the positive trend in global mean LHF during the last 25 years lies within the derived uncertainty boundaries ~~.~~

~~A new version of HOAPS-3.3, that is HOAPS-4.0, will be released in mid 2017. Major changes compared to HOAPS-3.3 include a temporal extension up to 2014, a new SST product (Version 2 of the NOAA Optimum Interpolation SST (OISST) product, Reynolds et al. (2007)), and the implementation of a 1D-Var retrieval for several geophysical parameters. Preliminary results suggest that the new U estimates have improved compared to HOAPS-3.3 in terms of bias and RMSD behaviour relative to *in situ* ground reference data. In consequence, estimates of LHF and E will be updated, along with LHF-related uncertainty estimates~~needs to therefore be treated with caution.

Results of the Q-term analysis presented in Sect. 4.5 and other studies suggest that more effort is necessary to improve the q_a retrieval. This would ultimately reduce the overall LHF uncertainty, which, according to e.g. Bourras (2006), ought to be below 10 W m^{-2} for a quantitative use over the global oceans. ~~In the framework of the HOAPS-4.0 release, this value has also been declared as the target requirement for the global mean LHF.~~ An increase in the reliability of HOAPS-3.3 LHF-related parameters could for example be achieved by referring to a new ground truth reference. Freeman et al. (2016), for example, recently presented a new version of ICOADS (release 3.0, up to 2014), highlighting its improvements compared to earlier versions, which target topics such as data quality, data traceability, and data base extension. Apart from new *in situ* reference data, the effect of approximations in bulk flux parameterizations should also be picked up, as has been done in detail in Brodeau et al. (2017). Amongst others, this concerns implications of sensor height corrections, algorithm choices, the q_s reduction due to the salinity effect, cool skin/ warm layer effects, and the assumption of constant sea level pressure.

According to Andersson et al. (2011), the E-P budget of HOAPS-3.2 is not closed. This also accounts for HOAPS-3.3, with a climatological mean value of 0.45 mm d^{-1} (1988–2012, 70° S-70° N). Long-term run-off estimates are summarized and published by the Global Runoff Data Center (GRDC), adding up to a mean value of 0.34 mm d^{-1} (Wilkinson et al., 2014). According to Andersson et al. (2011), the uncertainty of these run-off estimates is in the order of 10–20 %. Comparing these values to the HOAPS-3.3 global freshwater flux leaves an imbalance of approximately 0.10 mm d^{-1}, which is 0.30 mm d^{-1} below the HOAPS-3.2 estimate and can be evaluated as an improvement towards closing the global freshwater flux imbalance.

As $E_{clim}(E)$ is in the order of $\pm$ 0.6 mm d^{-1}, the imbalance clearly lies in the range of freshwater flux uncertainty. Keeping this uncertainty range in mind sheds new light on the conclusion by Iwasaki et al. (2014) that the HOAPS-3 freshwater budget (including river run off) is largest compared to the remaining data sets. A unit conversion from mm d^{-1} to kg year^{-1} allows for qualitatively estimating, whether the intercompared data sets in Iwasaki et al. (2014) (their Figure 6a) lie within the derived uncertainty range of HOAPS. As 0.6 mm d^{-1} corresponds to roughly 0.8*10^{17} kg year^{-1}, we conclude that all satellite and hybrid related time series lie within the uncertainty range. This does not account for the reanalyses; according to the authors, these tend to overestimate E, which is associated with the underlying bulk flux algorithm.

Recall, however, that uncertainty estimates of HOAPS-3.3 precipitation have not been accounted for in this quantitative estimation. Generally, the availability of remotely sensed precipitation uncertainty estimates is complicated by sparse reference data and its intermittency. ~~Tian and Peters-Lidard (2010) , for example, have taken on the challenge of creating global maps of uncertainties in satellite-based (i.e., six TRMM-era data sets) precipitation measurements. In conclusion, overall uncertainties range between 40–60 over the tropical oceans, whereas uncertainties may exceed 100 over the higher-latitudinal regimes poleward of 40° N/S.~~ A recent study by Burdanowitz et al. (2016) presents an automatic phase distinction algorithm for optical disdrometer data. Together with a continuously growing high-quality *in situ* data base of ship-based precipitation measurements (OceanRAIN, Klepp (2015)), it will serve as a valuable basis for a characterization of HOAPS-3.3 precipitation and hence freshwater flux uncertainty ranges in the near future. ~~Accuracy assessments of global rainfall estimates can also be achieved by means of triple collocation analysis, as is demonstrated in Massari et al. (2017) .~~

Future work also aims at investigating trends in water vapour transports (WVT), using HOAPS-3.3 monthly mean freshwater fluxes. Sohn and Park (2010), for example, demonstrated that trends in WVT can be used to examine circulation changes and conclude that the large-scale Hadley Circulation has experienced an increase in strength since 1979. Similarly, Durack et al. (2012) recently highlighted a considerable water cycle intensification during global warming. Available uncertainty estimates will allow for quantifying the WVT uncertainty range, the necessity of which has been picked up by e.g. Sohn et al. (2004).

A new version of HOAPS-3.3, that is HOAPS 4.0, has been released in October 2017 (Andersson et al., 2017) . Major changes compared to HOAPS-3.3 include a temporal extension up to 2014, a new SST product (Version 2 of the NOAA Optimum Interpolation SST (OISST) product, Reynolds et al. (2007)), and the implementation of a 1D-Var retrieval for several geophysical parameters. Preliminary results suggest that the new U estimates have improved compared to HOAPS-3.3 in terms of bias and RMSD behaviour relative to *in situ* ground reference data. In consequence, estimates of LHF and E have been updated, along with LHF-related uncertainty estimates.

Data availability: HOAPS-3.3 is a prolongation of HOAPS-3.2 and is based on a pre-release of the CM SAF SSM/I and SSMIS FCDR. It was created in the framework of the DFG FOR1740 research activity for internal use. The monthly mean HOAPS-3.2 climatology and the respective FCDR are publicly available and may be downloaded free of charge (http://www.cmsaf.eu/EN/Products/DOI Instantaneous and gridded HOAPS-3.3 data are available upon request from the author.

Competing Interests: The authors declare that they have no conflict of interest.

Acknowledgements. J. K. is funded by the German ~~Science~~ Research Foundation (DFG FOR1740/FOR21740). The funding for the development and implementation of the collocation software was provided by the German Meteorological Service (DWD). HOAPS-3.3 was generated within DFG FOR1740. DWD-ICOADS data was gratefully obtained from the Marine Climate Data Center (DWD).

References

Andersson, A., Fennig, K., Klepp, C., Bakan, S., Grassl, H., and Schulz, J.: The Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data - HOAPS-3, Earth System Science Data, 2, 215–234, doi:10.5194/essd-2-215-2010, 2010.

Andersson, A., Klepp, C., Fennig, K., Bakan, S., Grassl, H., and Schulz, J.: Evaluation of HOAPS-3 Ocean Surface Freshwater Flux Components, J. Appl. Meteor. Climatol., 50, 379–398, doi:10.1175/2010JAMC2341.1, 2011.

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

Baumgartner, A. and Reichel, E.: The World Water Balance, Elsevier, New York, USA, 1975.

Bentamy, A., Katsaros, K. B., Mestas-Nuñez, A. M., Drennan, W. M., Forde, E. B., and Roquet, H.: Satellite Estimates of Wind Speed and Latent Heat Flux over the Global Oceans, J. Climate, 16, 637–656, doi:10.1175/1520-0442(2003)016<0637:SEOWSA>2.0.CO;2, 2003.

Bentamy, A., Ayina, L.-H., Drennan, W., Katsaros, K., Mestas-Nuñez, A. M., and Pinker, R. T.: 15 years of ocean surface momentum and heat fluxes from remotely sensed observations, FLUXNEWS, 5, 14–16, [available online at http://sail.msk.ru/newsletter/fluxnews_5_final.pdf]., 2008.

Bentamy, A., Grodsky, S. A., Katsaros, K. B., Mestas-Nuñez, A. M., Blanke, B., and Desbiolles, F.: Improvement in air-sea flux estimates derived from satellite observations, International Journal of Remote Sensing, 34, 5243–5261, doi:10.1080/01431161.2013.787502, 2013.

Berry, D. I. and Kent, E. C.: A new Air-Sea Interaction Gridded Dataset from ICOADS with Uncertainty Estimates, Bull. Amer. Meteor. Soc., 90, 645–656, doi:10.1175/2008BAMS2639.1, 2009.

Bourassa, M. A., Gille, S. T., Bitz, C., Carlson, D., Cerovecki, I., Clayson, C. A., Cronin, M. F., Drennan, W. M., Fairall, C. W., Hoffman, R. N., Magnusdottir, G., Pinker, R. T., Renfrew, I. A., Serreze, M., Speer, K., Talley, L. D., and Wick, G. A.: High-Latitude Ocean and Sea Ice Surface Fluxes: Challenges for Climate Research, Bull. Amer. Meteor. Soc., 94, 401–423, doi:10.1175/BAMS-D-11-00244.1, 2013.

Bourras, D.: Comparison of Five Satellite-Derived Latent Heat Flux Products to Moored Buoy Data, J. Climate, 19, 6291–6313, doi:10.1175/JCLI3977.1, 2006.

Brodeau, L., Barnier, B., Gulev, S. K., and Woods, C.: Climatologically Significant Effects of Some Approximations in the Bulk Paramererizations of Turbulent Air-Sea Fluxes, Journal of Physical Oceanography, 47, 5–28, doi:10.1175/JPO-D-16-0169.1, 2017.

Brunke, M. A., Zeng, X., and Anderson, S.: Uncertainties in sea surface turbulent flux algorithms and data sets, J. Geophys. Res., 107, C10 3141, doi:10.1029/2001JC000992, 2002.

Brunke, M. A., Fairall, C. W., Zeng, X., Eymard, L., and Curry, J. A.: Which Bulk Aerodynamic Algorithms are Least Problematic in Computing Ocean Surface Turbulent Fluxes?, J. Climate, 16, 619–635, doi:10.1175/1520-0442(2003)016<0619:WBAAAL>2.0.CO;2, 2003.

Brunke, M. A., Wang, Z., Zeng, X., Bosilovich, M., and Shie, C.-L.: An Asessment of the Uncertainties in Ocean Surface Turbulent Fluxes in 11 Reanalysis, Satellite-Derived, and Combined Global Datasets, J. Climate, 24, 5469–5493, doi:10.1175/2011JCLI4223.1, 2011.

Burdanowitz, J., Klepp, C., and Bakan, S.: An automatic precipitation-phase distinction algorithm for optical disdrometer data over the global ocean, Atmos. Meas. Tech., 9, 1637–1652, doi:10.5194/amt-9-1637-2016, 2016.

Casey, K. S., Brandon, T. B., Cornillon, P., and Evans, R.: The Past, Present and Future of the AVHRR Pathfinder SST Program, in Oceanography from Space: Revisited, edited by V. Barale, J. F. R. Gower, and L. Alberotanza, Springer, Dordrecht, Netherlands, doi:10.1007/978-90-481-8681-5_16, 2010.

Cess, R. D. and Udelhofen, P. M.: Climate change during 1985-1999: cloud interactions determined from satellite measurements, Geophys. Res. Lett., 30, 1019, doi:10.1029/2002GL016128, 2003.

Chou, S.-H., Nelkin, E. J., Ardizzone, J., and Atlas, R. M.: A Comparison of Latent Heat Fluxes over Global Oceans for Four Flux Products, J. Climate, 17, 3973–3989, doi:10.1175/1520-0442(2004)017<3973:ACOLHF>2.0.CO;2, 2004.

Clayson, C. A., Roberts, J. B., and Bogdanoff, A. S.: The SeaFlux Turbulent Flux Dataset Version 1.0 Documentation (Version 1.2), Tech. rep., Woods Hole Oceanographic Institution, Woods Hole, MA, [available online at http://seaflux.org/seaflux_data/DOCUMENTATION/SeaFluxV1.0Documentation.pdf]., 2015.

Curry, J. A., Bentamy, A., Bourassa, M. A., Bourras, D., Bradley, E. F., Brunke, M., Castro, S., Chou, S. H., Clayson, C. A., Emery, W. J., Eymard, L., Fairall, C. W., Kubota, M. K., Lin, B., Perrie, W., Reeder, R. A., Renfrew, I. A., Rossow, W. B., Schulz, J., Smith, S. R., Webster, P. J., Wick, G. A., and Zeng, X.: SEAFLUX, Bull. Amer. Meteor. Soc., 85, 409–424, doi:10.1175/BAMS-85-3-409, 2004.

de Kloe, J., Stoffelen, A., and Verhoef, A.: Improved Use of Scatterometer Measurementrs by Using Stress-Equivalent Reference Winds, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10, 2340–2347, doi:10.1109/JSTARS.2017.2685242, 2017.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Quart. J. Roy. Meteor. Soc., 137, 553–597, doi:10.1002/qj.828, 2011.

Donlon, C. J., Minnett, P. J., Gentemann, C., Nightingale, T. J., Barton, I. J., Ward, B., and Murray, M. J.: Toward Improved Validation of Satellite Sea Surface Skin Temperature Measurements for Climate Research, J. Climate, 15, 353–369, doi:10.1175/1520-0442(2002)015<0353:TIVOSS>2.0.CO;2, 2002.

Durack, P. J., Wijffels, S. E., and Matear, R. J.: Ocean Salinities Reveal Strong Global Water Cycle Intensification during 1950 to 2000, Science, 336, 455–458, doi:10.1126/science.1212222, 2012.

Fairall, C. W., Bradley, E. F., Hare, J. E., Grachev, A. A., and Edson, J. B.: Bulk Parameterization of Air-Sea Fluxes: Updates and Verification for the COARE Algorithm, J. Climate, 16, 571–591, doi:10.1175/1520-0442(2003)016<0571:BPOASF>2.0.CO;2, 2003.

Fennig, K., Bakan, S., Grassl, H., Klepp, C., and Schulz, J.: Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data - HOAPS II - monthly mean, doi:10.1594/WDCC/HOAPS2_MONTHLY, 2006.

Fennig, K., Andersson, A., Bakan, S., Klepp, C., and Schröder, M.: Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data - HOAPS 3.2 - Monthly Means / 6-Hourly Composites, doi:10.5676/EUM_SAF_CM/HOAPS/V001, 2012.

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

Freeman, E., Woodruff, S. D., Worley, S. J., Lubker, S. J., Kent, E. C., Angel, W. E., Berry, D. I., Brohan, P., Eastman, R., Gates, L., Gloeden, W., Ji, Z., Lawrimore, J., Rayner, N. A., Rosenhagen, G., and Smith, S. R.: ICOADS Release 3.0: a major update to the historical marine climate record, Int. J. Climatol., p. in press, doi:10.1002/joc.4775, 2016.

Fuhrhop, R. and Simmer, C.: SSM/I Brightness Termperature Corrections for Incidence Angle Variations, J. Atmos. Oceanic Technol., 13, 246–254, doi:10.1175/1520-0426(1996)013<0246:SBTCFI>2.0.CO;2, 1996.

Gao, S., Chiu, L. S., and Shie, C.-L.: Trends and variations of ocean surface latent heat flux: Results from GSSTF2c data set, J. Geophys. Res., 40, 380–385, doi:10.1029/2012GL054620, 2013.

Gleckler, P. J. and Weare, B. C.: Uncertainties in Global Ocean Surface Heat Flux Climatologies Derived from Ship Observations, J. Climate, 16, 2764–2781, doi:10.1175/1520-0442(1997)010<2764:UIGOSH>2.0.CO;2, 1997.

Grodsky, S. A., Bentamy, A., Carton, J. A., and Pinker, R. T.: Intraseasonal Latent Heat Flux Based on Satellite Observations, J. Climate, 22, 4539–4556, doi:10.1175/2009JCLI2901.1, 2009.

Gulev, S., Jung, T., and Ruprecht, E.: Estimation of the Impact of Sampling Errors in the VOS Observations on Air-Sea Fluxes. Part I: Uncertainties in Climate Means, J. Climate, 20, 279–301, doi:10.1175/JCLI4010.1, 2007.

Gulev, S. K., Josey, S. A., Bourassa, M., Breivik, L.-A., Cronin, M. F., Fairall, C., Gille, S., Kent, E. C., Lee, C. M., McPhaden, M. J., Monteiro, P. M. S., Schuster, U., Smith, S., Trenberth, K. E., Wallace, D., and Woodruff, S. D.: Surface Energy, CO2 Fluxes and Sea Ice, in: Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society, edited by Hall, J., Harrison, D. E., and Stammer, D., pp. 193–211, European Space Agency, ESA Publication WPP-306, doi:10.5270/OceanObs09.pp.19, 2010.

Held, I. M. and Soden, B. J.: Robust responses of the hydrological cycle to global warming, J. Climate, 19, 5686–5699, doi:10.1175/JCLI3990.1, 2006.

Immler, F. J., Dykema, J., Gardiner, T., Whiteman, D. N., Thorne, P. W., and Vömel, H.: Reference Quality Upper-Air Measurements: guidance for developing GRUAN data products, Atmospheric Measurement Techniques, 3, 1217–1231, doi:10.5194/amt-3-1217-2010, 2010.

IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Chapter 3.4.1 [Stocker, T. F., D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P. M. Midgley (eds.)], Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

Iwasaki, S., Kubota, M. K., 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.

Jackson, D. L., Wick, G. A., and Robertson, F. R.: Improved multisensor approach to satellite-retrieved near-surface specific humidity observations, J. Geophys. Res., 114, D16 303, doi:10.1029/2008JD011341, 2009.

Josey, S. A.: Air-Sea Fluxes of Heat, Freshwater and Momentum, in Operational Oceanography in the 21st Century [Schiller, A. and Brasington, G. B. (eds.)], Springer, Dordrecht, Netherlands, doi:10.1007/978-94-007-0332-2_6, 2011.

Kanamitsu, M., Ebisuzaki, W., Woolen, J., Potter, J., and Fiorino, M.: NCEP/DOE AMIP-II reanalysis (R-2), Bull. Amer. Meteor. Soc., 83, 1631–1643, doi:10.1175/BAMS-83-11-1631, 2002.

Kelly, B. C.: Some Aspects of Measurement Error in Linear Regression of Astronomical Data, The Astrophysical Journal, 665, 1489–1506, doi:10.1086/519947, 2007.

Kent, E. C. and Berry, D. I.: Quantifying random measurement errors in voluntary observing ships' meteorological observations, Int. J. Climatol., 25, 843–856, doi:10.1002/joc.1167, 2005.

Kent, E. C. and Taylor, P. K.: Accuracy of Humidity Measurements on Ships: Consideration of Solar Radiation Effects, J. Atmos. Oceanic Technol., 13, 1317–1321, doi:10.1175/1520-0426(1996)013<1317:AOHMOS>2.0.CO;2, 1996.

Kent, E. C. and Taylor, P. K.: Toward estimating climatic trends in SST. Part II: Random Errors, J. Atmos. Oceanic Technol., 23, 475–486, doi:10.1175/JTECH1844.1, 2006.

Kent, E. C., Taylor, P. K., Truscott, B. S., and Hopkins, J. S.: The Accuracy of Voluntary Observing Ship's Meteorological Observations - Results of the VSOP-NA, J. Atmos. Oceanic Technol., 10, 591–608, doi:10.1175/1520-0426(1993)010<0591:TAOVOS>2.0.CO;2, 1993.

Kent, E. C., Woodruff, S. D., and Berry, D. I.: Metadata from WMO Publication No. 47 and an Assessment of Voluntary Observing Ship Observation Heights in ICOADS, J. Atmos. Oceanic Technol., 24, 214–234, doi:10.1175/JTECH1949.1, 2007.

Kent, E. C., Berry, D. I., Prytherch, J., and Roberts, J. B.: A comparison of global marine surface-specific humidity datasets from in situ observations and atmospheric reanalysis, Int. J. Climatol., 34, 355–376, doi:10.1002/joc.3691, 2014.

Kinzel, J.: Validation of HOAPS latent heat fluxes against parameterizations applied to R/V Polarstern data for 1995-1997, Master's thesis, University of Kiel, Germany, [available online at http://core.kmi.open.ac.uk/display/16271577]., 2013.

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.

Klepp, C.: The Oceanic Shipboard Precipitation Measurement Network for Surface Validation - OceanRAIN, Atmos. Res., Special issue of the International Precipitation Working Group, 163, 74–90, doi:10.1016/j.atmosres.2014.12.014, 2015.

Klepp, C., Andersson, A., and Bakan, S.: The HOAPS climatology: evaluation of latent heat flux, Flux News: Newsletter of the WCRP Working Group on Surface Fluxes, 5, 30–32, [available online at http://hdl.handle.net/11858/00-001M-0000-0011-FA76-C]., 2008.

Köhl, A.: Evaluation of the GECCO2 Ocean Synthesis: Transports of Volume, Heat and Freshwater in the Atlantic, Quart. J. Roy. Meteor. Soc., 141, 166–181, doi:10.1002/qj.2347, 2015.

Köhl, A. and Stammer, D.: Variability of the Meridional Overturning in the North Atlantic from the 50 years GECCO State Estimation, J. Phys. Oceanogr., 38, 1913–1930, doi:10.1175/2008JPO3775.1, 2008.

Krasnopolsky, V. M., Breaker, L. C., and Gemmil, W. H.: A Neural Network as a Nonlinear Transfer Function Model for Retrieving Surface Wind Speeds from the Special Sensor Microwave Imager, J. Geophys. Res. Oceans, 100, 11 003–11 045, doi:10.1029/95JC00857, 1995.

Kubota, M. K., Iwasaka, N., Kizu, S., Konda, M., and Kutsuwada, K.: Japanese Ocean Flux Data Sets with Use of Remote Sensing Observations (J-OFURO), J. Oceanogr., 58, 213–225, doi:10.1023/A:1015845321836, 2002.

Levitus, S., Antonov, J. I., and Boyer, T. P.: Warming of the world ocean, 1995-2003, Geophys. Res. Lett., 32, L02 604, doi:10.1029/2004GL021592, 2005.

Liu, J. and Curry, J. A.: Variability of the tropical and subtropical ocean surface latent heat flux during 1989-2000, Geophys. Res. Lett., 33, L05 706, doi:10.1029/2005GL024809, 2006.

Liu, W., Zhang, A., and Bishop, J.: Evaporation and solar irradiance as regulators of sea surface temperature in annual and interannual changes, J. Geophys. Res., 99, 12 623–12 638, doi:10.1029/94JC00604, 1994.

Loew, A., Bell, W., Brocca, L., Bulgin, C., Burdanowith, J., Calbet, X., Donner, R. V., Ghent, D., Gruber, A., Kaminski, T., Kinzel, J., Klepp, C., Lambert, J.-C., Schaepman-Strub, G., Schröder, M., and Verhoelst, T.: Validation Practices for Earth Observation Data Across Communities, Reviews of Geophysics, 2017.

Massari, C., Crow, W., and Brocca, L.: An assessment of the accuracy of global rainfall estimates without ground-based observations, Hydrology and Earth System Sciences Discussions, 21, 4347–4361, doi:10.5194/hess-21-4347-2017, 2017.

McClain, E. P.: Global sea surface temperatures and cloud clearing for aerosol optical depth estimates, International Journal of Remote Sensing, 10, 763–769, doi:10.1080/01431168908903917, 1989.

Mehta, V. M., DeCandis, A. J., and Mehta, A. V.: Remote-sensing-based estimates of the fundamental global water cycle: annual cycle, J. Geophys. Res., 110, D22 103, doi:10.1029/2004JD005672, 2005.

Mohanty, U. C., Ramesh, K. J., and Pant, M. C.: Certain Seasonal Characteristic Features of Oceanic Heat Budget Components over the Indian Seas in Relation to the Summer Monsoon Activity over India, Int. J. Climatol., 16, 243–264, doi:10.1002/(SICI)1097-0088(199603)16:3<243::AID-JOC2>3.0.CO;2-B, 1996.

Murray, F. W.: On the computation of saturation vapor pressure, J. Appl. Meteor., 6, 203–204, doi:10.1175/1520-0450(1967)006<0203:OTCOSV>2.0.CO;2, 1967.

O'Carroll, A. G., Eyre, J. R., and Saunders, R. W.: Three-way Error Analysis Between AATSR, AMSR-E, and In Situ Sea Surface Temperature Observations, J. Atmos. Oceanic Technol., 25, 1197–1207, doi:10.1175/2007JTECHO542.1, 2008.

Pierce, D. W., Barnett, T. P., Achutarao, K. M., Gleckler, P. J., Gregory, J. M., and Washington, W. M.: Anthropogenic warming of the oceans: Observations and model results, J. Climate, 19, 1873–1900, doi:10.1175/JCLI3723.1, 2006.

Prytherch, J., Kent, E. C., Fangohr, S., and Berry, D. I.: A comparison of SSM/I-derived global marine surface-specific humidity datasets, Int. J. Climatol., doi:10.1002/joc.4150, 2014.

Rahul, S. and Gnanaseelan, C.: Net Heat Flux Over the Indian Ocean: Trends, Driving Mechanisms, and Uncertainties, IEEE Transactions on Geoscience and Remote Sensing, 10, 776–780, doi:10.1109/LGRS.2012.2223194, 2013.

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.

Roberts, J. B., Clayson, C. A., Robertson, F. R., and Jackson, D. L.: Prediciting near-surface atmospheric variables from Special Sensor Microwave/Imager using neural networks with a first-guess approach, J. Geophys. Res., 115, D19 113, doi:10.1029/2009JD013099, 2010.

Romanova, V., Köhl, A., Stammer, D., Klepp, C., and Andersson, A.: Sea surface freshwater flux estimates from GECCO, HOAPS and NCEP, Tellus, 62, 435–452, doi:10.1111/j.1600-0870.2010.00447.x, 2010.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.-T., Chuang, H.-Y., Juang, H.-M. H., Sela, J., Iredell, M., Treadon, R., Kleist, D., Delst, P. V., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., Dool, H. V. D., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., chemm, J.-K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.-Z., Liu, Q., Chen, Y., Cucurull, L., Reynolds, R. W., Rutledge, G., and Goldberg, M.: The NCEP Climate Forecast System Reanalysis, Bull. Amer. Meteor. Soc., 91, 1015–1057, doi:10.1175/2010BAMS3001.1, 2010.

Santorelli, A., Pinker, R. T., Bentamy, A., Katsaros, K. B., Drennan, W. M., nez, A. M. M.-N., and Carton, J. A.: Differences between two estimates of air-sea turbulent heat fluxes over the Atlantic Ocean, J. Geophys. Res., 116, C09 028, doi:10.1029/2010JC006927, 2011.

Schlosser, A. and Houser, R.: Assessing a satellite-era perspective of the global water cycle, J. Climate, 20, 1316–1338, doi:10.1175/JCLI4057.1, 2006.

Schlüssel, P.: Satellite Remote Sensing of Evaporation over Sea, in: Radiation and Water in the Climate System: Remote measurements, Vol. 45, NATO ASI Series, pp. 431–461, Springer-Verlag, Berlin, Germany, 1996.

Schulz, J., Schlüssel, P., and Grassl, H.: Water vapor in the atmospheric boundary layer over oceans from SSM/I measurements, International Journal of Remote Sensing, 14, 2773–2789, doi:10.1080/01431169308904308, 1993.

Shearman, R. J. and Zelenko, A. A.: Wind measurements reduction to a standard level. Marine Meteorology and Related Oceanographic Activities Rep. 22, Tech. rep., WMO TD 311, Geneva, CH, 1989.

Shie, C.-L., Hilburn, K., Chiu, L. S., Adler, R., Lin, I.-I., Nelkin, E. J., Ardizzone, J., and Gao, S.: Goddard Satellite-Based Surface Turbulent Fluxes, Daily Grid F13, Version 3 [Savtchenko, A. (ed.)], Tech. rep., Goddhard Earth Science Data and Information Services Center (GES DISC), Greenbelt, MD, USA, doi:10.5067/MEASURES/GSSTF/DATA304, 2012.

Smith, S. R., Hughes, P. J., and Bourassa, M. A.: A comparison of nine monthly air-sea flux products, Int. J. Climatol., 31, 1002–1027, doi:10.1002/joc.2225, 2011.

Sohn, B.-J. and Park, S.-C.: Strengthened tropical circulations in past three decades inferred from water vapor transport, J. Geophys. Res., 115, D15 112, doi:10.1029/2009JD013713, 2010.

Sohn, B.-J., Smith, E. A., Robertson, F. R., and Park, S.-C.: Derived Over-Ocean Water Vapor Transports from Satellite-Retrieved E-P Datasets, J. Climate, 17, 1352–1365, doi:10.1175/1520-0442(2004)017<1352:DOWVTF>2.0.CO;2, 2004.

Stendardo, I., Rhein, M., and Hollmann, R.: A high resolution salinity time series 1993-2012 in the North Atlantic from Argo and Altimeter data, J. Geophys. Res. Oceans, 121, 2523–2551, doi:10.1002/2015JC011439, 2016.

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

Tian, Y. and Peters-Lidard, C.: A global map of uncertainties in satellite-based precipitation measurements, Geophys. Res. Lett., 37, L24 407, doi:10.1029/2010GL046008, 2010.

Tomita, H. and Kubota, M. K.: An analysis of the accuracy of Japanese ocean flux data sets with use of remote sensing observations (J-OFURO) satellite-derived latent heat flux using moored buoy data, J. Geophys. Res., 32, C07 007, doi:10,1029/2005 JC003013, 2006.

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.

Trenberth, K. E., Smith, L., Qian, T., Dai, A., and Fasullo, J.: Estimates of the global water budget and its annual cycle using observational and model data, Journal of Hydrometeorology, 8, 758–769, doi:10.1175/JHM600.1, 2007.

Trenberth, K. E., Fasullo, J. T., and Kiehl, J.: Earth's Global Energy Budget, Bull. Amer. Meteor. Soc., 90, 311–323, doi:10.1175/2008BAMS2634.1, 2009.

Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Berg, L. V. D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., Mcnally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Wollen, J.: The ERA-40 re-analysis, Quart. J. Roy. Meteor. Soc., 131, 2961–3012, doi:10.1256/qj.04.176, 2005.

Verhoef, A., Vogelzang, J., Verspeek, J., and Stoffelen, A.: Long-Term Scatterometer Wind Climate Data Records, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10, 2186–2194, doi:10.1109/JSTARS.2016.2615873, 2017.

Wang, W. and McPhaden, M. J.: What is the mean seasonal cycle of surface heat flux in the equatorial Pacific?, J. of Geophys. Res., 106, 837–857, doi:10.1029/1999JC000076, 2001.

Weller, R. A., Bradley, E. F., Edson, J. B., Fairall, C. W., Brooks, I., Yelland, M. J., and Pascal, R. W.: Sensors for physical fluxes at the sea surface: energy, heat, water, salt, Ocean Science, 4, 5247–263, doi:10.5194/os-4-247-2008, 2008.

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.

Wilkinson, K., von Zabern, M., and Scherzer, J.: Global Freshwater Fluxes into the World Oceans), Tech. rep., Federal Institute of Hydrology, Koblenz, doi:10.5675/GRDC_Report_44, 2014.

Winterfeldt, J. A., Andersson, A., Klepp, C., Bakan, S., and Weisse, R.: Comparison of HOAPS, QuikSCAT, and Buoy Wind Speed in the Eastern North Atlantic and the North Sea, IEEE Transactions on Geoscience and Remote Sensing, 48, 338–348, doi:10.1109/TGRS.2009.2023982, 2010.

Table 1. Absolute and relative random statistical measures resulting from the multi-dimensional LUTs, i.e., MTC and random uncertainty decomposition (Sect. 3.2, 3.3). 'stddev' = standard deviation, 'abs' = absolute, 'rel' = relative. Apart from the LHF-related bulk parameters themselves (U, q_s, and q_a), global mean ranges of the random retrieval- (E_{retr}^{ran}), random collocation- (E_c), and random *in situ* measurement uncertainty (E_{ins}) are shown. Relative measures result from bin-wise relative uncertainty calculations. For comparison, the asterisks indicate respective estimates published in Kent and Berry (2005), which are based on a semivariogram approach.

parameter / stat. measure	**mean**	**stddev**	**min (abs)**	**min (rel)**	**max (abs)**	**max (rel)**
q_a [g kg^{-1}]	8.8	4.4	2.8	/	19.3	/
$E_{retr}^{ran}(q_a)$	1.0	0.3	0.7	6 %	1.8	24 %
$E_c(q_a)$	0.5	0.1	0.4	3 %	0.7	18 %
$E_{ins}(q_a)$	0.5 [0.9*]	0.3 [0.3*]	0.1	4 %	1.2	7 %
U [m s^{-1}]	7.9	3.6	1.8	/	15.4	/
$E_{retr}^{ran}(U)$	1.4	0.4	1.0	12 %	2.6	63 %
$E_c(U)$	1.4	0.3	0.8	12 %	2.0	44 %
$E_{ins}(U)$	1.8 [2.5*]	0.2 [0.4*]	1.5	15 %	2.3	111 %
q_s [g kg^{-1}]	10.2	5.7	4.5	/	24.3	/
$E_{retr}^{ran}(q_s)$	0.5	0.2	0.2	2 %	0.9	9 %
$E_c(q_s)$	0.5	0.1	0.4	2 %	0.6	14 %
$E_{ins}(q_s)$	0.6	0.5	< 0.1	1 %	1.5	8 %

Woodruff, S. D., Worley, S. J., Lubker, S. J., Ji, Z., Freeman, J. E., Berry, D. I., Brohan, P., Kent, E. C., Reynolds, R. W., Smith, S. R., and Wilkinson, C.: ICOADS Release 2.5: extensions an enhancements to the surface marine meteorological archive, Int. J. Climatol., 31, 951–967, doi:10.1002/joc.2103, 2011.

Yu, L. and Weller, R. A.: Objectively Analyzed Air–Sea Heat Fluxes for the Global Ice-Free Oceans (1981–2005), Bull. Amer. Meteor. Soc., 88, 527–539, doi:10.1175/BAMS-88-4-527, 2007.

Yu, L., Zhang, Z., Zhong, S., Zhou, M., Gao, Z., Wu, H., and Sun, B.: An inter-comparison of six latent and sensible heat flux products over the Southern Ocean, Polar Research, 30, 10 167, doi:10.3402/polar.v30i0.10167, 2011.

Zeng, X., Zhao, M., and Dickinson, R. E.: Intercomparison of bulk aerodynamic algorithms for the computation of sea surface fluxes using TOGA COARE and TAO data, J. Climate, 11, 2628–2644, doi:doi:10.1175/1520-0442(1998)011<2628:IOBAAF>2.0.CO;2, 1998.

Table 2. ~~Monthly~~ Average of monthly mean HOAPS-3.3 LHF-related sampling uncertainties (E_{smp}) as a function of simultaneously operating SSM/I instruments (1995-2008). q_a = "hair", U = "wind", q_s = "hsea", LHF = "late", SST = "asst", E = "evap", air temperature = "tair". All magnitudes are negligible compared to the instantaneous random (E_{retr}^{ran}) and climatological uncertainties (E_{clim}) presented in Sect. 4.2 and 4.4.

# of satellites / parameters	"hair" [g kg^{-1}]	"wind" [m s^{-1}]	"hsea" [g kg^{-1}]	"late" [W m^{-2}]	"asst" [K]	"evap" [mm d^{-1}]	"tair" [K]
1	0.05	0.14	0.04	2.3	0.04	0.08	0.08
2	0.03	0.12	0.04	1.9	0.03	0.07	0.05
3	0.03	0.11	0.05	1.8	0.04	0.06	0.04

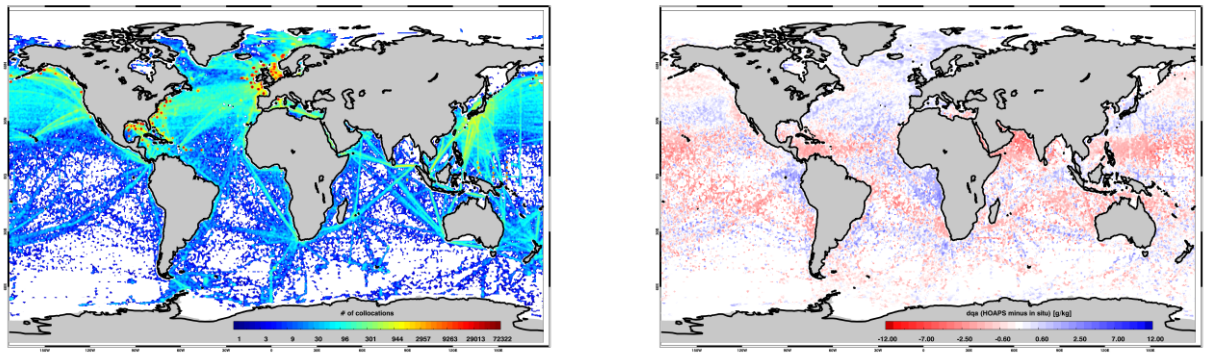


Figure 1. *Left panel*: Global map showing the distribution of collocated q_a measurements (HOAPS versus *in situ*) between 2001-2008. Overall, more than 13.8 million matchups contribute to this density map. Note that the colorbar is logarithmic. *Right panel*: Two-dimensional illustration of the near-surface humidity biases dq_a (HOAPS minus in situ, 2001-2008) shown in Fig. 2. Note that the colorbar is not linear.

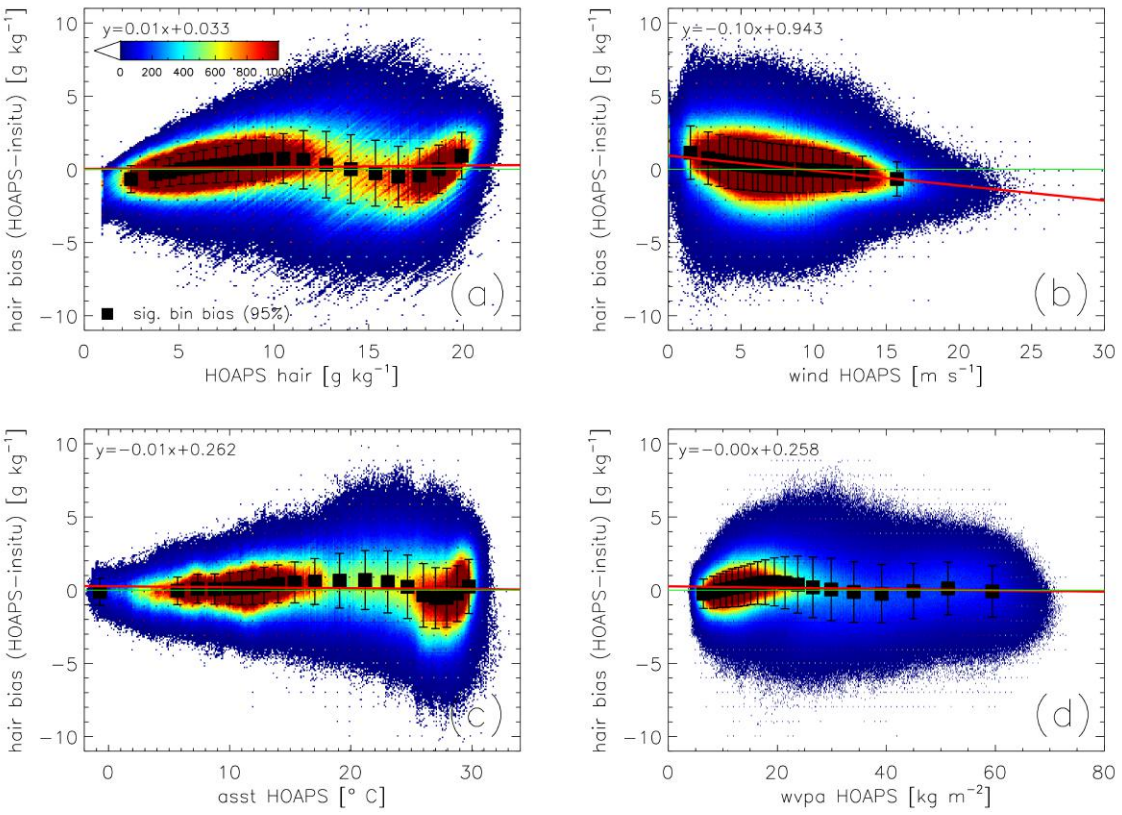


Figure 2. Scatter density plots of q_a bias (HOAPS-3.3 minus *in situ*, g kg^{-1}) as a function of (a) q_a ("hair"), (b) U ("wind"), (c) SST ("asst"), and (d) water vapour path ("wvpa"), based on global double collocations between 2001 and 2008. The black squares and error bars represent bin-averaged systematic uncertainties (significant at the 95 % level) and their standard deviations, whereby each bin contains 5 % of all double collocated matchups. Note that the bars include random uncertainty contributions by the satellite retrieval, the collocation procedure, and the *in situ* measurement uncertainty. (a) is a revised version of Fig. 3 published in Kinzel et al. (2016).

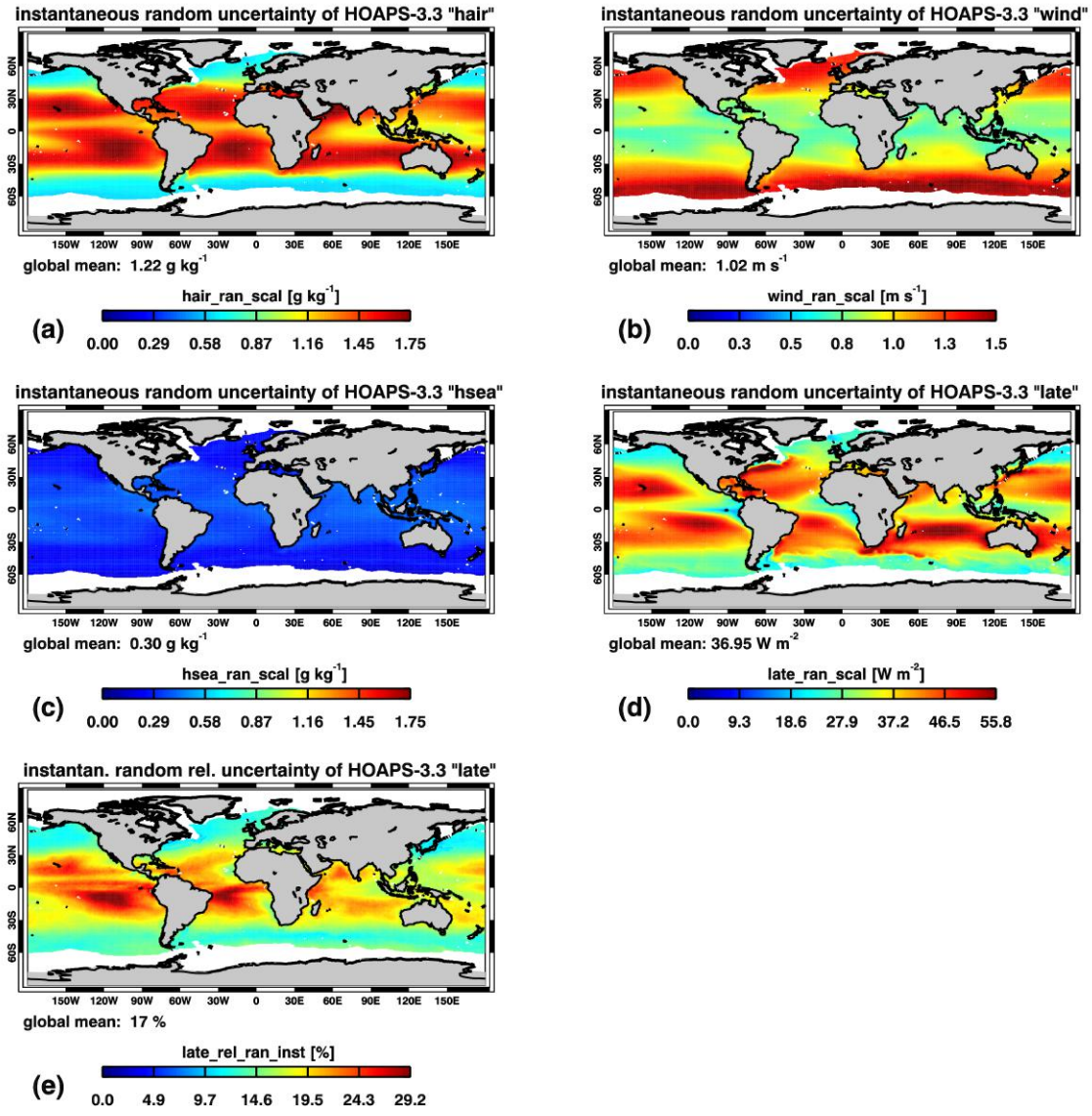


Figure 3. Temporal averages (1988-2012) of HOAPS-3.3 instantaneous E_{retr}^{ran} of (a) q_a ("hair"), (b) U ("wind"), (c) q_s ("hsea"), and (d) LHF ("late"). (e) Relative random retrieval uncertainty of HOAPS-3.3 LHF with respect to its natural variability. This variability is defined as the range between the 5th and 95th percentile of instantaneous LHF between 2000-2008. The global averages (text strings) were derived by considering a latitudinal cosine-dependency. All patterns result from the multi-dimensional bias analyses, MTC, random uncertainty decompositions, and, in case of (d), uncertainty propagation described in Sect. 3.2-~~3.3~~3.4. Note that the color bar ranges of (a) and (c) are identical to allow for direct comparisons.

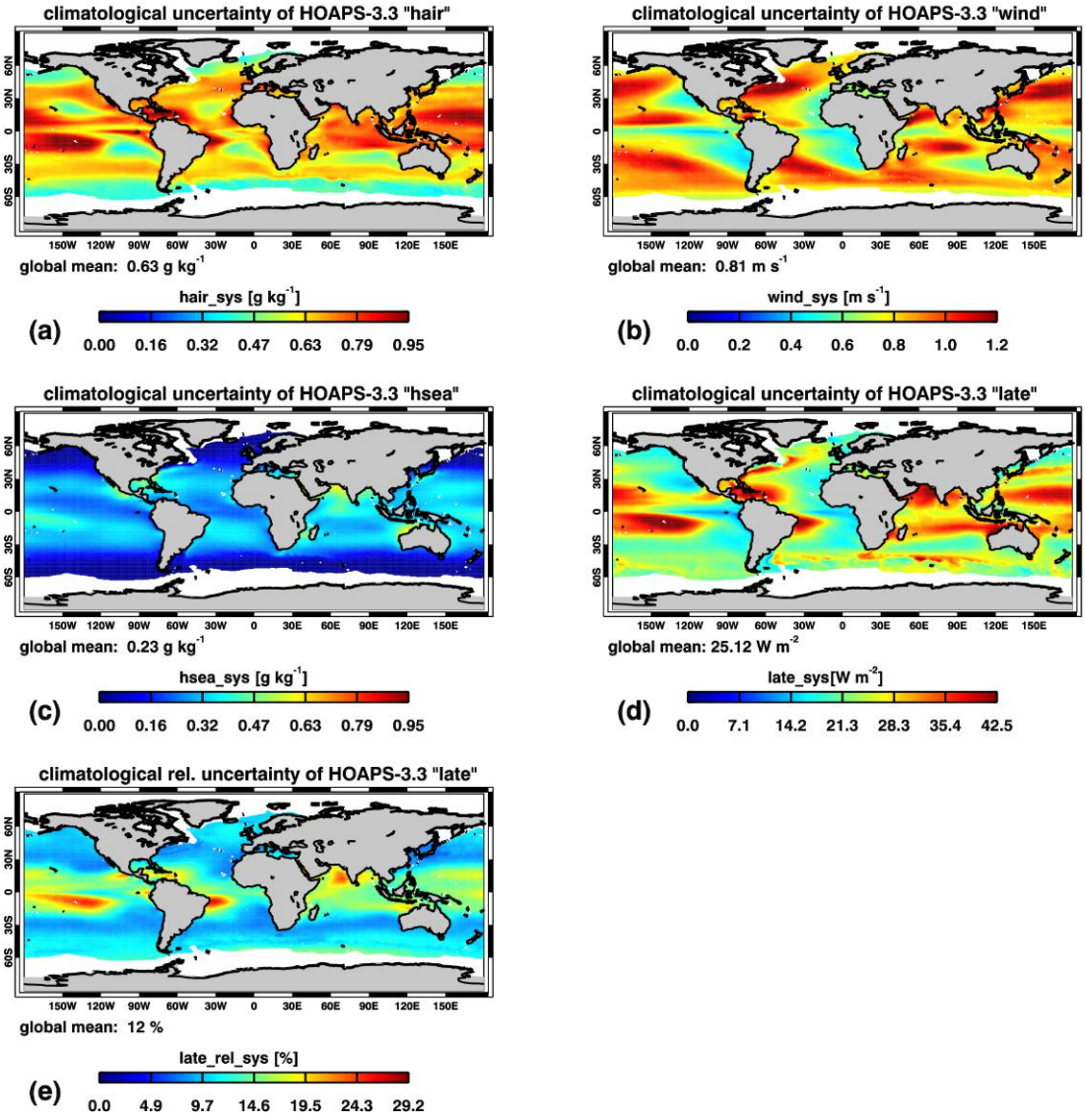


Figure 4. ~~Temporal averages (1988-2012) of~~ HOAPS-3.3 climatological total uncertainties (E_{clim}) of (a) q_a ("hair"), (b) U ("wind"), (c) q_s ("hsea"), and (d) LHF ("late"). E_{clim} is defined as the mean root mean squared sum of E_{sys}, E^{ran}_{retr}, and E_{smp} (1988-2012). (e) Climatological mean relative $E_{clim}(LHF)$ with respect to its natural variability. This variability is defined as the range between the 5th and 95th percentile of instantaneous LHF between 2000-2008. The global averages (text strings) were derived by considering a latitudinal cosine-dependency. All patterns result from the multi-dimensional bias analyses and subsequent uncertainty propagations described in Sect. 3.2 ~~-~~and 3.4. Note that the color bar ranges of (a) and (c) are identical to allow for direct comparisons.

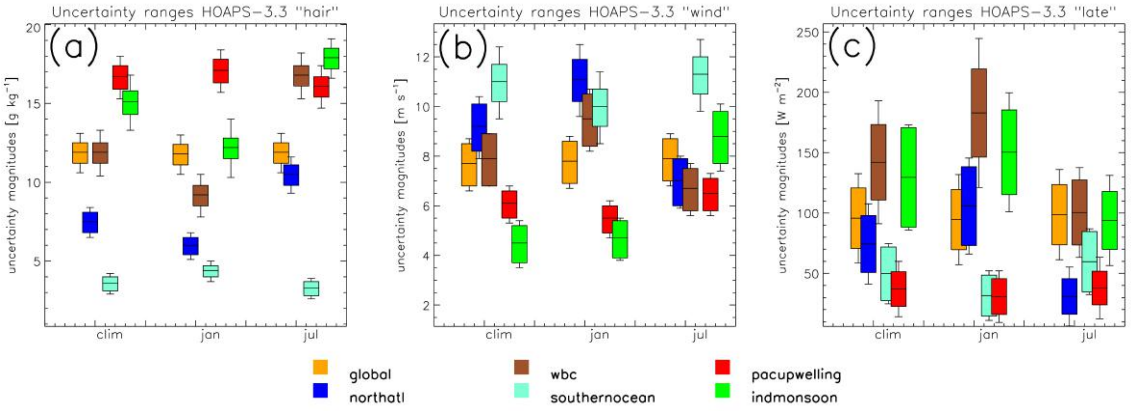


Figure 5. (a) Expected ranges of q_a ("hair") as a function of different regions and seasons~~, while considering both~~ . The color–coded boxes show E_{clim} ~~and~~ (1988-2012), whereas the bars indicate the average *instantaneous* random uncertainty component E_{retr}^{ran} (1988-2012). The following regions are presented: global (orange), North Atlantic (60° W–5° E, 35–65° N, dark blue), North Atlantic Western boundary current (WBC, 60–80° W, 30–40° N, brown), Southern Ocean (50–60° S, cyan), Pacific upwelling regime (80–100° W, 5° N–5° S, red), and Indian Monsoon region (50–75° E, 15–30° N, green). ~~Whereas the color–coded boxes show the expected systematic uncertainty, the bars indicate the random uncertainty component.~~ (b) As for (a), but for U ("wind"). (c) As for (a), but for LHF ("late").

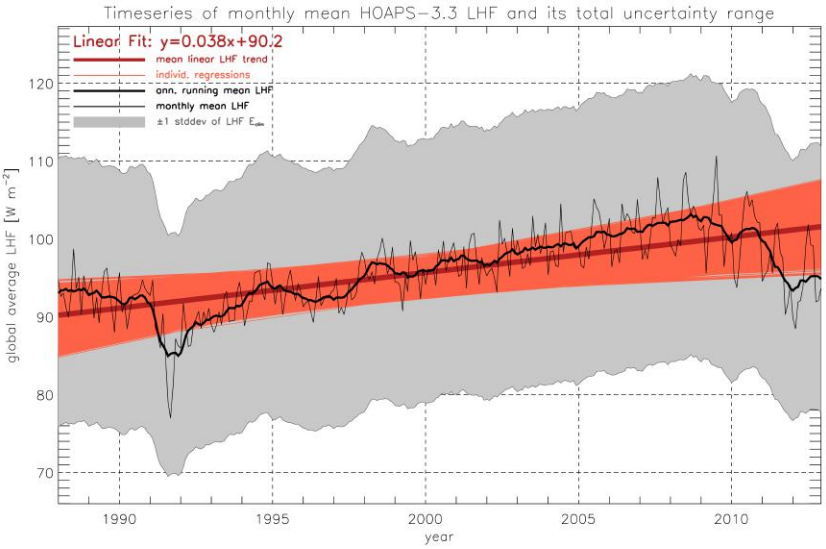


Figure 6. The thin (thick) black line shows the monthly (annual running mean) time series of HOAPS-3.3 LHF (70° S-70° N, cosine-weighted average). The dark red line illustrates the linear trend, which takes on a value of 4.5 W m^{-2} per decade (p<0.00001, based on a two-tailed t-test). The grey shading represents $\pm$ 1 standard deviation ("stddev") of the annual running mean E_{clim} (global average). The light red regression lines were iteratively derived following Kelly (2007) by taking $\pm$ 1 stddev of E_{clim} into account.