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: Marc Schröder, Email: marc.schroeder@dwd.de

Abstract. Latent heat flux (LHF) is 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, and sampling uncertainties, all of which are essential for assessing their quality. Here, the challenge is taken on by matching LHF-related pixel-level data of the Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite (HOAPS) climatology (version 3.3) to *in situ* measurements originating from a high-quality data archive of buoys and selected ships. Assuming the ground reference to be bias free, this allows for deriving instantaneous systematic uncertainties as a function of four atmospheric predictor variables. The approach is regionally independent and therefore overcomes the issue of sparse *in situ* data densities over large oceanic areas. Likewise, random uncertainties are derived, which not only include a retrieval component, but also contributions from *in situ* measurement noise and the collocation procedure. A recently published random uncertainty decomposition approach is applied to isolate the random retrieval uncertainty of all LHF-related HOAPS parameters. It makes use of two combinations of independent data triplets of both satellite and *in situ* data, which are analysed in terms of their pairwise variances of differences. Instantaneous uncertainties are finally aggregated, allowing for uncertainty characterisations on monthly to multi-annual time scales. 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 from 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. Regional and seasonal analyses suggest that largest total 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. 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. Specifically, latent heat fluxes (LHF) significantly control the surface energy budget and are, in addition to radiative fluxes, one of the main contributors to heating and cooling of the oceans. 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. Additionally, LHF modifies the atmospheric stability distribution and triggers convection, which in turn strongly impacts cloud formation and precipitation. 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 Global Energy and Water Cycle Experiment (GEWEX), and the Climate Variations (CLIVAR) Science Steering Group (e.g. Curry et al., 2004). 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 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).

Several LHF data records exist, which differ in instrumentation, creation process, data density, as well as spatial and temporal extent. These are either based on *in situ* measurements, reanalysis, remotely sensed data, or a merged version of these. Apart from isolated direct *in situ* measurements using e.g. sonic anemometers, all data methods share a need of bulk flux algorithms 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.

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 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). 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, 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). Thus, atmospheric reanalysis suffer from problems in their freshwater budgets (e.g. Schlosser and Houser, 2006; Trenberth et al., 2007).

Similarly, remotely sensed LHF climatologies are also prone to uncertainties. In addition to calibration uncertainties and aliasing problems (Bentamy et al., 2003), uncertainty sources either originate from uncertainties in the parameterization (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 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). 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 date, disagreements and/or weaknesses in data sets are often revealed by performing intercomparison studies, such as those presented by Kubota et al. (2003), 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 that comprehensive uncertainty estimates are valuable for climate research purposes. To date, none of the above-listed, satellite-based data records are accompanied by LHF-related uncertainty estimates, which hampers a quality assessment of the air–sea fluxes and related parameters. Such uncertainty assessments go beyond conventional LHF intercomparison studies, 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 (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 (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. 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 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 specifying systematic, random retrieval, and sampling uncertainties 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.

Whereas Sect. 2 introduces the data sets, Sect. 3.1 describes the procedure of matching HOAPS pixel-level data to *in situ* records (*double collocation analysis*). This results in estimates of systematic uncertainties of LHF-related parameters, assuming the ground reference to be bias free. It is assumed that these biases depend on the unique combination of four atmospheric predictor variables (q_a, U, sea surface temperature, and vertically integrated water vapour), all of which are observed simultaneously from space. The results from double collocation are then binned as a function of these four state variables (regionally independent *multi-dimensional bias analysis*, Sect. 3.2), resulting in bin-wise mean systematic uncertainties and, owing to their spread, random uncertainties. The random uncertainty estimates are not only related to the satellite retrieval, but also include contributions from the *in situ* source as well as the spatial and temporal matching. They can be decomposed into individual uncertainty components (*random error decomposition*, Sect. 3.3) following the method published in Kinzel et al. (2016). The approach is based on two combinations of data triplets originating from three independent sources (both HOAPS and in situ data, *multiple triple collocation*), which are evaluated in terms of their variances of differences and permit the isolation of the required retrieval-related uncertainty component. Rigorous error propagation to the instantaneous LHF-related data is performed subsequently, which allows to quantify both systematic and random retrieval uncertainties of LHF themselves (Sect. 3.4). Aggregating these instantaneous uncertainty measures allows for presenting monthly to multi-annual uncertainty distributions. Specifically regarding monthly mean sampling uncertainties (Sect. 3.5), the approach by Tomita and Kubota (2011) is employed. All uncertainty components are presented in Sect. 4, which includes regional and seasonal differentiations. Section

4 also comprises a trend analysis applying the derived uncertainty estimates. 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. The described sequence particularly allows for assigning LHF-related systematic and random uncertainties to instantaneous HOAPS-3.3 satellite data, which is not available for any other satellite data record to date. It extends the procedure described in Kinzel et al. (2016), as it is not restricted to q_a-related uncertainties, presents aggregated uncertainty distributions, and (next to random uncertainties) captures both systematic and sampling components.

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 HOAPS-3.3, 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. 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). 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 brightness temperatures (TBs) and ship data between 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 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 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.

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-3.0a by Fairall et al. (2003). It includes atmospheric stability calculations, which necessitate surface air temperatures as input. 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). 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 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~~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. In a first step, HOAPS LHF-related pixel-level records are matched to DWD-ICOADS measurements (*double collocation analysis*, Sect. 3.1). Assuming the ground reference to be bias free, this allows for investigating the systematic uncertainty structure as a function of four atmospheric state variables, namely q_a, U, sea surface temperature, and vertically integrated water vapour (*multi-dimensional bias analysis*, Sect. 3.2). Resulting random uncertainties, however, are not exlusively satellite-related, as they include contributions from *in situ* measurement noise and collocation. They can be corrected for, following the recently published approach of Kinzel et al. (2016) (*random uncertainty decomposition*, Sect. 3.3). The method is based on two combinations of independent data triplets including both pixel-level HOAPS-3.3 data as well as *in situ* records, which are analysed in terms of their variances of differences. In consequence, all HOAPS LHF-related instantaneous data is equipped with both systematic and random retrieval uncertainty estimates, which can be aggregated for gridding purposes and displayed as e.g. monthly or multi-annual means. When aggregating, sampling uncertainties additionally become important. However, it will be shown that they receive considerably less weight compared to the systematic uncertainty measures (Sect. 3.5). The sequence of analyses allows for a complete HOAPS-3.3 uncertainty characterization of LHF-related parameters on various time scales (Sect. 4), which goes beyond what has been published on LHF-related climatologies to date.

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 vertically integrated water vapour ("wvpa"), resulting from the double collocation analyses. Overall, 13.8 million matchups contribute to each subplot. 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 values between 7-12 g kg^{-1} , 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 for polar (< 5 g kg^{-1}) and subtropical (12–17 g kg^{-1}) humidity regimes. The latter 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). 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-)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 from artificial biases by illustrating dq_a as a function of *in situ* q_a, U, and 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. However, we are aware of these pseudo biases due to errors in the *in situ* records (e.g. Stoffelen, 1998) , specifically in the tail regimes, which consequently leads to an increase of the

HOAPS uncertainty estimates presented in Sect. 4. Two-sided regression analyses could further reduce these spurious biases, which are envisaged for future HOAPS uncertainty characterizations.

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* 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 (Fig. 2, x-axes) represent one dimension.

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, the uncertainties are assumed to depend on the concurrent atmospheric state. The combination of q_a, U, SST, and 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. 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~~ is therefore geophysically motivated, but implemented in a statistical manner. 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 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.

3.3 Random Uncertainty Decomposition

The *total* random uncertainties introduced in Sect. 3.2 (and 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 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 (Kinzel et al., 2016) . 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}, while explicitly accounting for error correlation terms, 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 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 required random satellite retrieval uncertainty, which is derived for all 20 bins as a function of U, q_s, and q_a.

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 from 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'.

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:

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

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}.

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

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

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

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 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. 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. 3.1 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). 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 0.7 and 1.8 g kg^{-1}, where minima (maxima) are found 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 from 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 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. 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. The magnitudes presented in Figure 3a are below those shown in Fig. 2a, as the random uncertainties have been corrected for the impact of $E_{ins}(q_a)$ and $E_c(q_a)$ (Sect. 3.3). 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 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. 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. 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. For each grid box, this variability is derived as the difference between the 5th and 95th percentile of instantaneous LHF observations between 2000–2008 (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 Global Patterns of HOAPS-3.3 Climatological Uncertainties

Figure 4 shows the distribution of the climatological uncertainties (E_{clim}) 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 from E_{retr}^{ran} and 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. 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 $E_{clim}(U)$ shown in Fig. 4b 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 exception 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 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~~ 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, that is 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 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 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 from $E_{clim}(q_a)$ 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 from $E_{clim}(U)$ 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. 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 HOAPS-3.3 Uncertainty Analyses

Global mean E^{ran}_{retr} and E_{clim} of LHF-related HOAPS-3.3 parameters are fairly constant in time throughout the whole climatology (Figs. 3-4). Absolute deviations from the global mean LHF (q_a, U) uncertainty become as large as 18 % (3 %, 8 %). 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, 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^{ran}_{retr}. 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. 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}, $E^{ran}_{retr}(q_a)$ 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 $E^{ran}_{retr}(LHF)$ are confined to boreal autumn 1991, taking on a mean value of 20 W m^{-2} and 33 W m^{-2}, respectively. These estimates are 20 % and 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 multi-annual (1988-2012) means of systematic uncertainty contributions, that is E_{clim}. At the same time, the error bars indicate the *instantaneous* random uncertainty components, that is E^{ran}_{retr}. 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 and 4.

Figure 5a indicates that the total 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 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 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 (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). 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)), is in the order of 10 %, which is in line with findings of Santorelli et al. (2011) and those illustrated in Fig. 6 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. 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 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 from 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. 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 considered 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 and 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. 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. 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.

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 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, https://doi.org/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, https://doi.org/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.

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, https://doi.org/10.1175/1520-0442(2003)016<0637:SEOWSA>2.0.CO;2, 2003.

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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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 Valudation of Satellite Sea Surface Skin Temperature Measurements for Climate Research, J. Climate, 15, 353–369, https://doi.org/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, https://doi.org/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, https://doi.org/10.1175/1520-0442(2003)016<0571:BPOASF>2.0.CO;2, 2003.

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, https://doi.org/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, https://doi.org/10.1175/1520-0426(1996)013<0246:SBTCFI>2.0.CO;2, 1996.

Gleckler, P. J. and Weare, B. C.: Uncertainties in Global Ocean Surface Heat Flux Climatologies Derived from Ship Observations, J. Climate, 16, 2764–2781, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/10.5270/OceanObs09.pp.19, 2010.

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, https://doi.org/10.5194/amt-3-1217-2010, 2010.

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, https://doi.org/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, https://doi.org/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, https://doi.org/10.1007/978-94-007-0332-2_6, 2011.

Kelly, B. C.: Some Aspects of Measurement Error in Linear Regression of Astronomical Data, The Astrophysical Journal, 665, 1489–1506, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/10.1023/A:1015845321836, 2002.

Kubota, M. K., Kano, A., Muramatsu, H., and Tomita, H.: Intercomparison of Various Surface Latent Heat Flux Fields, J. Climate, 16, 670–678, https://doi.org/10.1175/1520-0442(2003)016<0670:IOVSLH>2.0.CO;2, 2003.

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, https://doi.org/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, https://doi.org/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.

McClain, E. P.: Global sea surface temperatures and cloud clearing for aerosol optical depth estimates, International Journal of Remote Sensing, 10, 763–769, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/10.1175/2007JTECHO542.1, 2008.

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., https://doi.org/10.1002/joc.4150, 2014.

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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/10.1029/2010JC006927, 2011.

Schlosser, A. and Houser, R.: Assessing a satellite-era perspective of the global water cycle, J. Climate, 20, 1316–1338, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/10.1029/97JC03180, 1998.

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, https://doi.org/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, https://doi.org/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, https://doi.org/10.1175/JHM600.1, 2007.

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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/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, https://doi.org/10.1109/TGRS.2009.2023982, 2010.

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, https://doi.org/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, https://doi.org/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, https://doi.org/10.3402/polar.v30i0.10167, 2011.

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 %

Table 2. 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 4.3.

# 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

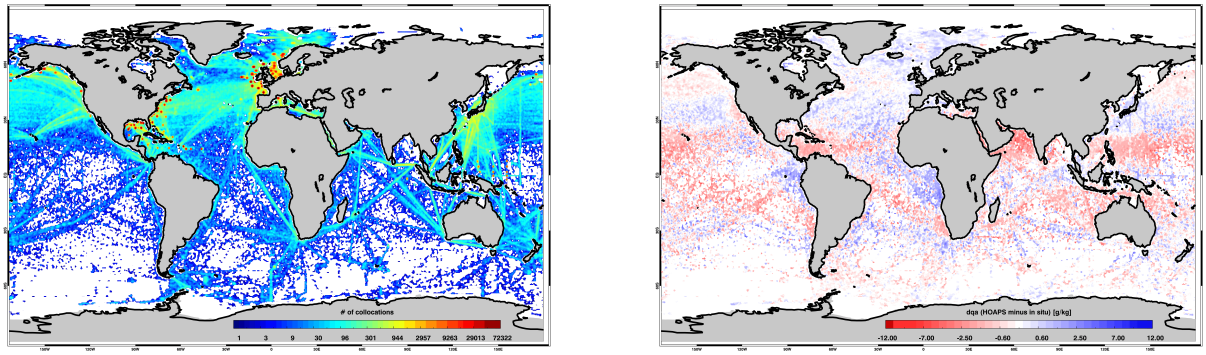

Figure 1. *Left panel*: Global map showing the distribution of collocated q_a measurements (HOAPS versus high-quality *in situ*) between 2001-2008. Overall, more than 13.8 million matchups contribute to this density map. Note that the colorbar is logarithmic. *Right panel*: Two-dimensional illustration 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.

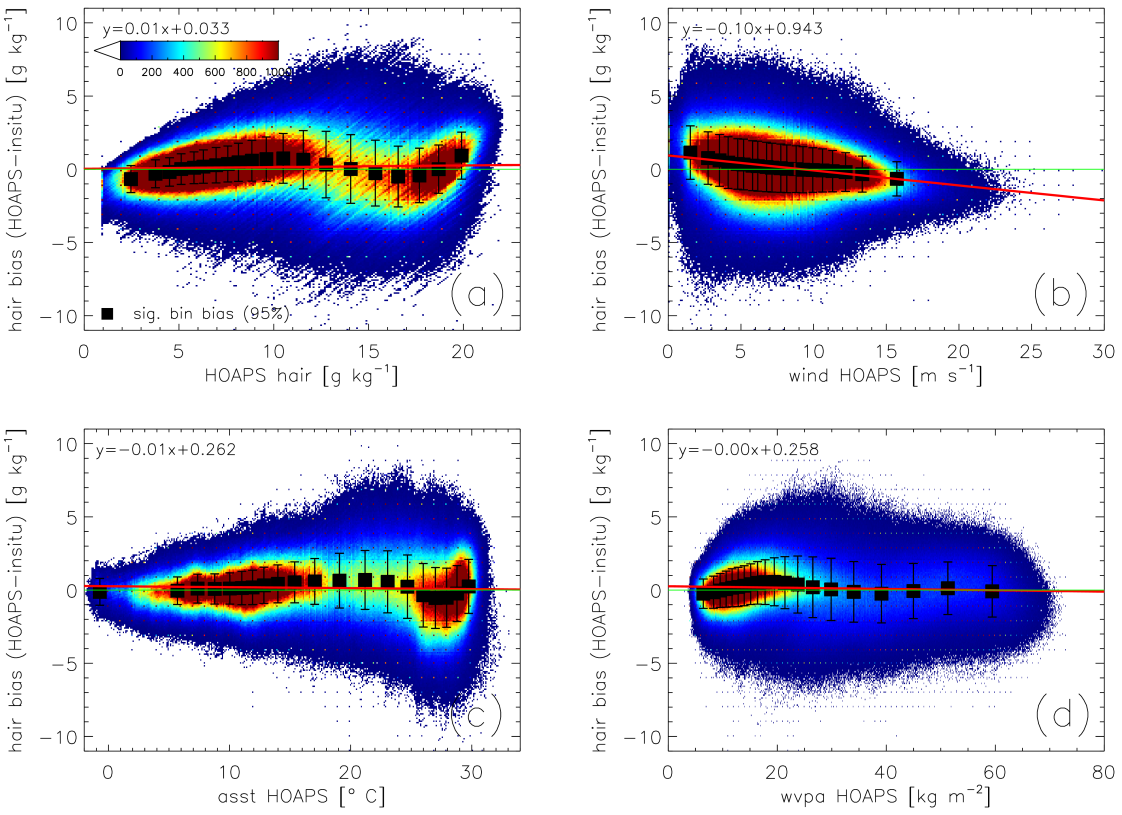

Figure 2. Scatter density plots of q_a bias (HOAPS-3.3 minus *in situ*, g kg^{-1}) as a function of (a) q_a ("hair"), (b) U ("wind"), (c) SST ("asst"), and (d) vertically integrated water vapour ("wvpa"), based on global double collocations between 2001 and 2008. The black squares and error bars represent 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 from 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).

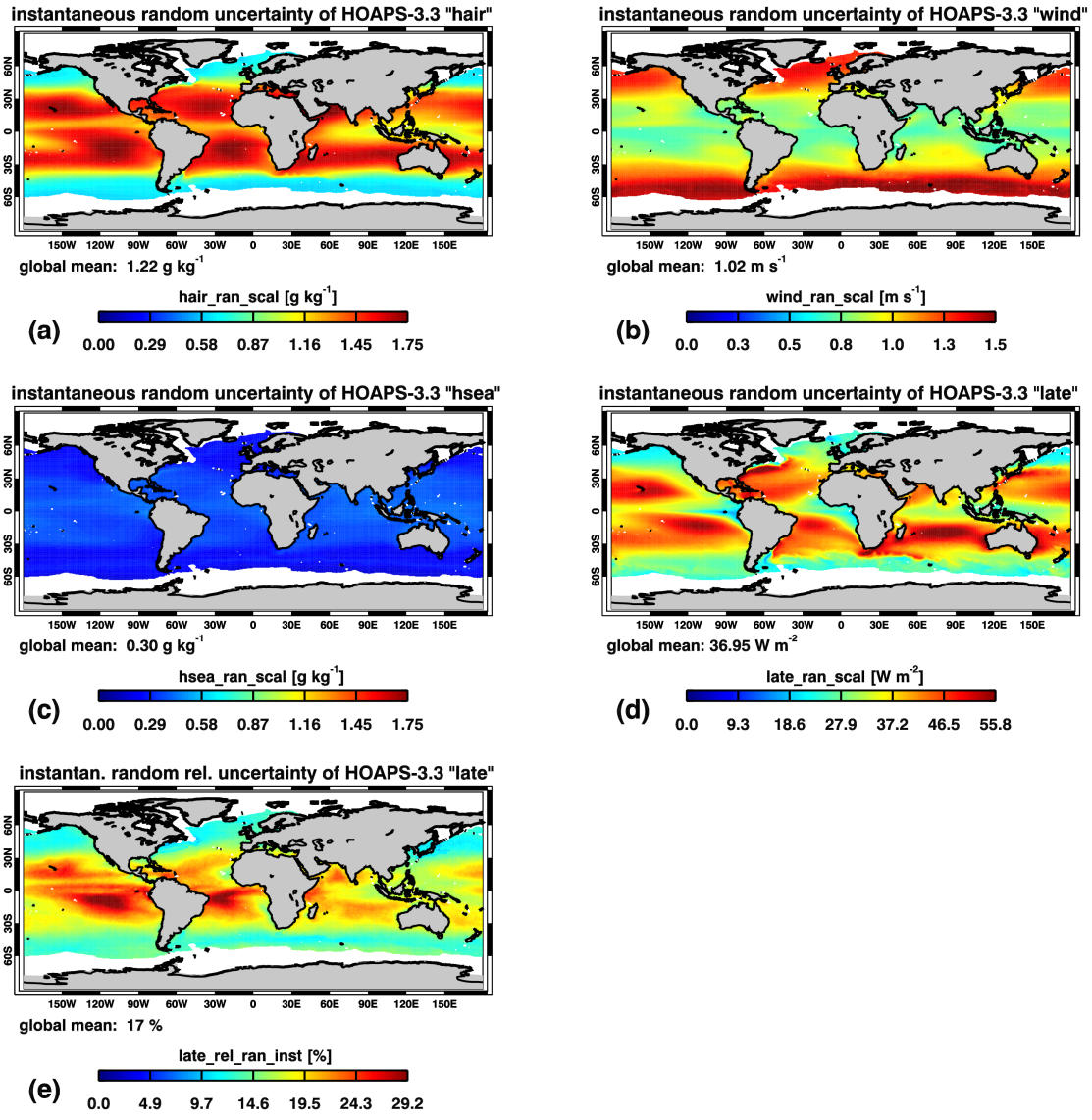

Figure 3. Temporal averages (1988-2012) of HOAPS-3.3 instantaneous E_{retr}^{ran} of (a) q_a ("hair"), (b) U ("wind"), (c) q_s ("hsea"), and (d) LHF ("late"). (e) Relative random retrieval uncertainty of HOAPS-3.3 LHF with respect to its natural variability. This variability is defined as the range between 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. 3.2-3.4. Note that the color bar ranges of (a) and (c) are identical to allow for direct comparisons.

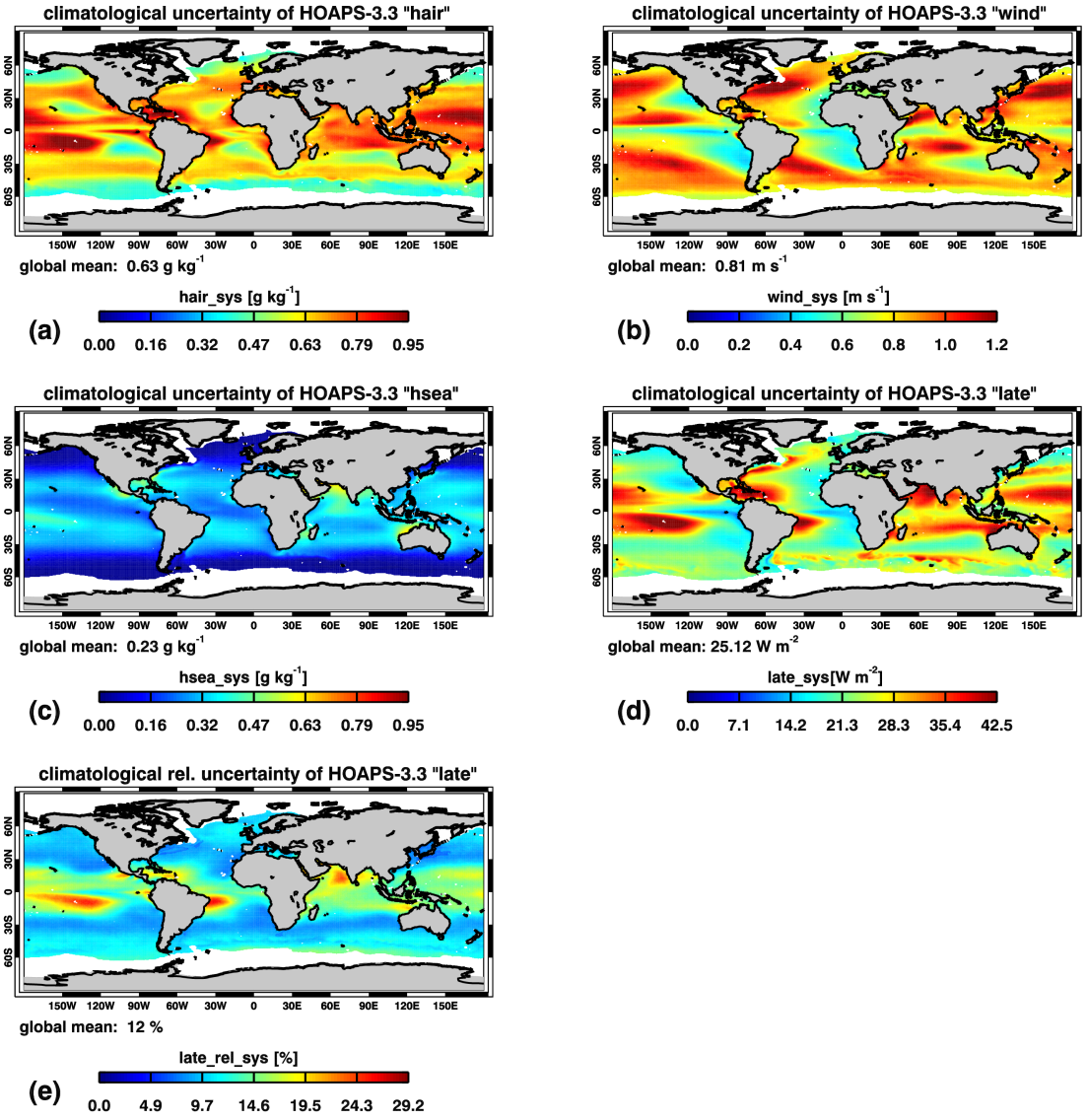

Figure 4. HOAPS-3.3 climatological total uncertainties (E_{clim}) of (a) q_a ("hair"), (b) U ("wind"), (c) q_s ("hsea"), and (d) LHF ("late"). E_{clim} is defined as the mean root mean squared sum of E_{sys}, E_{retr}^{ran}, and E_{smp} (1988-2012). (e) Climatological mean relative $E_{clim}(LHF)$ with 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.

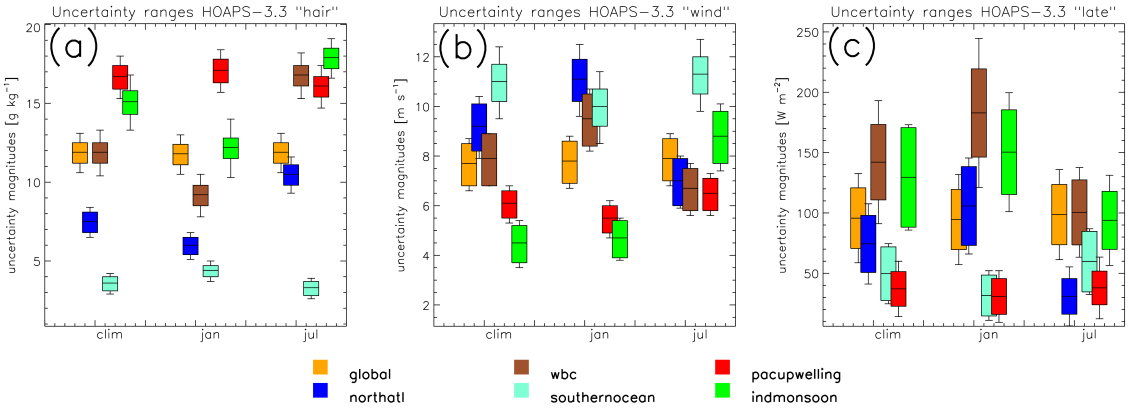

Figure 5. (a) Expected ranges of q_a ("hair") as a function of different regions and seasons. The color–coded boxes show E_{clim} (1988-2012), whereas the bars indicate the average *instantaneous* random uncertainty component E_{retr}^{ran} (1988-2012). The following regions are presented: global (orange), 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). (b) As for (a), but for U ("wind"). (c) As for (a), but for LHF ("late").

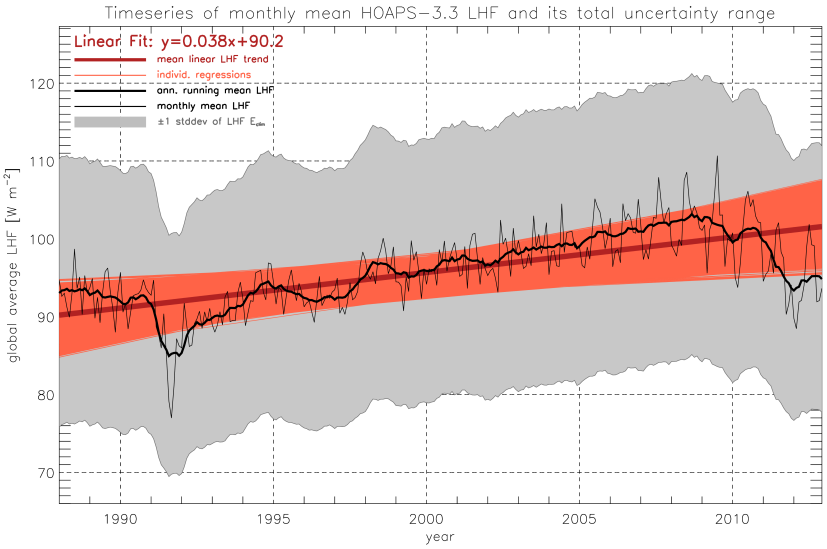

Figure 6. The thin (thick) black line shows the monthly (annual running mean) time series of HOAPS-3.3 LHF (70° S-70° N, cosine-weighted average). The dark red line illustrates the linear trend, which takes on a value of 4.5 W m^{-2} per decade (p<0.00001, based on a two-tailed t-test). The grey shading 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.