AMERICAN
METEOROLOGICAL
SOCIETY

Journal of Atmospheric and Oceanic Technology

EARLY ONLINE RELEASE

This is a preliminary PDF of the author-produced manuscript that has been peer-reviewed and accepted for publication. Since it is being posted so soon after acceptance, it has not yet been copyedited, formatted, or processed by AMS Publications. This preliminary version of the manuscript may be downloaded, distributed, and cited, but please be aware that there will be visual differences and possibly some content differences between this version and the final published version.

The DOI for this manuscript is doi: 10.1175/JTECH-D-15-0122.1

The final published version of this manuscript will replace the preliminary version at the above DOI once it is available.

If you would like to cite this EOR in a separate work, please use the following full citation:

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

É 2016 American Meteorological Society

latex .tex file
Click here to download LaTeX File (.tex, .sty, .cls, .bst, .bib) manuscriptKinzel.tex

AMERICAN METEOROLOGICAL SOCIETY 1919
AGRICULTURE · ENGINEERING · INDUSTRY · COMMERCE · AERIAL AND MARINE NAVIGATION · PUBLIC HEALTH

Decomposition of Random Errors Inherent to HOAPS-3.2 Near-Surface Humidity Estimates Using Multiple Triple Collocation Analysis

Julian Kinzel* and Karsten Fennig and Marc Schröder and Axel Andersson

Deutscher Wetterdienst, Satellite-based Climate Monitoring, Offenbach Germany.

Karl Bumke

GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany.

Rainer Hollmann

Deutscher Wetterdienst, Satellite-based Climate Monitoring, Offenbach, Germany.

PRELIMINARY ACCEPTED VERSION

**Corresponding author address:* Julian Kinzel, Deutscher Wetterdienst, Frankfurter Strasse 135, 63067 Offenbach, Germany.

E-mail: julian.kinzel@dwd.de

ABSTRACT

Latent heat fluxes (LHF) play an essential role in the global energy budget and are thus important for understanding the climate system. Satellite-based remote sensing permits a large-scale determination of LHF, which, amongst others, are based on near-surface specific humidity q_a. However, the q_a random retrieval error (E_{tot}) remains unknown. Here, a novel approach is presented to quantify the error contributions to pixel-level q_a of the Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite (HOAPS, version 3.2) dataset. The methodology makes use of multiple triple collocation (MTC) analysis between 1995-2008 over the global ice-free oceans. Apart from satellite records, these datasets include selected ship records extracted from the Seewetteramt Hamburg (SWA) archive and the International Comprehensive Ocean-Atmosphere Data Set (ICOADS), serving as the *in-situ* ground reference. The MTC approach permits the derivation of E_{tot} as the sum of model uncertainty E_M and sensor noise E_N, while random uncertainties due to *in-situ* measurement errors (E_{ins}) and collocation (E_C) are isolated concurrently. Results show an E_{tot} average of 1.1 $\pm$ 0.3 g kg^{-1}, whereas the mean E_C (E_{ins}) is in the order of 0.5 $\pm$ 0.1 g kg^{-1} (0.5 $\pm$ 0.3 g kg^{-1}). Regional analyses indicate a maximum of E_{tot} exceeding 1.5 g kg^{-1} within humidity regimes of 12-17 g kg^{-1}, associated with the single-parameter, multilinear q_a retrieval applied in HOAPS. Multi-dimensional bias analysis reveals that global maxima are located off the Arabian Peninsula.

1. Introduction

Besides short-wave and long-wave radiative fluxes, the heat transfer between ocean and atmosphere is composed of turbulent sensible (SHF) and latent (LHF) heat fluxes. On a global average, LHF represents the primary contributor for compensation of the ocean's energy gain by radiation fluxes over the ocean (Schulz et al. 1997) and hence for the closure of the surface energy budget. LHF considerably influences the oceanic heat balance and represents a vital source in terms of altering the atmospheric circulation and the overall hydrological cycle on seasonal to multi-decadal timescales (Chou et al. 2004). The understanding of the underlying physical processes crucially depends on the ability to accurately measure the ocean-surface heat fluxes. The latest assessment report of the Intergovernmental Panel on Climate Change (IPCC), for example, underpins the role of heat transfer between ocean and atmosphere in driving the oceanic circulation. It stresses that flux anomalies can impact water mass formation rates and alter oceanic and atmospheric circulation (IPCC 2013).

Thus, reliable long-term global LHF climate data records are needed to overcome this issue, serving as a verification source for coupled atmosphere-ocean general circulation models and climate analysis (Schulz et al. 1997). Similarly, LHF datasets represent a substantial input component to assimilation experiments, such as the oceanic synthesis performed by the German contribution to Estimating the Circulation and Climate of the Ocean (GECCO, e.g. Köhl and Stammer 2008).

Owing to a large spatial and interannual variability as well as spatial and temporal undersampling, Andersson et al. (2011) elucidate that *in-situ* LHF measurements remain troublesome over the global ocean. Conclusions within the AR5 assessment report (IPCC 2013) also mention the insufficient quality of *in-situ* observations when it comes to an assessment of turbulent heat flux changes. Although voluntary observing ships (VOS) provide the longest available *in-situ* record, Gulev et al. (2007) stress that VOS-based surface fluxes suffer from uncertainties associated with the ship observations, applied bulk aerodynamic algorithms, and the approach used to produce surface flux fields. Owing to this, random sampling uncertainties in LHF amount to several tens of W m^{-2} in poorly sampled high latitudes (Gulev et al. 2007).

Despite global coverage and high temporal resolutions, global atmospheric reanalyses have weaknesses as well as, e.g. associated with a lack of spatial detail (Winterfeldt et al. 2010). Reanalysis products are known to exhibit shortcomings in remote regions due to little *in-situ* ground reference data. In consequence, they are dominated by the atmospheric model (Gulev et al. 2007). In well-sampled regions, by contrast, the reanalysis fields are strongly constrained by observations.

In order to overcome the addressed issues, high-quality remote sensing datasets are of supplementary need. Several of these are currently available, incorporating LHF-related parameters. They comprise, for example, data of the Climate Goddard Satellite-Based Surface Turbulent Fluxes Version 3 (GSSTF3, Shie et al. 2012), the French Research Institute for Exploitation of the Sea (IFREMER, Bentamy et al. 2003), the Japanese Ocean Flux Data Sets with Use of Remote Sensing Observations (J-OFURO2, Kubota et al. 2002), the SeaFlux Version 1 dataset (Clayson et al. 2015), and the Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite (HOAPS) dataset (Andersson et al. 2010; Fennig et al. 2012). Their retrievals include a bulk aerodynamic algorithm to parameterize LHF in terms of observed mean quantities, i.e. bulk variables (e.g. Fairall et al. 2003).

HOAPS is a completely satellite-based climatology of precipitation, evaporation, related turbulent heat fluxes, and atmospheric state variables over the global ice-free oceans. The usefulness of the HOAPS climatology has been tested among numerous intercomparison studies and promising results have been published within Kubota et al. (2003), Bourras (2006), Klepp et al. (2008), Winterfeldt et al. (2010), and Andersson et al. (2011).

Bulk aerodynamic algorithms have a primary dependency on specific humidity q_a. Its accuracy directly impacts the uncertainty of the derived LHF. The Global Climate Observing System (GCOS 2010) has declared the near-surface specific humidity as an essential climate variable (ECV), indicating its prominent role in the context of climate analysis (Prytherch et al. 2014). However, the remote sensing of q_a remains challenging. The retrieval process is complicated, as the measured signal originates from relatively thick atmospheric layers (e.g. Schulz et al. 1997). Several studies have highlighted the importance of the uncertainties in q_a when investigating satellite-based LHF discrepancies (e.g. Andersson et al. 2011; Bentamy et al. 2013; Bourras 2006; Smith et al. 2011), implying a high potential for improvement. Furthermore, satellite validation analysis is per se difficult due to the lack of knowledge of the 'truth' (e.g. Zwieback et al. 2012) and the introduction of representativeness and collocation errors, owing to poor spatial coverages of *in-situ* measurements (Scipal et al. 2010).

In order to improve our understanding of uncertainties in satellite products, the triple collocation (TC) technique (e.g. O'Carroll et al. 2008) has been developed and applied. TC is based on three individual datasets and allows to isolate uncertainties of the underlying datasets. The set of equations resulting from such a single TC analysis permits to solve for a maximum of three unknown errors. However, the amount of random uncertainties inherent to the SSM/I instruments (model error E_M and noise error E_N) as well as the collocation procedure (random *in-situ* error E_{ins} and collocation error E_C) equals to four.

Within the framework of a random error characterization of HOAPS q_a, it will be demonstrated how to overcome this issue by extending the traditional TC analysis of O'Carroll et al. (2008) to a *multiple* TC (MTC), based on two triplets of SSM/I and *in-situ* records. This allows the decomposition of the overall random uncertainty in q_a into estimates of E_M and E_N. Their sum represents the random retrieval error E_{tot}. E_{ins} and E_C are quantified analogously. The results constitute a fundamental basis for a full error characterization of HOAPS LHF-related parameters, which will enhance HOAPS' analysis potential in future scientific studies.

Section 2 presents the applied data sources in more detail and introduces the MTC method. Section 3 shows results of the analyses, which include investigations of latitudinal and seasonal error dependencies as well as their hotspots. Findings are related to recent publications within Section 4, which also includes a qualitative comparison of the advantages and drawbacks of the applied data and the MTC approach.

2. Data and Methodology

a. Data

1) HOAPS-S Data Records

Apart from the sea surface temperature (SST), all HOAPS parameters are derived from inter-calibrated SSM/I (Special Sensor Microwave/Imager) passive microwave radiometers, which are installed aboard the satellites of the United States Air Force Defense Meteorological Satellite Program (DMSP). Therefore, HOAPS provides consistently derived global fields of freshwater flux related parameters, avoiding cross calibration uncertainties between different types of instruments. The current HOAPS version includes SSM/I records between 1987 and 2008, during which a total number of six instruments were in operational mode.

The SSM/I measurements are characterized by a conical scan pattern, where the antenna beam intersects the Earth's surface at an incidence angle of 53.1° and the swath width spans roughly 1400 km. The radiometers measure emitted and reflected thermal radiation from the Earth's surface and the atmosphere in form of upwelling microwave brightness temperatures (T_B's) at four different frequencies, namely 19.35 GHz, 22.2 GHz, 37.0 GHz, and 85.8 GHz. Whereas the 22.2

GHz channel only considers the vertically polarized signal, the remaining three channels measure both horizontal and vertical polarized signals (Hollinger et al. 1990). The channel footprints vary with frequency, ranging from elliptic 43x69 km^2 (cross-track/ along-track) at 19.35 GHz to rather circular 13x15 km^2 at 85.5 GHz. Each instrument completes one orbit within 102 minutes, implying that approximately 14 orbits per day are performed, allowing for 82% of global coverage between 87.5°S and 87.5°N within 24 hours. Due to the inclined orbit of the satellites, a spatial coverage of 100% is reached after three days.

Here, the focus lies on the HOAPS-S Version 3.2 data record (in the following HOAPS, Andersson et al. 2010; Fennig et al. 2012), which contains the HOAPS geophysical parameters in the SSM/I sensor resolution. HOAPS-S is based on a pre-release of the CM SAF SSM/I FCDR. Its extensive documentation, including Product User Manual, Validation Report, and Algorithm Theoretical Basis Document, is available online (Fennig et al. 2013). Compared to HOAPS-3, HOAPS-3.2 has been temporally extended until 2008 and is based on a reprocessed SSM/I 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). Starting with the most recent release (HOAPS-3.2), the HOAPS freshwater flux climatology is now 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 (MPI-M), Hamburg.

The HOAPS near-surface specific humidity 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). The underlying inverse model is based on linear regression between ship-based q_a and T_B, the former being linearly related to the integrated water vapor content. In comparison to earlier q_a model versions, considerable regional and seasonal biases were removed due to revised regression coefficients. Compared to Schulz et al. (1993, 1997), Bentamy et al. (2003) achieved a bias reduction of 15% and registered an overall root mean square error (RMSE) of 1.4 g kg^{-1} (originally 1.70 g kg^{-1}).

From 1995 onwards, records of up to three simultaneously operating SSM/I instruments are available (see Figure 2 in Andersson et al. 2010). As the MTC method relies on multiple SSM/I being in operational mode concurrently, the analysis is restricted to the time period from 1995 to 2008, excluding data prior to 1995 due to a comparatively poor *in-situ* data coverage.

2) SWA-ICOADS Ship Data Records

Hourly *in-situ* data originate from the marine meteorological data archive of the German Meteorological Service (DWD), supervised by the Seewetteramt Hamburg (SWA, part of DWD). It comprises global high-quality shipborne measurements as well as data provided by drifted and moored buoys. In case of data gaps within the SWA archive, the *in-situ* data basis was extended at SWA by available International Comprehensive Ocean-Atmosphere Data Set (ICOADS) measurements (Version 2.5, Woodruff et al. 2011). These records contain hourly global measurements obtained from ships, moored and drifting buoys as well as near-surface measurements of oceanographic profiles.

ICOADS 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). Depending on the period, the thermometers are also placed in aspirated and whirling psychrometers. q_a is eventually derived by applying the psychrometric formula. More information on VOS metadata and sensor types is given in Kent et al. (2007).

Several quality checks were performed at SWA prior to the merged SWA-ICOADS data's usage, which permitted a quality index assignment to each observation. The procedure is briefly described in the following.

To ensure a maximum degree of reliance, the SWA-ICOADS dataset underwent a flagging procedure based on a verification scheme. Investigated and possibly corrected features included a verification of the geographical position and, if given, the direction of travel. A subsequent calculation of the ship speed allowed for a consistency check of the spatial distances between subsequent measurements. Distances exceeding individually defined tolerance levels were discarded from further analysis. Next, climatological threshold checks were performed for the parameters air temperature, dew point temperature, sea surface pressure, SST, and wind speed. These thresholds were defined on the basis of the ERA-Interim dataset (Dee et al. 2011). Temporal outliers as well as repetitive values were identified and removed. Subsequently, inner consistency checks were carried out, which also involved an identification of unphysical relations between different parameters. In a final step, spatial checks were applied to aforementioned parameters to reject values which exceeded a maximum distance (individually defined for each parameter) to neighboring ship reports. The final outcome of all consistency checks was converted to internationally recognized quality flags (see standards defined by the World Meteorological Organization (WMO)).

Only ship records from the merged SWA-ICOADS database are selected for the subsequent analysis, in order to have a consistent, globally distributed data set as the ground reference. This decision is legitimate due to the vast amount of available *in-situ* measurements and prevents from blending data originating from different kinds of platforms. The approach of ship measurements (*in-situ*, as of now) as a ground comparison has been widely accepted and forms the basis of numerous other collocation analyses performed to date (e.g. Iwasaki and Kubota 2012; Jackson et al. 2006). To minimize their underlying error, only so-called 'special' (amongst others research vessels) and merchant vessels are extracted. Compare WMO (2013) for more information on the ship categorizations. In addition, only elements that appear to be correct (WMO Quality Flag 1) are considered during further analysis.

For comparison, MTC analysis using only buoy records was performed, which did not change the magnitudes of the decomposed random errors noteworthy (not shown). This conclusion may not apply to systematic uncertainties, suggesting the inclusion of buoy records when it comes to HOAPS bias analysis.

A height correction of the *in-situ* humidities to the HOAPS reference (10 m above sea level (ASL), assuming neutral stability) is not performed, although this could be done by means of VOS metadata (WMO 2013). The correction is not performed, as the introduced uncertainty, owing to the intermittent violation of the equivalent neutral stability assumption, may mask or even exceed the expected improvement associated with the bias correction. To qualitatively assess the impact of height adjustments of different complexity on E_{ins}, an investigation of collocated ship-based q_a values originating from match-ups of a subset of SWA-ICOADS and ERA-Interim data between 1995 and 2004 was carried out. An average ship-based q_a measurement height of 18 m was chosen (Kent et al. 2014). Over the Baltic Sea, which is representative for an extratropical ocean basin, the absolute q_a correction to 10 m results in an increase of only 0.1 $\pm$ 0.2 g kg^{-1} (full stability correction) (0.1 $\pm$ 0.1 g kg^{-1} (neutral stability correction)), performed on the basis of a turbulence algorithm without SST correction (Bumke et al. 2014). This correction-induced q_a increase lies within the uncertainty range suggested by Kent et al. (2014).

Indeed, Jackson et al. (2009) found an increase of q_a by more than 0.2 g kg^{-1} when comparing inversion-corrected AMMI (AMMIc) retrievals to original and subsequently to height-corrected ICOADS ship-based q_a. However, it led to an even larger bias of -0.29 g kg^{-1} (0.47 g kg^{-1}) and slightly larger RMSE in comparison to uncorrected *in-situ* measurements. This supports the argument that random variability is introduced by the height correction itself due to its dependency on the correction algorithm and associated (estimated) input bulk variables. Similar findings are published in Berry and Kent (2011), who argue that the height adjustment may be masked by natural variability of q_a (Figure 6 therein). A respective noise increase is also presented Prytherch et al. (2014). Kent and Berry (2005) show that the random error estimates are on average reduced by 8% (or 7%), if the full stability-dependent height correction is carried out (or assuming neutral stability). However, in comparison to the calculated total random error of 1.1 ± 0.1 g kg^{-1} published in Kent et al. (1999), this corresponds to an error reduction of just 0.1 g kg^{-1}. This finding, combined with those presented in Jackson et al. (2009) and Berry and Kent (2011), justifies the conservation of the original *in-situ* q_a within this work.

b. Previous publications involving TC

The need for TC-based error estimates related to different geophysical datasets was first realized by Stoffelen (1998), who suggested its application for the calibration of the European Remote-Sensing Satellite (ERS-1) scatterometer winds using wind speeds originating from the National Oceanic and Atmospheric Administration (NOAA) buoys and forecast model winds from the National Centers for Environmental Prediction (NCEP). Similarly, Caires and Sterl (2003) carried out TC analysis to validate significant wave height and wind speed fields from ERA-40 against altimeter measurements of buoys, ERS-1, and the Ocean Topography Experiment (Topex/Poseidon, NASA). Janssen et al. (2007) applied the TC method for wave height analyses. The introduction of the TC method into the field of satellite-based soil moisture research (Scipal et al. 2010) demonstrates the approach's potential for a wide range of applications.

The strategy of this study to apply MTC analysis to HOAPS q_a follows that of O'Carroll et al. (2008), who collocated data from the Advanced Along Track Scanning Radiometer (AATSR), Advanced Microwave Scanning Radiometer - Earth Observing System (AMSR-E), and buoy SST to successively derive the standard deviation of error on each observation type.

c. MTC methodology

The satellite error decomposition based on MTC analysis relies on match-ups of triplets involving both SSM/I and *in-situ* records. These triplets are created on the basis of conventional double collocation in a first step, resulting in paired match-ups of HOAPS and ship q_a records between 60°S and 60°N. The collocated pairs are based on the so-called nearest neighbor approach, i.e. HOAPS q_a pixels are assigned to respective ship observations closest in time and space.

Ship records and up to three simultaneously available SSM/I instruments eventually allow for performing MTC analysis. A setup sketch of the triplets contributing to the MTC is shown in Figure 1 (left panel). Triplets incorporating two independent ship measurements and one HOAPS pixel represent the first TC setup (top left panel, *V1* as of now), whereas a single ship record and two HOAPS pixels of independent SSM/I instruments form the second triplet structure (top right panel, *V2* as of now). In case of *V1*, match-ups incorporating two separate measurements obtained from the same vessel are excluded from further analysis. Although representing a major constraint in terms of amounts of available data, this approach ensures a complete independence of both *in-situ* records. Figure 1 (right panel) shows the distribution of the overall *V1* triplet amounts. Clearly, the *in-situ* data density is highest in mid-latitudinal, coastal regions.

Temporal and spatial collocation thresholds are set to 180 minutes and 50 km, following a statistical investigation by Kinzel (2013). For this, the author analyzed temporal decorrelation lengths of hourly ship q_a between 1995-1997, exemplarily for *R/V* Polarstern. The analysis was confined to the mid-latitudes, as these regions cover the tracks of extra-tropical storms, which are associated with largest fluctuations of LHF-related parameters in time (e.g. Romanou et al. 2006). Specifically for q_a, Kinzel (2013) obtained a temporal decorrelation scale of approximately six hours. Assuming an average ship speed of 15-20 km h^{-1}, this resulted in a spatial decorrelation scale of 90-120 km. These numbers are well above the chosen collocation thresholds.

As the representation of various atmospheric states should be the same for both *V1* and *V2*, TC *V2* triplets are only considered, if their ship record and either one of the participating HOAPS pixels contribute to *V1* as well.

Triplets including outliers are rejected from further analysis on the basis of 3σ standard deviation tests. Ship measurements within *V1* and *V2* represent the *in-situ* ground reference during this filtering process.

Subsequently, a bias correction with respect to the *in-situ* source is performed. Its importance for TC analysis is highlighted in e.g. O'Carroll et al. (2008). It implies that the results of the q_a error decomposition exclusively contain *random* uncertainties, as the systematic error is removed.

In preparation for the satellite error decomposition, the variances of differences between two data sources x and y, V_{xy}, are quantified, following O'Carroll et al. (2008):

$$V_{xy} = var(x) + var(y) - 2 \cdot cov(x,y). \quad (1)$$

That is, V_{xy} is given by the sum of the individual variances, corrected by the error covariance. In case the errors of x and y are not totally independent, respective covariance terms differ from zero and hence impact the satellite error decomposition.

At this stage, the MTC approach requires the assumption of an error model underlying every data source, which allows for expressing each term shown in Eq. (1) as a sum of supposedly contributing random errors. The following error model setup for ships (s) and satellites (sat) is formulated:

$$E_s = E_{ins}, \tag{2a}$$

$$E_{sat} = E_M + E_N. \tag{2b}$$

The collocation error (E_C) is neglected at this stage, as only those random error sources are listed in Eq. (2a) - (2b), which are *always* inherent to ship and satellite data.

Recall that E_{ins}, E_M, and E_N denote the random errors associated with the *in-situ* measurement, the satellite retrieval model as well as the sensor noise, respectively.

Given three independent data sources per TC version, Eq. (1) can be applied six times, requiring contributions of E_C. For this, the relative contribution of each data source to E_C does not need to be specified for the MTC application and is thus arbitrarily assigned to either Eq. (2a) or Eq. (2b) before utilizing Eq. (1).

On the basis of Eq. (2a) - Eq. (2b), the application of Eq. (1) yields the following variances of differences for TC *V1* (Eq. (3a) - Eq. (3c)) and TC *V2* (Eq. (4a) - Eq. (4c)):

$$V_{s1,s2} = 2(E_{ins})^2 + (E_C)^2, \tag{3a}$$

$$V_{s1,sat} = (E_{ins})^2 + (E_M)^2 + (E_N)^2 + (E_C)^2, \tag{3b}$$

$$V_{s2,sat} = (E_{ins})^2 + (E_M)^2 + (E_N)^2 + (E_C)^2, \tag{3c}$$

$$V_{s,sat1} = (E_{ins})^2 + (E_M)^2 + (E_N)^2 + (E_C)^2, \tag{4a}$$

$$V_{s,sat2} = (E_{ins})^2 + (E_M)^2 + (E_N)^2 + (E_C)^2, \tag{4b}$$

$$V_{sat1,sat2} = 2(E_N)^2 + (E_C)^2. \tag{4c}$$

E_M, E_N, and E_C are assumed to be satellite-independent. Regarding E_M, this is straightforward, as the exact same algorithm is applied to all SSM/I measurements to retrieve q_a. Concerning E_N, the SSM/I sensor sensitivities are shown in the aforementioned Validation Report (Figure 2 in Fennig et al. 2013). The referenced figure does not indicate a E_N-dependency on the instruments. As to E_C, the double and triple collocations rely on constant collocation criteria and the channel-dependent footprint sizes do not differ among the instruments.

Given the magnitude of V_{xy} on the left-hand side of Eq. (3a) - Eq. (4c), the individual random errors can be quantified successively. In order to solve Eq. (4c) for E_C, it is a prerequisite to calculate E_N synthetically by means of an arbitrary daily HOAPS-S record of T_B's. For this, a random Gaussian noise with zero mean and a variance equal to the channel noise is simulated and subsequently added to the daily T_B record. The assumption of Gaussian-distributed sensor sensitivities is widely accepted in literature and e.g. applied in Carsey (1992). E_N represents the standard deviation of the difference between the original and the synthetically derived q_a with a value of 0.3 g kg^{-1}. As E_N is a feature of the radiometer itself, it is independent of both platform and regime. Given E_C, E_{ins} is derived via Eq. (3a). Subsequently, both E_C and E_N suffice as input to solve Eq. (3b) - Eq. (4b) for E_M. The resulting arithmetic mean of all four solutions is assumed to be the best estimate of E_M. This is reasonable, as a separate analysis revealed that the standard deviations among the four E_M solutions are in the order of 0.02 g kg^{-1} to 0.18 g kg^{-1} , corresponding to only 1-16% of E_{tot} (not shown).

Due to the independence of the individual uncertainty components, the retrieval error E_{tot} results from:

$$E_{tot} = \sqrt{(E_M)^2 + (E_N)^2}, \tag{5}$$

which is dominated by E_M due to the relatively small E_N.

As expressed by Eq. (3a) - Eq. (4c), E_{tot} cannot be isolated using a *single* TC approach, i.e. a system of only three equations. This demonstrates the advantage of the applied MTC analysis regarding a successful decomposition of all random errors inherent to HOAPS q_a.

In preparation for applying Eq. (1) - Eq. (5), all triplets contributing to the MTC analysis are sorted in an ascending order (with respect to 'sat' in *V1* and 'sat1' in *V2*) and divided into 20 bins, respectively. All bins contain an equal amount of match-ups, whereas the amount contributing to *V1* differs from that of *V2*. Consequently, the bin widths are not constant, ranging from 0.37 g kg^{-1} to 1.86 g kg^{-1}. The uncertainty decomposition using Eq. (1) - Eq. (5), including the bias correction, is carried out separately for each bin. The resulting bin-dependent error magnitudes shown in Section 3a and Section 3b are arithmetic means of ten individual error decomposition analyses, whereby 30% of bin data are randomly drawn to derive V_{xy}, respectively. More precisely, the decomposition is based on 18005 triplets per *TC* version per bin.

3. Results of Random Error Decomposition

First, the focus lies on the q_a-dependent random uncertainty decomposition. To assess the regional dependency of the decomposed errors, a differentiation between tropics (0° - 30° N/S) and extratropics (30° - 60° N/S) is presented next. To investigate the temporal impact on the error statistics, winter (DJF), spring (MAM), summer (JJA), and autumn (SON) are considered separately. Furthermore, a multi-dimensional bias analysis approach helps to localize q_a uncertainty hotspots in space.

a. q_a-Dependent Random Error Decomposition

Figure 2 shows the result of the HOAPS q_a error decomposition as a function of q_a itself. The retrieval error E_{tot} (in red) converges to a minimum of approximately 0.7 g kg^{-1} for smallest q_a (relative uncertainty of 23%) and a global maximum partly exceeding 1.5 g kg^{-1} (relative uncertainty up to 13%) for q_a between 12-17 g kg^{-1}. Its global average value is given by 1.1 $\pm$ 0.3 g kg^{-1} (14% of relative uncertainty).

Due to the minor impact of E_N on E_{tot} (Eq. (5)), the satellite's retrieval model uncertainty E_M (shown in blue) closely resembles E_{tot} throughout the range of q_a and its mean is given by of 1.0 $\pm$ 0.3 g kg^{-1}.

The q_a error decomposition further reveals that E_C, shown in black, fluctuates around 0.5 g kg^{-1} for q_a below 10 g kg^{-1}, above which a positive trend causes E_C to maximize locally (0.7 g kg^{-1}) within a q_a regime of 14-17 g kg^{-1}. Its average value is given by 0.5 $\pm$ 0.1 g kg^{-1}, representing a relative uncertainty of 7%. In comparison to E_{tot}, the overall stability of E_C is noticeable and was to be expected, as the collocation criteria were kept constant. However, its maximum for q_a values of 14-17 g kg^{-1} indicates the largest uncertainty due to the collocation process and in consequence the MTC approach. This humidity regime is confined to rather narrow latitudinal bands over the subtropical ocean basins and extratropical fronts. These strong gradients point out the limits of the chosen collocation criteria. They become smaller in the vicinity of the equator, as is reflected in declining E_C for largest q_a.

Whereas 0.4 $\pm$ 0.1 g kg^{-1} represents the mean of E_{ins} (shown in yellow) for q_a below 10 g kg^{-1}, its average within (sub-) tropical surface humidity regimes is 0.9 $\pm$ 0.1 g kg^{-1}. In the inner tropics, it even exceeds E_{tot}. Overall, relative uncertainties range between 4-8%, emphasizing a linear relationship between *in-situ* measurement uncertainties and the magnitude of q_a. Its absolute average is given by 0.6 ± 0.3 g kg^{-1}.

The increase of E_{tot} from 0.7 g kg^{-1} in low-humidity regimes up to 1.8 g kg^{-1} close to 14 g kg^{-1} and its subsequent gradual decay is also mirrored in Figure 3, showing the bias of q_a (HOAPS minus *in-situ*) and its standard deviation as a function of HOAPS q_a. Accordingly, it is evident that these standard deviations, which are shown as black bars, maximize for q_a ranging between 12-14 g kg^{-1} ($\approx$ 2.3 g kg^{-1}), similar to Jackson et al. (2009) (their Figure 6b). The smallest spread of 1 g kg^{-1} occurs for q_a of 3 g kg^{-1}. As in Figure 2, the spread of the q_a bias clearly reduces to $\approx$ 1.7 g kg^{-1} (Figure 3) in tropical q_a regimes, implying a reduction in E_{tot}. The slope of the best-fit shown in Figure 3 is virtually zero, supporting the validity of the underlying retrieval model on a global scale. Yet, regime-dependent retrieval weaknesses exist. In contrast to E_{tot} in Figure 2, the bars shown in Figure 3 reflect the *overall* bin-dependent random uncertainty. Apart from the retrieval error E_{tot}, it also incorporates the effects of E_C and E_{ins}. This can be considered as a disadvantage in the representation of Figure 3 and again strengthens the information content resulting from the MTC analysis (Figure 2), which allows for a successive error decomposition. An accumulation of E_{tot}, E_C, and E_{ins} for the critical q_a range in Figure 2 results in an overall random uncertainty of 2.2 g kg^{-1} (i.e. E_{sum}), which closely resembles the observed equivalent of 2.3 g kg^{-1} in Figure 3.

Bentamy et al. (2013) and Roberts et al. (2010) demonstrate that their SSM/I q_a retrievals exhibit an explicit SST-dependency. The authors show that an inclusion of SST into their neural network (Roberts et al. 2010) and multi-parameter (Bentamy et al. 2013) approach considerably reduces the noise of q_a differences. To determine the overall impact of SST on the q_a retrieval error within the underlying work, a SST bias correction with respect to the *in-situ* data was performed and the analyses presented in Section 2c were repeated. The results indicate that E_{tot} is reduced by just 2% within the critical humidity regime between 12-17 g kg^{-1} (not shown), suggesting a multi-paramater approach to be of secondary importance in this q_a range. However, for small (3-5 g kg^{-1}) and large (18-20 g kg^{-1}) q_a margins, the retrieval uncertainty is on average reduced by 9% and 5%, respectively. SST-related q_a uncertainty hotspots (in an absolute sense) are found along the coasts of Western Australia and Northern Chile (SST $\approx$ 20°C) , where the total random q_a uncertainty associated with SST is up to 0.2 g kg^{-1}, i.e. $\approx$ 10% of the underlying total uncertainty (not shown).

b. Seasonal and Regional Random Error Decomposition

The distribution of E_{tot} (Figure 2) suggests that the underlying model for retrieving q_a exhibits both strengths (small q_a) and weaknesses (q_a between 12-17 g kg^{-1}), supporting the necessity of differentiating between different surface moisture and hence geographical regimes when it comes to q_a error decomposition. To highlight regional error dependencies, Figure 4 exemplarily confront time series of decomposed errors during boreal winter (DJF) within the extratropics (30° - 60° N/S, left panel) and the tropics (0° - 30° N/S, right panel). Table 1 summarizes all decomposed error magnitudes, along with their standard deviation and relative contributions (to the basin-mean q_a) as a function of region and season.

Focusing on the extratropics first (left panel), the average value of E_{tot} is 0.8 $\pm$ 0.1 g kg^{-1} (16% relative error). This order of magnitude is expected for an average q_a of 5.2 g kg^{-1} $\pm$ 0.4 g kg^{-1} (Figure 2). The overall uncertainty introduced by E_{ins} (by E_C) is given by 0.3 $\pm$ 0.1 g kg^{-1} (5% rel. error) (0.6 $\pm$ 0.1 g kg^{-1} (11% rel. error)). A closer look at the different seasons for extratropical latitudes (Table 1) indicates that retrieval errors maximize during boreal autumn (SON, 1.1 $\pm$ 0.1 g kg^{-1}, yet only 13% rel. uncertainty). E_{tot} associated with the largest average q_a during boreal summer months (JJA, 10.0 g kg^{-1}) remains 0.1 g kg^{-1} below the SON average. According to the constant increase in retrieval errors with increasing q_a, as illustrated in Figure 2, this was not to be expected. Strong positive outliers in boreal autumn E_{tot}, specifically in 1997 during the evolving El Niño event, may explain this feature (see further below for explanation). As also suggested by Figure 2, E_{ins} maximizes during boreal summer (JJA, 0.7 g kg^{-1}), along with the temporal q_a maximum in the course of a year. The local reduction in E_C for q_a values of 9-10 g kg^{-1}, as seen in Figure 2, is well represented in the seasonal analysis. Hence, E_C has a maximum of 0.7 g kg^{-1} in SON, whereas 0.6 g kg^{-1} are representative for extratropical boreal summer months.

Comparing extratropical error characteristics to the tropical counterpart (right panel) clearly demonstrates the retrieval error dependency on boundary-layer moisture content. During boreal winter (Figure 4, right panel), the average tropical retrieval uncertainty is given by 1.6 $\pm$ 0.2 g kg^{-1} (11% rel. error), where the average of q_a is 13.9 $\pm$ 0.8 g kg^{-1}. This humidity range corresponds to the moisture regime of largest retrieval discrepancies (Figure 2) and explains why E_{tot} is 0.2 g kg^{-1} to 0.4 g kg^{-1} larger in comparison to the remaining seasons. During boreal winter, *in-situ* (collocation) uncertainties are on average 0.8 g kg^{-1} (0.1 g kg^{-1}) larger in comparison to the extratropical counterpart, yet having relative contributions of only 7% (5%).

The regional confrontation of decomposed errors shown in Figure 4 and Table 1 clearly mirrors the error dependency on the q_a regime. In case of tropical latitudes, this goes along with interannual variability in error magnitudes, due to their pronounced sensitivity to q_a, as is illustrated in Figure 2.

In general, outliers within seasonal and regional time series could possibly be linked to strong El Niño and La Niña events, which are identified by means of the Oceanic Niño Index (Climate Prediction Center, NOAA), representing SST anomalies within the Niño-3.4 region (5°S - 5°N; 170°W - 120°W). Such a link may exist for the tropical boreal autumn in 2007 (E_{tot} 0.4 g kg^{-1} larger than seasonal average, not shown), associated with a moderate La Niña event. Anomalously low SSTs within the Niño-3.4 region, which are associated with these events, were already persistent during the preceding eight months. This supports the hypothesis that anomaly patterns may have propagated towards the Atlantic Ocean (where the *in-situ* data density is highest) via atmospheric planetary Rossby Waves and may have caused a q_a shift into humidity regimes associated with larger q_a retrieval uncertainties. This mechanism may also be attributed to the tropical boreal winter (1998) and the extratropical boreal autumn (1997) (E_{tot} being 0.2 g kg^{-1} larger than the seasonal averages, respectively), in line with the strong El Niño event established several months earlier. The effects of El Niño Southern Oscillation (ENSO) teleconnections on air-sea interaction on a global scale have been investigated by Alexander et al. (2002), for example.

c. Regional Random Uncertainty Hotspots

Figure 2 to Figure 4 demonstrate the behavior of the decomposed errors as a function of q_a only. To localize *true* hotspots of E_{tot} in space, however, the q_a-dependent error magnitudes shown in Figure 2 cannot simply be transferred to a global map, knowing only the average near-surface humidity distribution. The E_{tot} uncertainty pattern rather depends on the dominating sources of uncertainty, which are introduced by further atmospheric state variables. A specific region may for example be exposed to prevailing wind speeds, which enhance or dampen the E_{tot} illustrated in Figure 2.

To overcome this issue and hence capture the overall random q_a uncertainty as a function of the simultaneous atmospheric state, the analysis shown in Figure 3 is expanded by deriving q_a biases as a function of wind speed, SST, and water vapor path by means of double collocation (not shown). These three parameters are available from HOAPS and allow a distinction of different atmospheric regimes. As in Figure 3, this supplemental analysis results in bin-specific q_a biases. Given all four one-dimensional bias analyses, a four-dimensional bias look-up table is constructed, where the dimensions correspond to q_a, wind speed, SST, and water vapor path. Figure 5 (left panel) shows a sketch of this table in three-dimensional space. Subsequently, all instantaneous biases resulting from the double collocation procedure are assigned to one of the 20^4 = 160000 bins. If less than 100 bias values are assigned to a bin, its content is considered as non-representative and an interpolation is carried out along all dimensions. The overall random q_a uncertainty for every bin (equivalent to E_{sum} in Figure 2) is defined as the spread of all instantaneous biases underlying every bin. In a last step, these random uncertainties in q_a are corrected for the relative contributions of E_{ins} and E_C (bin-dependent, according to Figure 2) to exclusively focus on the random retrieval error E_{tot}. Applying all instantaneous HOAPS data to this four-dimensional random retrieval uncertainty table leads to a global q_a random retrieval uncertainty distribution, which is shown in Figure 5 (right panel) for 1995-2008. Its area-weighted global average is 0.82 g kg^{-1}.

As can be seen, largest retrieval uncertainties (with the exception of the global maximum off the Arabian Peninsula and India) are found along subtropical bands of both hemispheres, where they reach values up to 1.5 g kg^{-1}. More specifically, the maxima are located in regimes characterized by a mixture of trade and shallow cumulus with thin cirrus (Rossow et al. 2005; Oreopoulos and Rossow 2011), which seem to introduce an additional uncertainty within the q_a retrieval. At the same time, the average random retrieval error of q_a reduces towards the tropics, as is reflected in Figure 2 and Figure 3. Overall, the magnitudes are consistent with the total random uncertainties resulting from the error decomposition (Figure 2). This suggests that q_a itself has the largest influence on q_a-related E_{tot}, whereas the impacts of wind speed, SST, and water vapor path are of secondary order on a climatological scale.

The global q_a random uncertainty maximum within the Arabian Sea (up to 1.7 g kg^{-1}) is special, in as much as concurrent mean wind speeds remain below 5 m s^{-1} throughout most of the year (apart from boreal summer months, where monsoon-related wind speeds often exceed 12 m s^{-1}). Further analyses revealed that the spread of the q_a bias as a function of wind speed is largest for smallest wind speeds. This may be due to an enhanced decoupling of the vertical atmospheric column, introducing additional difficulties in the q_a retrieval, which could explain the amplification of the q_a-related E_{tot} in this region.

Summing up, the error characteristics show a clear regional (Figure 2, Figure 5 (right panel)) and seasonal (Figure 4, Table 1) dependency. Total uncertainties are especially large in subtropical latitudes (Figure 5, right panel), particularly during boreal winter (DJF), when q_a remains in a near-surface humidity range associated with largest q_a retrieval uncertainties (12-17 g kg^{-1}).

4. Discussion

a. q_a Retrieval Uncertainties

Figure 2 to Figure 4 suggest that the retrieval exhibits largest uncertainties for particular atmospheric and oceanic conditions. Possible explanations for this retrieval performance will be discussed in the following.

Note that all cited publications including RMSE estimates of q_a retrievals neither explicitly perform a bias correction with respect to the *in-situ* reference, nor have E_c and E_{ins} been removed. In consequence, the resulting random uncertainty estimates ($\widehat{=} E_{sum}$) exceed the *true* random retrieval error ($\widehat{=} E_{tot}$), which remains unknown. This highlights the benefit of the chosen MTC approach.

Numerous q_a retrievals have been presented to date and intercomparisons have been carried out in the past. The single-parameter, multilinear approach of Bentamy et al. (2003), which is used in HOAPS, considerably improved the accuracy of q_a in comparison to former attempts presented in e.g. Liu (1986). The latter took precipitable water as a proxy for the q_a retrieval. Revised regression coefficients within Bentamy et al. (2003), based on a more representative *in-situ* dataset, led to an average reduction in both q_a bias (15%) and its RMSE ($\approx$ 20%), favoring its successful implementation and/or tuning in further studies (e.g. Andersson et al. 2010; Jackson et al. 2009; Kubota and Hihara 2008).

A correlation coefficient of 0.96 between the integrated water vapor content (w) and the boundary layer humidity contribution (up to 500 m ASL) shown in Schulz et al. (1993) generally justifies the assumption of an underlying linear relationship between w and q_a. However, this linear relationship is challenged by Bourras (2006) (which in parts also applies to the algorithm of HOAPS), who elucidates two cases of vertical q_a profiles, where this linear dependency breaks down and in consequence introduces large errors in q_a. On the one hand, his considerations target the decoupling of the boundary layer moisture from higher atmospheric water vapor contents, which may be identified by means of local minima of vertical correlation profiles between both parameters. On the other hand, Bourras (2006) specifically addresses regions of deep convection and associated retrieval deficiencies (see also Bentamy et al. 2013), where the assumption of most water vapor being confined to the boundary layer is violated.

To overcome such retrieval errors, an inclusion of nonlinear terms within the retrieval algorithms, as presented in e.g. Jackson et al. (2009), can reduce the RMSE between remotely-sensed and *in-situ* records. Specifically, their AMMI (Advanced Microwave Sounding Unit (AMSU)-A and SSM/I) retrieval incorporates a quadratic term for the 52.8 GHz channel (not available in HOAPS). This channel not only provides somewhat more direct information on the lower troposphere, its quadratic weighting also allows for better describing the nonlinear relationship between lower tropospheric temperatures and water vapor.

Furthermore, Bentamy et al. (2013) argue that single-parameter, multilinear regressions may be too simple to capture the underlying physical mechanisms. The authors show that q_a seems to exhibit an explicit SST-dependency when investigating q_a biases between NOCv2.0 (Berry and Kent 2011) and SSM/I (their Figure 1). Including a SST- as well as a stability dependency (T_{air} minus SST) in their retrieval considerably reduces the noise (by up to 50%) of daily q_a differences (*in-situ* minus SSM/I) at 0.25° resolution on a global scale. Main discrepancies are confined to extratropical southern latitudes. Large-scale biases (dry tropics, wet subtropics), which were evident in former q_a retrievals, remain marginal within their multi-parameter approach.

Roberts et al. (2010) also picks up the influence of SST on the representativeness of the SSM/I retrieval output for q_a and presents a non-linear approach on the basis of a neural network. Applying SST as a first-guess input parameter to the retrieval and accounting for the regime-dependent effect of high cloud liquid water (CLW) on T_B's, the authors demonstrate that biases (RMSE) of q_a are reduced by 45% (27%) in comparison to e.g. Bentamy et al. (2003) (their Figure 5). The remaining bias (RMSE) is given by 0.16 g kg^{-1} (1.32 g kg^{-1}). Regarding the RMSE, its magnitude agrees with the average E_{sum} derived in this work (1.29 g kg^{-1}). Especially for very high CLW, the latter tends to effectively remove low-level humidity information from the satellite signal, which applies to most, yet not all compared satellite q_a datasets. Largest discrepancies between both approaches are evident for negative lapse rates (i.e. inversions) along with elevated moisture above 900 hPa. Similar conclusions involving the impact of inversions on T_B's are drawn in Jackson et al. (2006) (their Figure 3). Given traditional linear regression models, moist air masses aloft feign large boundary moistures and thus introduce large errors in T_B and consequently q_a. Roberts et al. (2010) present two case studies, for which the SST boundary condition is able to successfully distinguish inversion profiles from near-neutral or unstable stratifications. Regimes with damped SST associated with cold surface currents or upwelling regimes along with retrieval issues due to stratocumulus clouds (see Jackson et al. 2009; Smith et al. 2011) may be more effectively interpreted by their sophisticated retrieval. Furthermore, the authors demonstrate that warm SST in conjunction with high-level subsidence and hence little moisture (as frequently observed over the North Pacific during boreal summer within the descending branch of the Hadley Cell) do not necessarily lead to large biases in q_a, given their approach.

In order to further quantify q_a retrieval weaknesses, Iwasaki and Kubota (2012) developed two retrievals for estimating q_a using Tropical Rainfall Measuring Mission Microwave Imager (TMI) T_B data in comparison to ICOADS moored buoy data between 2003-2006. The essential difference between both linear retrievals was the amount of contributing TMI channels and thus their complexity. The authors show that their products yield a smaller RMSE specifically in the tropics, compared to those published in Schlüssel et al. (1995) (SSM/I), Kubota and Hihara (2008) (AMSR-E), and Schlüssel and Albert (2001) (TMI). The authors hold the inclusion of the 85 GHz polarized radiation for responsible, which is not included within the model of Bentamy et al. (2003) and hence HOAPS. This finding may be responsible for the negative bias along with largest RMSE within the subtropical high-pressure systems, which falls into critical q_a range of 12-17 g kg^{-1} (see Figure 3). Specifically for the subtropical highs, where CLW and rain rates remain small, the 85 GHz channels may include valuable boundary layer humidity information. However, one needs to keep in mind that their results are only representative for tropical regimes (due to the TMI orbit), in contrast to the approach of Bentamy et al. (2003).

Due to inherent deficiencies in single-sensor q_a retrievals (such as Bentamy et al. 2003), Jackson et al. (2006) and Jackson et al. (2009) elucidate the advantage of a multi-sensor approach, which, apart from SSM/I, utilizes temperature and humidity sounders (AMSU-A and SSM/T-2, respectively). Aiming at better evaluating the lower-tropospheric temperature and moisture characteristics, the authors reduce the RMSE differences (in comparison to ICOADS VOS and buoy measurements) by up to 0.4 g kg^{-1}, compared to single-sensor retrievals. This approach introduces additional information provided by the microwave sounders for q_a ranges of 16-20 g kg^{-1} and regimes of very low moisture content.

Prytherch et al. (2014) recently published results of an intercomparison involving different SSM/I-based q_a datasets and identified considerable discrepancies among the data records, where regional variations exceed 1 g kg^{-1} on an annual basis, despite relying on the same retrieval algorithm. Hence, differences among HOAPS, GSSTF3, and IFREMER, all of which rely on the algorithm of Bentamy et al. (2003), are bound to originate from varying data processing routines, intercalibration techniques, and quality controls. The different handling of hydrometeor contamination of the signal as well as humidity inversions are two procedures within these filtering routines, which introduce departures among the resulting q_a. In contrast to IFREMER e.g., HOAPS includes a humidity inversion correction, which is possibly the reason for the former being low-biased within regimes of smallest absolute q_a (Figure 9b within Prytherch et al. 2014). On the other hand, effects of inter-satellite calibrations on the T_B's may explain discrepancies among q_a based on HOAPS (intercalibration performed) and IFREMER (not subject to intercalibration).

b. In-Situ Uncertainties

Kent and Berry (2005) recall that VOS observations contain significant uncertainties and are of variable quality. They estimated random measurement errors in VOS between 1970-2002 using a semi-variogram approach, based on the ICOADS dataset (Woodruff et al. 1998). Their Figure 1d shows global maps of the uncorrelated uncertainty component of q_a averaged over the whole time frame. The spatial distribution of random variability components ranges between 0.7 ± 0.1 g kg^{-1} (Extratropical North Atlantic) to 1.7 ± 0.4 g kg^{-1} (near the Arabian Peninsula). A further investigation of latitudinal error dependencies in Kent and Berry (2005) indicates that the random error component constitutes the largest part of the total observational error within tropical regions. In contrast, the sampling error becomes considerably more important within the extratropics. These results imply that the random error component increases from larger (small q_a) to lower (large q_a) latitudes, as is also seen within Figure 2, with the exception of the inner tropics (Section 3).

The estimates published in Kent and Berry (2005) for the lower q_a boundary closely resemble the *in-situ* errors shown in Figure 2, given that most of the match-ups below 10 g kg^{-1} are constrained to extratropical northern latitudes along major shipping lanes (Bentamy et al. 2003). As discussed in Section 3a, moister regimes are subject to larger random *in-situ* errors, which agrees with results published in Kent and Berry (2005). Yet, their average random error in q_a is 1.1 $\pm$ 0.1 g kg^{-1}, which is $\approx$ 0.5 g kg^{-1} larger than the average estimate in this study (0.6 $\pm$ 0.3 g kg^{-1}). This discrepancy may be due to the strict filtering of non-appropriate ship records prior to the MTC analysis. Furthermore, the amount of contributing match-ups displayed in Kent and Berry (2005) is considerably lower than the collocated triplets forming the basis of this work. Additionally, Figure 1 in Kent and Berry (2005) includes *in-situ* data of 32 years. The *in-situ* quality in early years is likely to have been below today's measurement accuracies and particularly below the quality standard chosen for this study.

Kent and Taylor (1996) and Berry et al. (2004), amongst others, investigated the impact of solar radiation on the uncertainty of ship-based q_a. In this context, Berry et al. (2004) present a correction for radiative heating errors on the basis of an analytical solution of the heat budget for an idealized ship. They found a RMSE reduction of the air-sea temperature difference of 30% to $\approx$ 0.5°C, eventually reducing the RMSE of q_a.

The uncertainties introduced by different hygrometer types are explored by Kent et al. (1993) in the framework of the VOS Special Observing Project - North Atlantic (VSOP-NA), who suggest to apply an empirical correction to humidity measurements using marine screens. The authors argue that the latter tend to be high-biased in comparison to psychrometers, presumably due to their poor ventilation. Such a correction is presented by Kent and Taylor (1995) for screen-based dew point temperatures. Screen humidity corrections are also applied within Kent et al. (2014) among an intercomparison study of *in-situ* and reanalysis q_a.

Jackson et al. (2009) also focus on hygrometer- and radiation-induced uncertainties, based on ICOADS observations and AMMIc q_a retrievals. However, the authors conclude that both error sources contribute less than 0.05 g kg^{-1} to the overall uncertainty, suggesting their input with respect to the total error budget to be negligible.

c. Applied Methodology

Eq. (3a) - Eq. (4c) incorporate an error contribution associated with the collocation procedure (E_C). As this work's definition of E_C is only related to spatial and temporal mismatches, it is not specifically differentiated between E_C used in Eq. (3a) and Eq. (4c). However, it is likely that an additional random point-to-area uncertainty (error of representativeness, E_R) is inherent to the MTC matchups. This is accounted for, in as much as E_{ins} derived in Eq. (3a) is supplemented by a E_R contribution. However, E_R is not explicitly resolved, as this inhibits a complete error decomposition due to too many unknowns. Instead, the calculated E_{ins} (Eq. (3a) and hence Eq. (3b) - Eq. (4b)) remains slightly larger than in theory, whereas E_M becomes negligibly smaller. Although a quantification of E_R is not possible, the derived decorrelation length scale in Kinzel (2013) considerably exceeds the diameter of a SSM/I footprint, which is the scale of interest regarding the point-to-area issue. It is therefore concluded that E_R lies within the uncertainty of E_{ins} and is therefore negligible in comparison to the overall variances of differences (see note on this in O'Carroll et al. 2008). Equipping *in-situ* data sources with random uncertainty estimates (prior to using them in context of retrieval validation analysis) is strongly recommended, as this would allow to explicitly derive E_R.

One could also argue that the applied MTC method does not yield robust results for the critical q_a regime, which is subject to limited amounts of triplets due to narrow shipping lanes in the subtropical ocean basins. To quantify the robustness of the variances, Scipal et al. (2010) estimated the impact of constraining the TC analysis to small subsets of simulated time series subject to random noise. Results indicate that less than 100 match-ups (=N) lead to *systematic* uncertainties of up to 5%, which does not influence the present analysis. Zwieback et al. (2012), however, argued that the *relative* error, i.e. the standard error relative to the quantity of interest, exceeds 22% for N=100, assuming all error variances to be of similar size and the underlying noise to be normally distributed. If their Eq. (29) holds, at least 2000 match-ups are necessary to restrict the relative error contribution to 5%. For a single year on a seasonal basis, this may imply a reduced reliability of the MTC approach, as the tropical data coverage may temporarily fall below this target.

The chosen collocation criteria are identical to those applied by e.g. Jackson et al. (2006), who also investigated q_a using microwave satellite observations. However, modifications of the collocation criteria underlying this work were also carried out to treat the temporal deviation more strictly, removing collocated pairs where Δt exceeded 60 minutes. Specifically for the critical q_a regime of 12-17 g kg^{-1}, the results do not indicate a reduction of the satellite retrieval error. Instead, the temporal restriction leaves even less match-ups in the already poorly sampled regions, which further increases the random uncertainty of the variance estimates (Scipal et al. 2010). It is therefore concluded that the originally chosen collocation thresholds of 180 minutes and 50 km are adequate. Yet, large humidity gradients may occur along mid-latitudinal shipping routes, associated with frontal systems. However, these do not distort the error decomposition itself, as such outliers have been removed from the analyses (see Section 2c). A comparison of the error bar magnitudes shown in Figure 3 with E_{sum} in Figure 2 yield absolute differences in the order of only 5-10% throughout the whole q_a range. Keeping in mind that the temporal threshold for match-ups shown in Figure 3 is only $\pm$ 1 hour, this further supports the assumption that $\pm$ 3 hours is a reasonable temporal decorrelation scale. In general, the decorrelation time scale cannot be chosen arbitrarily small in preparation for the MTC analysis, because the temporal difference of SSM/I overpasses of two different instruments is in the order of 2-3 hours. This depends on the combination of SSM/I instruments (e.g. Andersson et al. 2010). Consequently, TC *V2* and hence the MTC analysis would often not be realizable if the temporal thresholds were set to e.g. $\pm$ 1 hour.

5. Conclusion and Outlook

Latent heat fluxes (LHF) play a key role in the context of energy exchange between ocean and atmosphere and thus impact the global energy cycle. Due to insufficient spatial sampling of *in-situ* measurements, remote sensing represents an indispensable technique to monitor parameterized LHF in high resolution. However, their uncertainty estimates, which find expression in the satellite's retrieval error E_{tot}, are not sufficiently quantified to date, which complicates their use in context of model validation, trend and variability analyses as well as process studies.

For the near-surface specific humidity q_a, which represents a key geophysical input parameter to parameterized LHF, the aim was to decompose overall satellite-based random q_a uncertainties into individual components to isolate the desired E_{tot}.

In this context, it was shown that the ordinary TC approach can be (and needs to be) extended by means of a novel, multiple TC (MTC) procedure, serving as a powerful tool to distinguish satellite-based random uncertainties associated with the underlying model (E_M) and sensor noise (E_N) from contributions of *in-situ* records (E_{ins}) and collocation (E_C). The MTC analysis was specifically performed for the HOAPS-3.2 q_a on pixel-level basis, based on an extensive match-up database of SWA-ICOADS ship records for the time period of 1995-2008.

The robust results of the MTC analysis indicate that the random retrieval error E_{tot} is on average 1.1 ± 0.3 g kg^{-1}, which is supplemented by averages of E_C (0.5 ± 0.1 g kg^{-1}) and E_{ins} (0.5 ± 0.3 g kg^{-1}). E_N was derived synthetically (0.3 g kg^{-1}). A q_a-dependent analysis shows that the retrieval has largest difficulties in the regime of 12-17 g kg^{-1}, where E_{tot} exceeds 1.5 g kg^{-1}. Largest E_C (0.7 g kg^{-1}) also fall into this range, which is representative for the subtropical domain encompassing the global oceans. On the contrary, E_{ins} increases rather linearly with q_a, taking on values between 0.2 - 1.2 g kg^{-1}. Local analysis on a global scale reveals absolute uncertainty maxima of approximately 1.7 g kg^{-1} off the Arabian Peninsula, where both q_a and wind speed remain in ranges susceptible for large random q_a errors (small wind speeds coupled to rather large, yet not tropical q_a).

Despite random *in-situ* measurement errors and possible deficits underlying the collocation approach, the results suggest that the largest random q_a uncertainties originate from the retrieval itself, which in case of HOAPS-3.2 is based on the linear, single-parameter regression retrieval by Bentamy et al. (2003). The MTC-based findings demonstrate how both regime-dependent retrieval uncertainties and *in-situ* measurement issues can be effectively isolated. This will prove very helpful in further advancing the satellite-based q_a retrieval to meet the desired q_a quality requirements. As discussed in Section 4, HOAPS q_a uncertainties could possibly be reduced by introducing new retrieval algorithms, which could rely on a multiple parameter approach and/or incorporate non-linear regression terms.

Similar to HOAPS-3.2, previous q_a retrievals have mostly been derived from regression analysis using training data sets of T_B's and *in-situ* point measurements. This implies that respective RMSE estimates typically include both E_{ins} and E_C and thus inhibit an explicit determination of the random retrieval uncertainty. This again emphasizes the benefit of the uncertainty decomposition approach. Assigning random uncertainty estimates to all contributing data sources, as done within this work, allows to evaluate the satellite retrieval precision. If only E_{sum} was given, a quantitative comparison between retrieval and *in-situ* random uncertainties to assess retrieval constraints cannot be carried out.

A step towards higher-quality q_a certainly also involves a more comprehensive *in-situ* validation dataset, in which all humidities are equally well represented. This task will be challenging, as the number of VOS is continuously declining (see Kent et al. 2014). Additionally, the ICOADS dataset does not contain call signs after December 2007 (Kent et al. 2013), which further hinders the validation of remotely sensed parameters, as platforms producing systematic measurement errors may no longer be excluded from error analyses.

Future work aims at quantifying E_{tot} of satellite-based wind speed and SST. Respective findings will help to derive E_{tot} of the remaining LHF-related bulk parameters and hence the retrieval uncertainty of HOAPS evaporation.

To better assess the quality of the satellite-based datasets, Prytherch et al. (2014) furthermore argue that grid box based q_a uncertainty estimates would be extremely beneficial, which are not available to date. This approach is currently undertaken at DWD and first results will be published in the near future. As a total error assessment involves the investigation of random error contributions, the presented work can therefore be understood as a first step towards this effort. A full error characterization of all HOAPS freshwater flux related parameters will be implemented in the next official HOAPS climatology, which will be released in late 2016.

Acknowledgments. The first author was funded by the German Science Foundation (DFG). Funding for the second to fourth author was covered by EUMETSAT. The funding for the development and implementation of the collocation softwares was provided by the German Meteorological Service (DWD). The HOAPS-3.2 data was kindly provided by EUMETSAT Satellite Application Facility on Climate Monitoring (CM SAF). SWA-ICOADS data was gratefully obtained from SWA (DWD).

References

Alexander, M. A., I. Bladé, M. Newman, J. R. Lanzante, N.-C. Lau, and J. D. Scott, 2002: The Atmospheric Bridge: The Influence of ENSO Teleconnections on Air-Sea Interaction over the Global Oceans. *J. Climate*, **15**, 2205–2231, doi:10.1175/1520-0442(2002)015⟨2205:TABTIO⟩2.0.CO;2.

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

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

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

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

Berry, D. I., and E. C. Kent, 2011: Air-Sea fluxes from ICOADS: the construction of a new gridded dataset with uncertainty estimates. *Int. J. Climatol.*, **31**, 987–1001, doi:10.1002/joc.2059.

Berry, D. I., E. C. Kent, and P. K. Taylor, 2004: An Analytic Model of Heating Errors in Marine Air Temperatures from Ships. *J. Atmos. Oceanic Technol.*, **21**, 1198–1215, doi: 10.1175/1520-0426(2004)021⟨1198:AAMOHE⟩2.0.CO;2.

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

Bumke, K., M. Schlundt, J. Kalisch, A. Macke, and H. Kleta, 2014: Measured and Parameterized Energy Fluxes Estimated for Atlantic Transects of R/V *Polarstern*. *J. Phys. Oceanogr.*, **44**, 482–491, doi:10.1175/JPO-D-13-0152.1.

Caires, S., and A. Sterl, 2003: Validation of ocean wind and wave data using triple collocation. *J. Geophys. Res.*, **108** (C3), 3098, doi:10.1029/2002JC001491.

Carsey, F. D., 1992: The estimation of geophysical parameters using passive microwave algorithms. *Microwave Remote Sensing of Sea Ice, Geophysical Monograph, Chapter 10.2.5.3*, American Geophysical Union, 214–215.

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

Clayson, C. A., J. B. Roberts, and A. S. Bogdanoff, 2015: The SeaFlux Turbulent Flux Dataset Version 1.0 Documentation (Version 1.2). Tech. rep., Woods Hole Oceanographic Institution, Woods Hole, MA, 5 pp. Available online at http://seaflux.org/seaflux_data/DOCUMENTATION/SeaFluxDocumentationV12.pdf (accessed 02 December 2015).

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, doi:10.1002/qj.828.

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

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

Fennig, K., A. Andersson, S. Bakan, and M. Schröder, 2013: Fundamental climate data record of SSM/I brightness temperatures. Satellite Application Facility on Climate Monitoring (CM SAF), doi:10.5676/EUM_SAF_CM/FCDR_SSMI/V001.

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

GCOS, 2010: Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC. GCOS-138. Tech. rep., WMO, Geneva, CH, 186 pp. Available online at http://www.wmo.int/pages/prog/gcos/Publications/gcos-138.pdf (accessed 02 December 2015).

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

Hollinger, J. P., J. L. Peirce, and G. A. Poe, 1990: SSM/I Instrument Evaluation. *IEEE Transactions on Geoscience and Remote Sensing*, **28**, 781–790, doi:10.1109/36.58964.

IPCC, 2013: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Chapter 3.4.1* [Stocker, T. F., D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P. M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.

Iwasaki, S., and M. K. Kubota, 2012: Algorithms for estimation of air-specific humidity using TMI data. *International Journal of Remote Sensing*, **33**, 7413–7430, doi:10.1080/01431161.2012.685974.

Jackson, D. L., G. A. Wick, and J. J. Bates, 2006: Near-surface retrieval of air temperature and specific humidity using multisensor microwave satellite observations. *J. Geophys. Res.*, **111**, D10 306, doi:10.1029/2005JD006431.

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

Janssen, P. A. E. M., S. Abdalla, H. Hersbach, and J.-R. Bidlot, 2007: Error Estimation of Buoy, Satellite, and Model Wave Height Data. *J. Atmos. Oceanic Technol.*, **24**, 1665–1677, doi:10.1175/JTECH2069.1.

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

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

Kent, E. C., P. G. Challenor, and P. K. Taylor, 1999: A Statistical Determination of the Random Observational Errors Present in Voluntary Observing Ships Meteorological Reports. *J. Atmos. Oceanic Technol.*, **16**, 905–914, doi:10.1175/1520-0426(1999)016⟨0905:ASDOTR⟩2.0.CO;2.

Kent, E. C., N. A. Rayner, D. I. Berry, M. Saunby, B. I. Moat, J. Kennedy, and D. Parker, 2013: Global analysis of night marine air temperature and its uncertainty since 1880: The HadNMAT2 data set. *J. Geophys. Res.*, **118**, 1281–1298, doi:10.1002/jgrd.50152.

Kent, E. C., and P. K. Taylor, 1995: A Comparison of Sensible and Latent Heat Flux Estimates for the North Atlantic Ocean. *J. Phys. Oceanogr.*, **25**, 1530–1549, doi:10.1175/1520-0485(1995)025⟨1530:ACOSAL⟩2.0.CO;2.

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

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

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

Kinzel, J., 2013: Validation of HOAPS latent heat fluxes against parmaeterizations applied to R/V *Polarstern* data for 1995-1997. M.S. thesis, Department of Marine Meteorology, University of Kiel, Germany, 102 pp., [Available online at http://core.kmi.open.ac.uk/display/16271577].

Klepp, C., A. Andersson, and S. Bakan, 2008: The HOAPS climatology: evaluation of latent heat flux. *Flux News: Newsletter of the WCRP Working Group on Surface Fluxes*, **5**, 30–32.

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

Kubota, M. K., and T. Hihara, 2008: Retrieval of Surface Air Specific Humidity Over the Ocean Using AMSR-E Measurements. *Sensors*, **8**, 8016–8026, doi:10.3390/s8128016.

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

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

Liu, W. T., 1986: Statistical Relation Between Monthly Mean Precipitable Water and Surface-Level Humidity over Global Oceans. *Mon. Wea. Rev.*, **14**, 1591–1602, doi:10.1175/1520-0493(1986)114⟨1591:SRBMMP⟩2.0.CO;2.

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

Oreopoulos, L., and W. B. Rossow, 2011: The cloud radiative effects of International Satellite Cloud Climatology Project weather states. *J. Geophys. Res.*, **116**, D12 202, doi:10.1029/2010JD015472.

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

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

Romanou, A., W. B. Rossow, and S.-H. Chou, 2006: Decorrelation Scales of High-Resolution Turbulent Fluxes at the Ocean Surface and a Method to Fill in Gaps in Satellite Data Products. *J. Climate*, **19**, 3378–3393, doi:10.1002/joc.1167.

Rossow, W. B., G. Tselioudis, A. Polak, and C. Jakob, 2005: Tropical climate described as a distribution of weather states indicated by distinct mesoscale cloud property mixtures. *Geophys. Res. Lett.*, **32**, L21 812, doi:10.1029/2005GL024584.

Schlüssel, P., 1996: Satellite remote sensing of evaporation over sea. *Radiation and Water in the Climate System: Remote measurements, Vol. 45, NATO ASI Series*, Springer-Verlag, Berlin, Germany, 431–461.

Schlüssel, P., and A. Albert, 2001: Latent heat flux at the sea surface retrieved from combined TMI and VIRS measurements of TRMM. *International Journal of Remote Sensing*, **22**, 1975–1998, doi:10.1080/01431160118529.

Schlüssel, P., L. Schanz, and G. Englisch, 1995: Retrieval of latent heat fluxes and longwave irradiance at the sea surface from SSM/I and AVHRR measurements. *Advances in Space Research*, **16**, 107–116, doi:10.1016/0273-1177(95)00389-V.

Schulz, J., J. Meywerk, S. Ewald, and P. Schlüssel, 1997: Evaluation of Satellite-Derived Latent Heat Fluxes. *J. Climate*, **10**, 2782–2795, doi:10.1175/1520-0442(1997)010⟨2782:EOSDLH⟩2.0.CO;2.

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

Scipal, K., W. Dorigo, and R. de Jeu, 2010: Triple collocation - a new tool to determine the error structure of global soil moisture products. *IEEE International Geoscience & Remote Sensing Symposium (IGARSS)*, Honolulu, Hawaii, USA, 4426-4429, doi:10.1109/IGARSS.2010.5652128.

Shie, C.-L., K. Hilburn, L. S. Chiu, R. Adler, I.-I. Lin, E. J. Nelkin, J. Ardizzone, and S. Gao, 2012: 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. Available online at doi:10.5067/MEASURES/GSSTF/DATA304 (accessed 02 December 2015).

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

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

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

WMO, 2013: International List of Selected, Supplementary and Auxiliary Ships, WMO Report No. 47. Tech. rep., WMO, Geneva, CH. Available online at https://www.wmo.int/pages/prog/www/ois/pub47/pub47-home.htm (accessed 02 December 2015).

Woodruff, S. D., H. F. Diaz, J. D. Elms, and S. J. Worley, 1998: COADS Release 2 Data and Metadata Enhancements for Improvements of Marine Surface Flux Fields. *Physics and Chemistry of the Earth*, **23**, 517–526, doi:10.1016/S0079-1946(98)00064-0.

Woodruff, S. D., and Coauthors, 2011: ICOADS Release 2.5: extensions an enhancements to the surface marine meteorological archive. *Int. J. Climatol.*, **31**, 951–967, doi:10.1002/joc.2103.

Zwieback, S., K. Scipal, W. Dorigo, and W. Wagner, 2012: Structural and statistical properties of the collocation technique for error characterization. *Nonlinear Processes in Geophysics*, **19**, 69–80, doi:10.5194/npg-19-69-2012.

LIST OF TABLES

TABLE 1. Results of the seasonally-dependent q_a error decomposition (top: extratropics, bottom: tropics). Next to HOAPS averages and their standard deviations (std) of q_a, random errors associated with the retrieval (E_{tot}), collocation (E_C), and *in-situ* source (E_{ins}) [g kg^{-1}] are shown. Relative contributions to the basin-mean q_a are given in brackets [%]. DJF = december-february, MAM = march-may, JJA = june-august, SON = september-november.

decomposed errors / seasons	DJF	MAM	JJA	SON
	Extratropics (30° - 60° N/S)			
HOAPS q_a (average + std)	5.2 ± 0.4	6.1 ± 0.6	10.0 ± 0.7	8.2 ± 0.4
E_{tot} (average + std (rel. contribution))	0.8 ± 0.1 (16%)	0.8 ± 0.1 (14%)	1.0 ± 0.1 (10%)	1.1 ± 0.1 (13%)
E_c (average + std (rel. contribution))	0.6 ± 0.1 (11%)	0.5 ± 0.1 (7%)	0.6 ± 0.1 (5%)	0.7 ± 0.1 (7%)
E_{ins} (average + std (rel. contribution))	0.3 ± 0.1 (5%)	0.4 ± 0.1 (7%)	0.7 ± 0.1 (6%)	0.6 ± 0.1 (6%)
	Tropics (0° - 30° N/S)			
HOAPS q_a (average + std)	13.9 ± 0.8	15.0 ± 1.0	17.4 ± 0.8	16.1 ± 0.7
E_{tot} (average + std (rel. contribution))	1.6 ± 0.2 (11%)	1.4 ± 0.3 (9%)	1.2 ± 0.1 (6%)	1.4 ± 0.3 (8%)
E_c (average + std (rel. contribution))	0.7 ± 0.1 (5%)	0.7 ± 0.1 (4%)	0.8 ± 0.4 (4%)	0.7 ± 0.1 (4%)
E_{ins} (average + std (rel. contribution))	1.1 ± 0.1 (7%)	1.2 ± 0.2 (7%)	1.3 ± 0.1 (7%)	1.2 ± 0.1 (7%)

LIST OF FIGURES

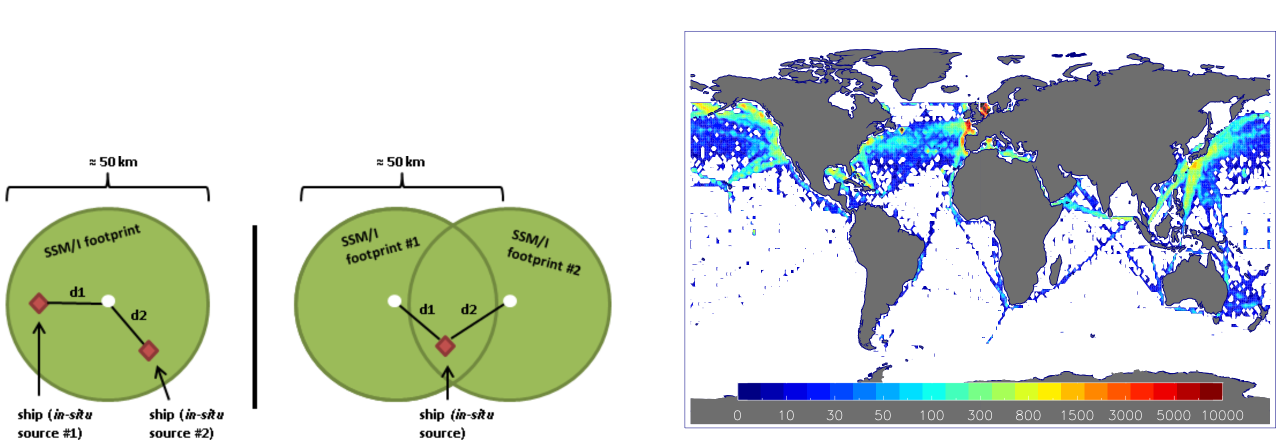

FIG. 1. *Left panel*: Sketch of the applied MTC *V1* and *V2* in preparation for the q_a error decomposition. The red diamonds represent a single ship record. Depending on the MTC version, a ship record is being collocated to a second, independent ship measurement and a HOAPS pixel (*V1*, left) or to pixels of two different satellite instruments (*V2*, right). Temporal and spatial collocation thresholds between the center of a HOAPS pixel and both *in-situ* sources (*V1*) as well as between *in-situ* measurement and both centers of the SSM/I records (*V2*) were set to 180 minutes and 50 km (d1, d2), respectively. *Right panel*: Distribution of TC *V1* triplets (#) between 1995-2008 throughout the global oceans. Note that the colorbar is nonlinear.

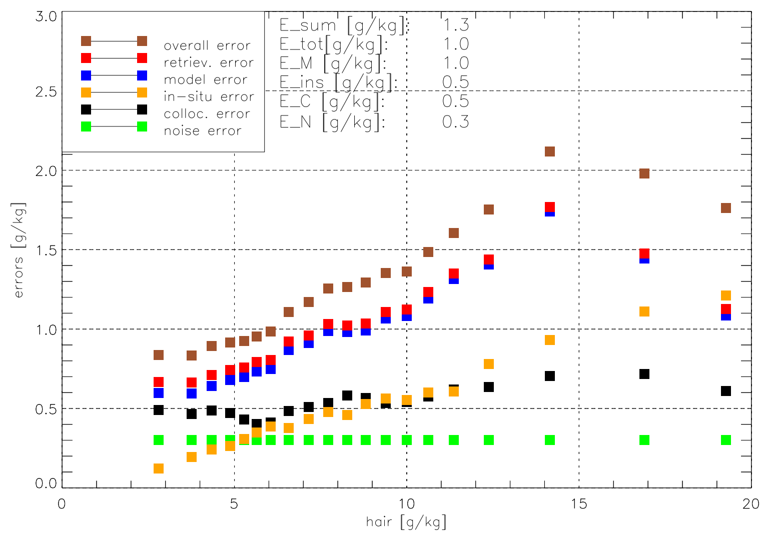

FIG. 2. Decomposition of satellite- and MTC related q_a error terms, based on MTC match-ups between 1995-2008, equatorward of 60°N/S. The decomposition is based on 18005 triplets per *TC* version per bin, which results in a total number of 720200 triplets. The x-axis values of the decomposed random uncertainties are the bin-dependent arithmetic means of the satellite records, which constitute a part of the *TC1* triplets. The strings at the top indicate overall arithmetic means of the individual random error contributions. E_{sum} represents the sum of E_{tot}, E_C, and E_{ins} (legitimate due to the independence of the individual uncertainty components) and allows for a direct comparison to the error bars shown in Figure 3. Recall that E_N was synthetically derived (compare text for further description on this) and thus remains constant throughout the q_a range. The *in-situ* component is based on selected, quality-controlled ship measurements only. Standard deviations (std) of all decomposed random uncertainties are not shown, as the bin-dependent decomposition is very stable and std maxima are in the order of 0.02 g kg^{-1} only.

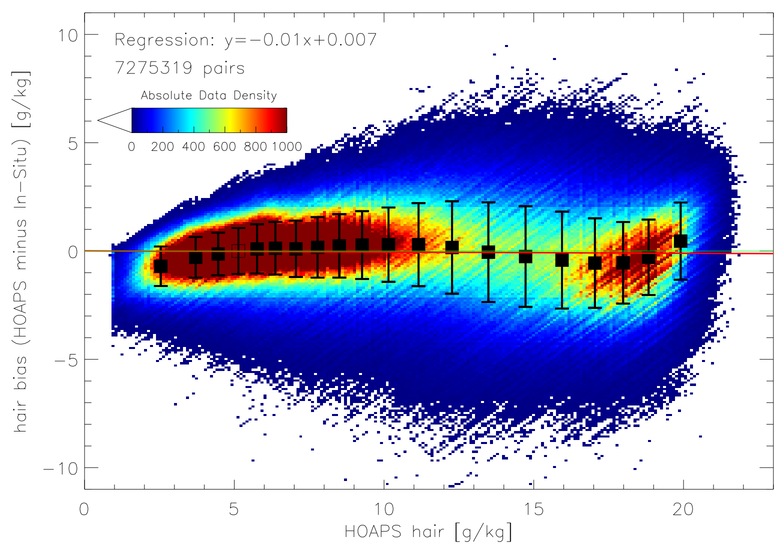

FIG. 3. Non-normalized scatter density plot of q_a bias (HOAPS minus *in-situ* measurements) [g kg^{-1}], based on global double collocations between 1995-2008. Again, the *in-situ* component is composed of selected, quality-controlled ships only. The temporal match-up threshold was set to $\pm$ 1 hour, in contrast to Figure 2. Black (transparent) squares indicate significant (insignificant) bin biases (at the 95% level). Their standard deviations are given by the black bars.

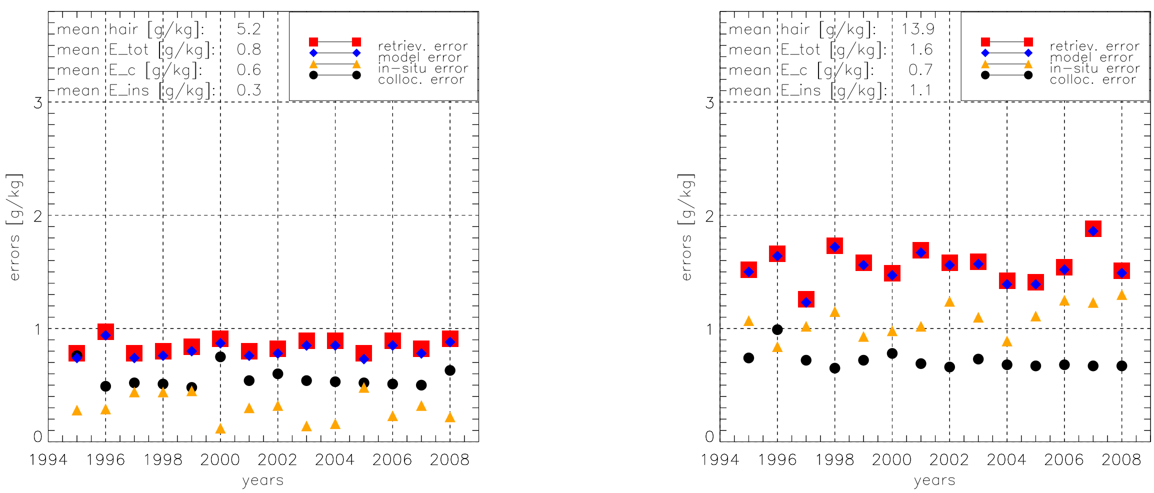

FIG. 4. Time series of decomposed q_a-related errors [g kg^{-1}] for wintertime (DJF) 1995-2008 within the extratropics (30°-60° N/S, *left panel*) and tropics (0°-30° N/S, *right panel*), based on MTC analysis. Statistical values shown on the upper left-hand side are based on the overall time period. Recall that the *in-situ* uncertainty is only based on selected ship measurements. For the sake of simplicity, E_N and E_{sum} are not shown. *Right panel*: As left, but for tropics (0°-30° N/S)

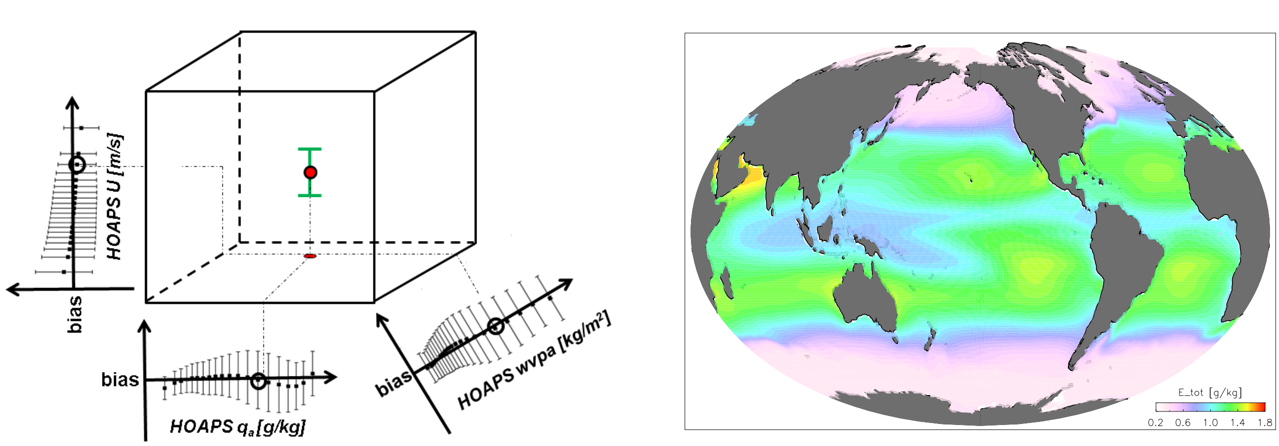

FIG. 5. *Left panel*: simple 3-dimensional sketch illustrating the procedure of assigning multi-dimensional mean biases (red circle) and respective spreads (green error bar) to instantaneous HOAPS pixels of q_a. The black circles along the three axes exemplarily represent the concurrent atmospheric q_a (x-axis), water vapor path (y-axis), and wind speed (z-axis), respectively. *Right panel*: Average instantaneous random retrieval uncertainty of HOAPS q_a [g kg^{-1}] for the time period 1995-2008. The illustrated estimates were derived from a four-dimensional look-up table encorporating the spread of instantaneous q_a biases [HOAPS minus *in-situ*], which was corrected for q_a-dependent contributions of E_C and E_{ins} (according to Figure 2). This table (its simpler, 3-dimensional version is shown on the left-hand side) was created to quantify the random retrieval uncertainty of each HOAPS q_a pixel, based on unique combinations of prevailing q_a, wind speed, SST, and water vapor path values. The averages are presented on a regular 1°x1° grid.