EUMETSAT Satellite Application Facility on Climate Monitoring

Validation Report SSM/I and SSMIS products

HOAPS version 4.0

Precipitation	CM-12611 (PRE_HOAPS)
Vertically Integrated Water Vapour	CM-12701 (HTW_HOAPS)
Evaporation	CM-12801 (EVA_HOAPS)
Latent Heat Flux	CM-12811 (LHF_HOAPS)
Freshwater Flux	CM-12821 (EMP_HOAPS)
Near Surface Specific Humidity	CM-12901 (NSH_HOAPS)
Near Surface Wind Speed	CM-12911 (SWS_HOAPS)

Document Signature Table

	Name	Function	Signature	Date
Author	Kathrin Graw Julian Kinzel	HOAPS Team scientist		31.01.2017
Author	Marc Schröder Karsten Fennig	CM SAF scientist		31.01.2017
Author	Axel Andersson	Former CM SAF scientist		31.01.2017
Editor	Rainer Hollmann	Science Coordinator		31.01.2017
Approval	Rainer Hollmann	Science Coordinator		
Release	Martin Werscheck	Project Manager		

Distribution List

Internal Distribution	
Name	**No. Copies**
DWD Archive	1
CM SAF Team	1

External Distribution		
Company	**Name**	**No. Copies**
PUBLIC		1

Document Change Record

Issue/ Revision	Date	DCN No.	Changed Pages/Paragraphs
1.0	14/11/2016	SAF/CM/DWD/VAL/HOAPS/2	Submitted for review.
1.1	31/01/2017	SAF/CM/DWD/VAL/HOAPS/2	RIDs from DRR2.7 implemented
1.2	15/03/2017	SAF/CM/DWD/VAL/HOAPS/2	Missing action from RR implemented

Applicable documents

Reference	Title	Code
AD 1	Memorandum of Understanding between CM SAF and the Max-Planck Institute for Meteorology and Meteorological Institute, University of Hamburg	
AD 2	Cooperation Agreement	

Reference Documents

Reference	Title	Code
RD 1	CM-SAF Product Requirements Document	SAF/CM/DWD/PRD/2.9
RD 2	Product User Manual	SAF/CM/PUM/HOAPS/2/1.1
RD 3	Algorithm Theoretical Basis Document HOAPS	SAF/CM/ATBD/HOAPS/2/2.3
RD 4	Fundamental Climate Data Record of Microwave Imager Radiances, Edition 3	DOI:10.5676/EUM_SAF_CM/FCDR_MWI/V003

Acronyms and Definitions

Table: List of acronyms and definitions.

Symbol	Comment
1D-Var	1 Dimensional Variational
AD	Applicable Document
AMSR-E	Advanced Microwave Scanning Radiometer-EOS
asst	SST from AVHRR measurements
ATBD	Algorithm Theoretical Baseline Document
AVHRR	Advanced Very-High-Resolution Radiometer
budg	Freshwater Flux
CDR	Climate Data Record
C_E	Transfer Coefficient
CFSR	Climate Forecast System Reanalysis
CM SAF	Satellite Application Facility on Climate Monitoring
COARE	Coupled Ocean Atmosphere Response Experiment
COSMIC	Constellation Observing System for Meteorology, Ionosphere, and Climate
CPC	Climate Prediction Center
CMAP	CPC Merged Analysis of Precipitation
DMSP	Defence Meteorological Satellite Program
DWD	Deutscher Wetterdienst (German MetService)
E	Evaporation
E_C	Random collocation errors
E_{ins}	In-situ measurement noise
E_{tot}	Satellite retrieval error
ECMWF	European Centre for Medium Range Forecast
ECV	Essential Climate Variable
EMP	Freshwater Flux

Symbol	Comment
EOS	Earth Observing System
ERA	ECMWF Re-Analysis
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
EVA	Evaporation
evap	Evaporation
FCDR	Fundamental Climate Data Record
FMI	Finnish Meteorological Institute
GCOS	Global Climate Observing System
GDAP	GEWEX Data and Assessments Panel
GEWEX	Global Energy and Water cycle EXperiment
G-VAP	GEWEX water vapor assessment
GPCP	Global Precipitation Climatology Project
GPROF	Goddard Profiling Algorithm
GPS	Global Positioning System
GPS RO	GPS Radio Occultation technique
GRDC	Global Runoff Data Center
GSMaP	Global Satellite Mapping of Precipitation
GSSTF3	Goddard Satellite-based Surface Turbulent Fluxes (GSSTF) Data Set for Global Water and Energy Cycle Research (GSSTF3)
hair	Near surface specific humidity
HOAPS	Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data
HOAPS-G	HOAPS Gridded product
HTW	Vertically integrated water vpour
ICOADS	International Comprehensive Ocean-Atmosphere Data Set
IFREMER	Institut français de recherche pour l'exploitation de la mer
IPWG	International Precipitation Working Group

Symbol	Comment
ITCZ	Inter-Tropical Convergence Zone
J-OFURO	Japanese Ocean Flux Data Sets with Use of Remote Sensing Observations
JRA-55	Japanese 55-year reanalysis
KNMI	Koninklijk Nederlands Meteorologisch Institut
late	Latent heat flux
lhv	Latent heat of vaporization
LHF	Latent Heat Flux
MERRA	Modern Era Retrospective-Analysis for Research and Applications
MIRS	Microwave Integrated Retrieval System
MTC	Multiple Triple Collocation
N	Number of valid pairs of data to be validated and reference data
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NCL	NCAR Command Language
NMHS	National Meteorological and Hydrological Services
NOAA	National Oceanic and Atmospheric Administration
NOCS	National Oceanography Centre Southampton
NSH	Near surface specific humidity
NWP	Numerical Weather Prediction
OceanRain	Ocean Rainfall And Ice-phase precipitation measurement Network
OAFlux	Objectively Analyzed Air-Sea Fluxes
P	Precipitation
PACRAIN	Pacific Rainfall Database
PERSIANN	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks

Symbol	Comment
PR	Precipitation Radar
PRD	Product Requirements Document
PRE	Precipitation
R	Pearson's correlation coefficient
ρ_a	Density of moist air
rain	Rain rate
RD	Reference Document
RMIB	Royal Meteorological Institute of Belgium
RMSD	Root Mean Square Deviation
RMSE	Root-Mean-Square Error
ROM SAF	Satellite Application Facility on Radio Occultation Meteorology
SMHI	Swedish Meteorological and Hydrological Institute
SPCZ	South Pacific conversion zone
SSM/I	Special Sensor Microwave/Imager
SSMIS	Special Sensor Microwave Imager/Sounder
SST	Sea Surface Temperaure
SWS	Near Surface Wind Speed
TAO	Tropical Atmosphere Ocean
TIROS	Television Infrared Observation Satellites
TOVS	TIROS Operational Vertical Sounder
TCWV	Total Column Water Vapour
TDRs	TMI brightness temperature data files
TMI	TRMM Microwave Imager
TMPA	TRMM Multisatellite Precipitation Analysis
TRMM	Tropical Rainfall Measuring Mission
U	Wind speed

Symbol	Comment
UMORA	Unified Microwave Ocean Retrieval Algorithm
VIM	International Vocabulary of Metrology
VOS	Voluntary Observing Ships
WCRP	World Climate Research Programme
wind	Near surface wind speed
WMO	World Meteorological Organization
wvpa	Vertically Integrated Water Vapour
q_a	Near-surface specific humidity
q_s	Sea surface saturation specific humidity
x_i	Estimation of a reference quantity according to y_i
y_i	Estimation of a quantity retrieved with the HOAPS retrieval

Table of Contents

List of Figures

List of Tables

1 Executive Summary

This report provides information on the validation (to avoid unclear terminology, a passage of the used terminology is included at the end of this section) of the Satellite Application Facility on Climate Monitoring (CM SAF) Hamburg Ocean Atmosphere Parameters and fluxes from Satellite data version 4.0 (HOAPS-4.0) data record. HOAPS-4.0 is derived from microwave satellite observations from the Special Sensor Microwave/Imager (SSM/I) onboard Defence Meteorological Satellite Program (DMSP) platforms F08, F10, F11, F13, F14 and F15 and from Special Sensor Microwave Imager/Sounder (SSMIS) observations onboard DMSP platforms F16, F17 and F18.

These data can be used to determine latent heat flux and precipitation over the global ocean surface. HOAPS-4.0 provides a satellite-based data record with consistently derived global fields of both evaporation and precipitation, and hence of freshwater flux, for the period 1987 to 2014. Fields are provided at monthly and 6-hourly temporal resolution on a 0.5˚ spatial grid, instantaneous/pixel values on the native swath are available on request. Validation results are presented based on comparisons of monthly mean and instantaneous values.

This report presents an evaluation of 7 parameters derived from the latest version of the CM SAF Fundamental Climate Data Record (FCDR) derived from SSM/I and SSMISS observations [RD 4]. An overview is given in Table 1-1.

Table 1-1: Overview of parameters from HOAPS-4.0.

Parameter	CM SAF ID	Abbreviation (as in PRD)	Abbreviation (as in file name and file)
Near surface specific humidity	CM-12901	NSH_HOAPS	hair
Wind speed at 10 m height	CM-12911	SWS_HOAPS	wind
Latent heat flux at sea surface	CM-12811	LHF_HOAPS	late
Precipitation	CM-12611	PRE_HOAPS	rain
Evaporation	CM-12801	EVA_HOAPS	evap
Freshwater flux	CM-12821	EMP_HOAPS	budg
Vertically integrated water vapour	CM-12701	HTW_HOAPS	wvpa

The CM SAF HOAPS-4.0 product is derived from a mixture of newly-developed methods and heritage algorithms from HOAPS-3.2. There are several improvements from HOAPS-3.2. HOAPS-4.0 uses the latest CM SAF FCDR and extends the record to 2014 using data from SSMIS. Uncertainty estimates have been provided for evaporation and the related parameters hair, wvpa, late, and budg. A new physical retrieval scheme, 1D-Var, is used to

estimate the 10 m wind speed and TCWV. All other parameters are calculated using the same method as in HOAPS-3.2 (Andersson et al. 2010; Fennig et al., 2013), but benefit from being calculated from the improved FCDR.

The evaluation of the HOAPS-4.0 data record largely follows Andersson et al. (2010) and Fennig et al. (2013). The gridded monthly means of HOAPS-4.0 products are compared to HOAPS-3.2 and to a range of data records from reanalysis, other satellite observation projects, and in -situ ship measurements as reference. Note that reanalysis data for validation should be used with caution. Additional sections cover the validation of latent heat flux and related parameters using Level 2 data and buoy and ship data as reference (see section 8), the validation of TCWV with Global Positioning System Radio Occultation technique (GPS RO) data, and results from the Global Energy and Water cycle EXperiment (GEWEX) water vapor assessment (G-VAP).

The comparison between HOAPS-4.0 and HOAPS-3.2 shows small changes in near surface specific humidity and precipitation relative to HOAPS-3.2. Wind, evaporation and freshwater flux generally exhibit negative differences (HOAPS-4.0 − HOAPS-3.2), with the main exception being an increase in wind over stratus regions.

In addition monthly mean, gridded HOAPS-4.0 products have been compared to a range of data records:

- Near surface specific humidity: ERA-Interim, IFREMER V2, NOCS V2

- Near surface wind speed: ERA-Interim, IFREMER V2, NOCS V2

- Evaporation: ERA-Interim, IFREMER V2, NOCS V2

- Latent heat flux: ERA-Interim, IFREMER V2, NOCS V2

- Precipitation: ERA-Interim, GPCP V2.2, TRMM 3B43 V7

- Freshwater flux: ERA-Interim, GPCP V2.2+IFREMER V2

- TCWV: COSMIC (beta-version, ROM SAF), ERA-Interim, REMSS(SSM/I+SSMIS) V7, TMI V7

While the general patterns are reproduced by all data records and global mean time series agree often within a range of 10% of the individual products, locally larger deviations occur for all parameters. Specifically regarding near-surface humidity and rainfall estimates, the satellite derived data records agree better with HOAPS than with the reanalyses and/or in-situ data. However, the compared satellite data records are not fully independent, as the satellite input data may be of the same origin and/or similar algorithms or parameterizations are used in retrieval procedures. This also accounts to some extent for ERA-Interim, which assimilates a wide range of satellite data.

The comparison of HOAPS-4.0 (near surface specific humidity, near surface wind speed, latent heat flux and evaporation) to buoy and ship observations, using instantaneous data, exhibits similar or improved quality relative to HOAPS-3.2. In particular, the quality of the wind retrieval has been improved. It is noticeable that this improvement is seen in regions where local maxima in bias relative to other references occur. Due to negligible differences in near surface specific humidity, this results in a reduction in HOAPS-4.0 evaporation. This is can be seen as an overall improvement, specifically regarding the reduction in the freshwater imbalance from 0.77 mm/d (HOAPS-3.2) to 0.50 mm/d (HOAPS-4.0).

The resulting freshwater flux estimates exhibit distinct differences in terms of global averages as well as regional biases. Relative to HOAPS-3.2 the freshwater imbalance has been reduced by more than 0.25 mm/d. Compared to long term mean global river runoff data, the ocean surface freshwater balance is not closed by any of the compared fields. The data records exhibit a positive bias in E-P (evaporation - precipitation) of 0.05 mm/d to 0.5 mm/d. However, more detailed validation efforts are needed to explain and potentially remove the remaining biases between the different data records.

Based on the comparisons presented in this report, we conclude that the HOAPS-4.0 data record is within target accuracies or better, as summarised in Table 9-1 and provides consistent fields of evaporation, precipitation, and the resulting freshwater flux as well as TCWV, near surface specific humidity and near surface (10 m) wind speed that are well suited for further studies on climatological and regional scale.

A description of the parameters, their dependency on additional input data sources, the SSM/I and SSMIS raw data handling and FCDR production, their continuation with future HOAPS versions and the HOAPS versioning approach is given in the product user manual [RD 2]. Basic accuracy requirements are defined in the product requirements document [], and the algorithm theoretical basis document describes the individual parameter algorithms [RD 3].

2   The EUMETSAT SAF on Climate Monitoring

The importance of climate monitoring with satellites was recognized in 2000 by EUMETSAT Member States when they amended the EUMETSAT Convention to affirm that the EUMETSAT mandate is also to "contribute to the operational monitoring of the climate and the detection of global climatic changes". Following this, EUMETSAT established within its Satellite Application Facility (SAF) network a dedicated centre, the SAF on Climate Monitoring (CM SAF, http://www.cmsaf.eu/).

The consortium of CM SAF currently comprises the Deutscher Wetterdienst (DWD) as host institute, and the partners from the Royal Meteorological Institute of Belgium (RMIB), the Finnish Meteorological Institute (FMI), the Royal Meteorological Institute of the Netherlands (KNMI), the Swedish Meteorological and Hydrological Institute (SMHI), the Meteorological Service of Switzerland (MeteoSwiss), and the Meteorological Service of the United Kingdom (UK MetOffice). Since the beginning in 1999, the EUMETSAT Satellite Application Facility on Climate Monitoring (CM SAF) has developed and will continue to develop capabilities for a sustained generation and provision of Climate Data Records (CDR's) derived from operational meteorological satellites.

In particular the generation of long-term data records is pursued. The ultimate aim is to make the resulting data records suitable for the analysis of climate variability and potentially the detection of climate trends. CM SAF works in close collaboration with the EUMETSAT Central Facility and liaises with other satellite operators to advance the availability, quality and usability of Fundamental Climate Data Records (FCDRs) as defined by the Global Climate Observing System (GCOS). As a major task the CM SAF utilizes FCDRs to produce records of Essential Climate Variables (ECVs) as defined by GCOS. Thematically, the focus of CM SAF is on ECVs associated with the global energy and water cycle.

The CM SAF data records can serve applications related to the new Global Framework of Climate Services initiated by the WMO World Climate Conference-3 in 2009. CM SAF is supporting climate services at national meteorological and hydrological services (NMHSs) with long-term data records but also with data records produced close to real time that can be used to prepare monthly/annual updates of the state of the climate. Both types of products together allow for a consistent description of mean values, anomalies, variability, and potential trends for the chosen ECVs. CM SAF ECV data records also serve the improvement of climate models both at global and regional scale.

A catalogue of all available CM SAF products is accessible via the CM SAF webpage, http://www.cmsaf.eu/. Here, detailed information about product ordering, add-on tools, sample programs and documentation is provided.

3   Introduction

The SSM/I and SSMIS radiometers aboard the DMSP satellites, available since 1987, became a foundation for the derivation of surface flux and precipitation time series by various international research groups. Depending on the data record application purpose, blending and pattern morphing techniques have been developed to combine different satellite and model data with the SSM/I and SSMIS time series.

Generally these data records fall into two categories providing either surface fluxes or precipitation estimates. Prominent surface flux products are the Goddard Satellite-Based Surface Turbulent Fluxes version 2 (GSSTF3; Shie et al., 2012), the Japanese Ocean Flux Data records with the Use of Remote Sensing Observations (J-OFURO 2; Kubota and Tomita 2007), the objectively analyzed air–sea fluxes (OAFlux; Yu and Weller 2007; Yu et al. 2008), the merged flux data record of the Institut Français de Recherche pour l'Exploration de la Mer (IFREMER; Bentamy et al. 2003, 2013), evaporation from Remote Sensing Systems (Hilburn, 2009), and SeaFlux (Clayson et al., 2012). Well-known and widely used precipitation products for a variety of applications are the Global Precipitation Climatology Project (GPCP; Huffman et al. 1997; Adler et al. 2003), the Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA; Huffman et al. 2007), the Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP; Xie and Arkin 1997), the Unified Microwave Ocean Retrieval Algorithm (UMORA; Hilburn and Wentz 2008), the Global Satellite Mapping of Precipitation (GSMaP; Kubota et al. 2007), and Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN; Hsu et al. 1997; PERSIANN-CDR, Ashouri et al., 2015). The NOAA Microwave Integrated Retrieval System (MIRS, Boukabara et al., 2011) also provides a 1D-Var based retrieval from passive microwave sensors, but currently starts in 2008 and does not include data from SSM/I.

The combination of such satellite-retrieved data records results in estimates of the global ocean freshwater flux. Schlosser and Houser (2007) state that this is a highly required but difficult task, as differently calibrated time series and inhomogeneous data sources have to be combined while there is no comprehensive in situ validation data available.

Alternatively, reanalysis data records, such as the European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis ERA-Interim (ERA-Interim; Dee et al., 2011), National Aeronautics and Space Administration Modern-Era Retrospective Analysis for Research and Applications (MERRA, Rienecker et al., 2011, MERRA2), National Centers for Environmental Prediction (NCEP) Climate Forecast System Reanalysis (CFSR, Saha et al., 2010), or the Japan Meteorological Agency Japanese 55-year Reanalysis (JRA-55) (Kobayashi et al. 2015), provide the relevant water cycle parameters.

For the ocean surface fluxes, ship observations give the opportunity to derive global ocean data records, such as the National Oceanography Centre Southampton (NOCS) surface flux data record (Berry and Kent, 2011).

In contrast, the Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data (HOAPS) has been developed with the goal to derive the parameters required to retrieve the global ocean surface freshwater flux components consistently within one entirely satellite based data record (Andersson et al., 2010b). For the sake of long-term homogeneity the approach for HOAPS is to use the SSM/I and SSMIS as the common data source for all retrievals instead of combining different data sources. This ensures a uniform sampling for all parameters and avoids complications with the cross calibration and the implementation of

retrieval procedures for different types of sensors. Another criterion for the design of the data record is to use stand-alone retrieval procedures that only rely on SSM/I and SSMIS brightness temperatures and the Advanced Very High Resolution Radiometer (AVHRR)-based SST as input and are independent of ancillary input data, such as additional first guess fields from model output. With HOAPS-4.0 a priori selection from a background profile data base is needed as input to the 1D-Var retrieval of TCWV and near surface wind speed.

Several recent readily available products are selected for comparison. These products are derived from different data sources and represent model-based estimates from a reanalysis data record, in situ measurements from ships, and different satellite-based data records that include sensors not utilized in HOAPS.

For the evaporation and related parameters, these are the ERA-Interim reanalysis, the ship-measurement based NOCS V2 data record, and the satellite-data based IFREMER V2 flux data record.

The HOAPS precipitation product is compared to ERA-Interim and the two satellite-based products, GPCP V2.2 and TRMM 3B43, version 7.

HOAPS TCWV products are compared to the TMI-based product from REMSS and to the combined microwave imager product from REMSS (both in version 7), ERA-Interim and COSMIC and Metop GPS RO data (beta-version from ROM SAF).

A recently combined buoy and ship data record from DWD and ICOADS is used as reference for comparison of meteorological and flux parameters using instantaneous data.

All used time series have a significant temporal overlap with the full HOAPS time series of 27 full years. However, the overlap is a function of reference data record and can be smaller. In general it may be concluded that the resulting comparisons are an advance over previous studies in terms of temporal coverage, in terms of number of reference data records and in terms of approaches.

Note, that some of the reference data is not fully independent as input data from the same observations were used. Scatterometer wind speeds or ship observations are, for example, assimilated into the ERA-Interim reanalysis and SSM/I brightness temperatures are used in ERA-Interim and also in the satellite-based products. These examples illustrate that there is a need for independent, high quality reference data enabling in the first place a profound validation of new climate data records (CDRs), e.g. the HOAPS products. To validate the HOAPS-4.0 products a variety of mature and widely used reference data have been chosen. The chosen references have been selected advisedly in a former requirements review, neither admitting that the references might not be globally bias free nor implying superior quality over HOAPS-4.0.

The validation performed in this validation report perceives as a comparison between different data sets to ideally get objective evidence that specified requirements are fulfilled. In terms of the International Vocabulary of Metrology (VIM) this is conform with a "verification". Furthermore a "validation" is defined in the VIM as a verification where the requirements are adequate for an intended use. Formally speaking this is the purpose of the review process associated with the release of CM SAF climate data records: verify that requirements in the field of climate applications are fulfilled. Therefore the term "validation" is used throughout this report.

An overview of the data records used as reference in this study is given in section 4. Comparisons to HOAPS-3.2 and climatological comparisons between HOAPS and the other

products are shown in section 5 and section 6 together with a discussion of the results and implications for the individual parameters. Section 7 presents the results from the decadal stability analysis. Results from a comparison to instantaneous data from buoys and ships are shown in section 8. Finally, conclusions and outlook are given in section 8.

4 Data records for Comparison with HOAPS-4.0

True values of the individual parameters are not known and a true reference for all HOAPS-4.0 products is not available. There is still a need for improved high quality reference data on which a profound validation can be based on. For this validation it was proposed to compare the HOAPS-4.0 data records with a variety of data records which have been advisedly chosen based on experience, literature and a review meeting. In this way a set of reference data records were identified and explicitly defined to allow a unique conclusion regarding the requirements.

4.1 IFREMER Satellite Derived Turbulent Fluxes V2

Bentamy (2003) developed a remotely sensed data record of wind stress and surface turbulent latent and sensible heat fluxes. It utilizes different input sources to derive the flux parameters using the COARE 3.0 algorithm (Fairall et al., 2003). As in HOAPS-4.0, q_a is derived from SSM/I data with the algorithm described in Bentamy (2003). In contrast to HOAPS-4.0, the IFREMER wind speed is derived from a combination of retrievals based on scatterometers and SSM/I data. The flux fields are retrieved using a kriging method to merge the various satellite estimates. The SST is taken from the NOAA - Optimum Interpolation (OI) weekly product. Here the version 2 is used, which is currently available on a regular latitude-longitude grid with 1° spatial resolution for the time period from March 1992 to December 2007 (Bentamy et al., 2008).

Note that a version 3 is available as well (Bentamy et al., 2013) which covers the QuickSat period from November 1999-November 2009 and has a spatial resolution of 0.25°.

It was chosen to rely on V2 as it covers 15 years instead of only 9 years. Thus, it provides a longer temporal overlap.

4.2 NOCS v2.0

The National Oceanography Centre Southampton (NOCS) v2.0 surface flux data record (Berry and Kent, 2009, 2011) is exclusively based on Voluntary Observing Ship (VOS) data from the International Comprehensive Ocean–Atmosphere Data Set (ICOADS; Woodruff et al., 2011). The NOCS data record provides fields of marine surface meteorology and fluxes over the global ice-free ocean that is constructed using a bias adjustment procedure and an optimum interpolation method. The turbulent fluxes are derived with the bulk parameterization of Smith (1980, 1988). The data is provided on a regular longitude-latitude grid with 1° spatial resolution and covers the period 1973 to 2014. For comparison HOAPS was regridded to meet the NOCS spatial resolution.

Note the sampling characteristics of NOCS due to the data sparse regions in the southern ocean. For example NOCS winds are underestimated because there are almost no winter measurements, rather than an avoidance of storms. Thus the variability can not be estimated in such cases.

4.3 GPCP V2.2

The GPCP version 2.2 combined product (Adler et al. 2003, Huffman et al., 2009) provides fields of satellite-derived precipitation retrieved from passive microwave including SSM/I and SSMIS and infrared data. For the merging procedure the infrared precipitation estimates from geosynchronous satellites are continuously calibrated with the passive microwave precipitation retrievals from polar orbiting satellites, which are considered to be of higher

accuracy. Different from HOAPS, GPCP provides also precipitation observations over land, where the analysis also makes use of surface data from rain gauges. Monthly means are available from 01 January 1979 to present on a cylindrical equal area grid of 2.5°x2.5°. For comparison HOAPS was regridded to meet the GPCP spatial resolution.

4.4 TRMM 3B43

In the 3B43 product, data from the TRMM Microwave Imager (TMI) and the Precipitation Radar (PR) are blended with SSM/I, geosynchronous precipitation and rain gauge data. In analogy to the GPCP product, microwave and infrared retrievals are used to complement the TRMM precipitation retrievals. The 3B43 product is available for the period 1 January 1998 to 31 March 2015 and is limited to the region between 50˚S and 50˚N. The monthly means are available in version 7 (V7) on a regular longitude-latitude grid with 0.25° resolution. More details can be found in Huffman and Bolvin (2014). The grid resolution has been adapted to HOAPS resolution by applying arithmetic averaging.

4.5 ERA-Interim

ERA-Interim (Dee et al. 2011) is a third-generation reanalysis and improves on previous versions, for example, by using a four-dimensional variational data assimilation (4D-Var) scheme for atmospheric analysis and variational bias correction. ERA-Interim output has a native horizontal resolution of about 0.758°, a temporal resolution of 6 h (3D fields) and 3 h (2D fields) and covers the period 01 January 1979 to present. It assimilates observations from a large variety of instruments, among them radiances from several satellites (see Dee et al., 2011).

Monthly means (of daily means) of TCWV, wind speed at 10 meters, dew point temperature and pressure (both at 2 m) were obtained online (http://apps.ecmwf.int/data-records/) with a spatial resolution of 0.5° longitude–latitude. Precipitation, evaporation and latent heat flux were obtained from two 12 hour forecasts and then added to get a monthly average of daily sums. Near surface specific humidity was computed from dew point temperature and surface pressure: first vapour pressure is calculated from dew point temperature following WMO standards (Hyland and Wexler, 1983) and then vapour pressure and pressure are used to estimate the near surface specific humidity. For all parameters the grid centre was shifted by half a degree, e.g., from (180°, 0°) to (179.75°, 0.25°).

The ocean surface freshwater flux fields were calculated by subtracting the respective evaporation and precipitation monthly mean grid values.

4.6 SSM/I and TMI products from REMSS

The TCWV values are retrieved from SSM/I (F-08, F-10, F-11, F-13, F-14, and F-15), the F-16 and F-17 SSMIS, AMSR-E (Aqua), and WindSat (Coriolis) instruments. These microwave radiometers have been carefully intercalibrated at the brightness temperature level and the version 7 (V7) ocean products have been produced using a consistent processing methodology for all sensors. The TCWV retrieval is described in Wentz (1997) and Wentz and Spencer (1998). The TCWV data record from REMSS is available over the global ice-free ocean on a regular longitude–latitude grid with 1° resolution and contains monthly means from January 1988 to the month of most recent processing. TMI data products are available for the time period from 7 December 1997 to 31 December 2014. TMI brightness temperature data files (TDRs) were obtained from NASA Goddard and were reverse engineered back to raw radiometer counts. Using a consistent processing scheme and a robust radiative transfer

model, the TMI data were intercalibrated with the other microwave radiometers. Then the ocean products were generated following Wentz (1997) and Wentz and Spencer (1998). The V7 products are available over the global ice-free ocean on a regular longitude-latitude grid with 0.25° resolution. The grid resolution has been adapted to HOAPS resolution by applying arithmetic averaging.

4.7 GPS RO

Within a Federated Activity CM SAF cooperates with the Satellite Application on Radio Occultation Meteorology (ROM SAF) on the evaluation of GPS RO and HOAPS data. A beta version of Level 2 specific humidity profiles from COSMIC and Metop have been provided by ROM SAF (Metop data provided on 04 October 2016 and COSMIC data provided 09 August 2016). The GPS RO data covers the periods September 2006 – December 2014 (COSMIC) and December 2008 - September 2015 (Metop).

The GPS RO instrument looks tangentially through the Earth's atmosphere and measures the time delay of the refracted GPS radio signals as the ray passes through the Earth's atmosphere on its way from the transmitting satellite to COSMIC or Metop. Thus, the data is not available on swath or grid basis but are quasi-randomly distributed over the whole globe, and the horizontal resolution for each individual atmospheric profile is between 100 km and 300 km. Within the troposphere the GPS RO data suffers from ambiguity because both temperature and humidity impact the refractive index and thus the time delay. This ambiguity is resolved by using a 1D-Var analysis with ERA-Interim temperature and humidity data as *a priori*. More information is available at http://www.romsaf.org.

4.8 Buoy and ship observations

To quantify both systematic and random uncertainties in HOAPS-4.0 latent heat flux (LHF) related parameters, i.e., near-surface specific humidity (q_a), wind speed (U), and sea surface saturation specific humidity (q_s), in-situ data from global high-quality shipborne measurements as well as data provided by drifting and moored buoys are used. The in-situ point measurements represent the ground reference and are available in up to hourly temporal resolution from the marine meteorological data archive of Deutscher Wetterdienst (DWD), supervised by the Seewetteramt Hamburg (part of DWD). In case of data gaps, records are complemented by data from the International Comprehensive Ocean-Atmosphere Data Set (ICOADS) (version 2.5; Woodruff et al. 2011). More details can be found in Kinzel et al. (2016).

5   Comparison of HOAPS-4.0 and HOAPS-3.2 products

Figure 1 shows climatological differences of the HOAPS-4.0 and HOAPS-3.2 parameters wind, evaporation, precipitation and freshwater budget for the time period 1988-2008. Regarding wind speed (top left), the output of the 1D-Var retrieval resolves larger wind values in comparison to the neural network based HOAPS-3.2 values along 20°N/S of the Eastern Atlantic and Pacific Ocean. Similar bias magnitudes are located off the Arabian Peninsula. Smaller positive differences of HOAPS-4.0 wind speeds are found along 30°S over the Indian Ocean. Maxima of up to 1 m/s are limited to coastal areas of Peru and Chile. By contrast, the remaining regions experience a wind speed reduction in HOAPS-4.0 on a climatological scale. This reduction becomes relatively large (down to -1.2 m/s) over the Central tropical Pacific and off eastern coasts of South America. A similar decrease is located over the tropical Indian Ocean. On average, HOAPS-4.0 wind speeds have decreased by 0.39 m/s in comparison to HOAPS-3.2.
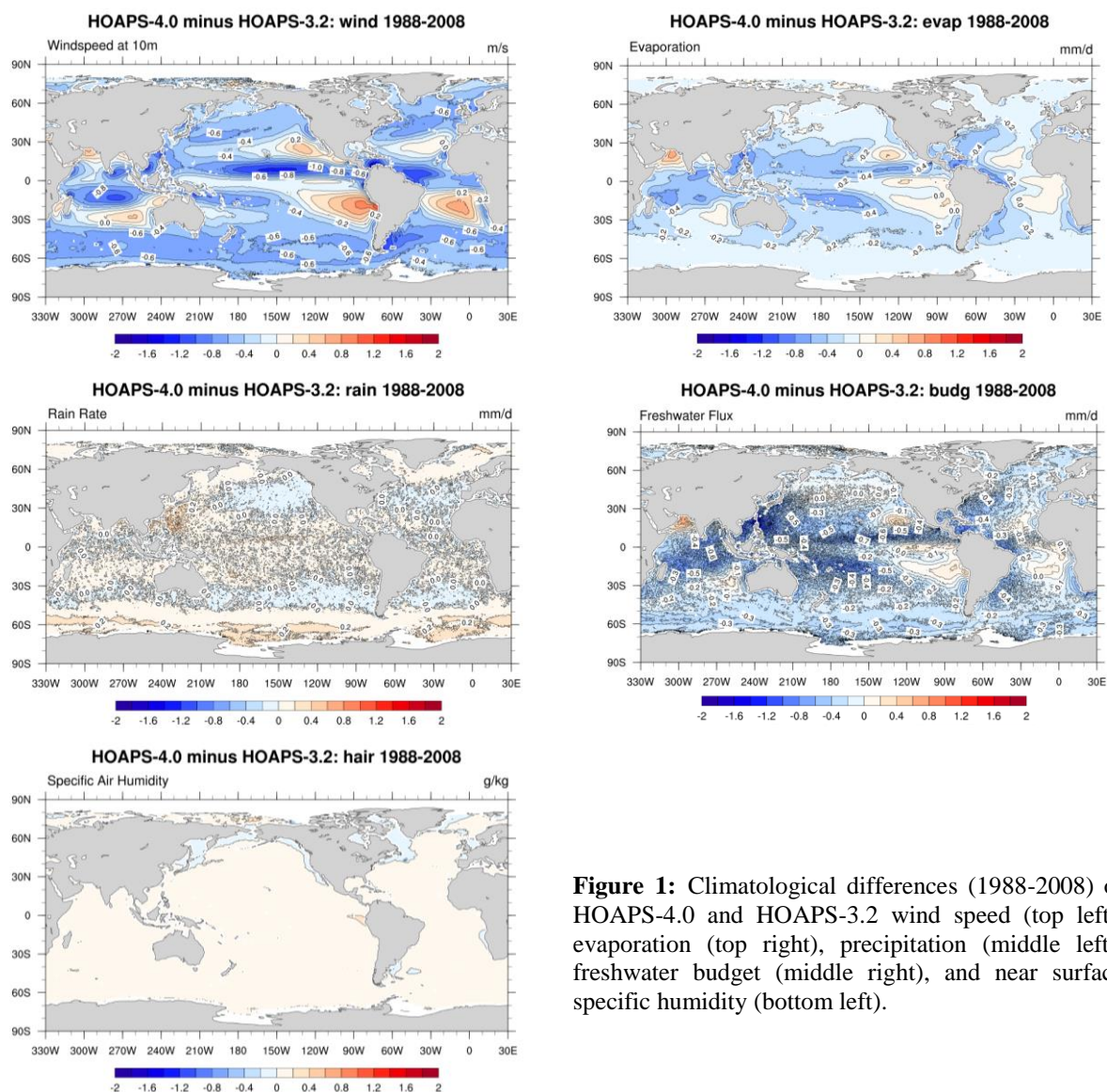
Figure 1: Climatological differences (1988-2008) of HOAPS-4.0 and HOAPS-3.2 wind speed (top left), evaporation (top right), precipitation (middle left), freshwater budget (middle right), and near surface specific humidity (bottom left).

A comparison to global average wind speeds U from HOAPS-4.0 (Figure 4, Section 6.3) indicates that the positive biases exclusively go along with climatologically smallest U. Comparing DWD-ICOADS data to instantaneous HOAPS-3.2 and HOAPS-4.0 records

indicates that the rather strong underestimation of HOAPS-3.2 U over these regions (not shown) has been considerably reduced compared to HOAPS-4.0 (compare Figure 21b, Section 8.3). Likewise, the overestimation of strong HOAPS-3.2 U (not shown) has also become smaller (compare Figure 21b), e.g. over the extratropical ocean basins of both hemispheres and over the Central Indian and Pacific Ocean. A note on the improvement of the HOAPS-4.0 U retrieval is also given in Section 8.5 later on. Regarding the local bias maximum off the Arabian Peninsula, Andersson et al. (2011) hypothesize that U differences among data records over this regime are likely to originate from incorrect representation of the atmosphere and sea surface properties. The increase of HOAPS-4.0 U in comparison to HOAPS-3.2 reduces the bias to in-situ measurements, which may imply that the 1D-Var retrieval is able to better capture the state of the atmosphere and the sea surface in comparison to the neural network implemented for HOAPS-3.2.

The HOAPS-4.0 parameters which have been retrieved with the 1D-Var can be considered as independent of the source of the background profiles used. The background profile database was constructed to get a wide spread of atmospheric states. It is important that no atmospheric states are systematically excluded in the database and that the profiles are physically consistent. It is a fixed database that is not linked to geographical coordinates, not to time and which is not updated according to space and time of the measurements during processing. The retrieved PDF of precipitation have been compared to the corresponding PDF from ERA-Interim (the origin of the profile database). The PDFs are clearly different (not shown here).

Deviations in near-surface specific humidities (q_a) among both data records are small and more or less uniform, showing a general increase of approximately 0.1 g/kg (not shown). Exceptions are confined to coastal regions of the northern hemispheric extratropical continents, where biases are of similar magnitude, yet negative (not shown). This was to be expected, as both HOAPS-3.2 and HOAPS-4.0 make use of the same humidity retrieval. The observed q_a deviations are therefore associated with the difference in the SST (compare ATBD [RD 3]).

Comparing climatological differences of evaporation (Figure 1, top right), the pattern closely resembles that of the wind speed biases discussed above, owing to the uniform, minor increase in HOAPS-4.0 q_a (global mean difference: 0.07 g/kg). Analogously, most oceanic regions show a reduction in HOAPS-4.0 evaporation compared to HOAPS-3.2, which locally exceeds -0.8 mm/d. By contrast, the overestimation of HOAPS-4.0 off the Arabian Peninsula (exceeding 0.6 mm/d) is striking. It remains unclear why the remaining positive biases in wind speeds (off the eastern margins of the subtropical continents, top left) do not produce equally large positive evaporation biases as over the Arabian Sea. This is presumed to be linked to differences among sea surface temperatures (SST) and thus sea surface saturation specific humidities (q_s), which essentially contributes to the evaporation estimates. Note that the cold bias in the SST data record, which according to Andersson et al. (2010b) caused the minimum in HOAPS-3.2 after the eruption of Mount Pinatubo, has considerably been reduced in HOAPS-4.0. The global mean bias in comparison to HOAPS-3.2 equals to -0.24 mm/d.

Differences in climatologically mean precipitation estimates (Figure 1, middle left) are minor and range between -0.2 mm/d to 0.4 mm/d (global mean difference: 0.03 mm/d). Extratropical storm track regions (and poleward) experience a slight increase in rain intensities in HOAPS-4.0, whereas small decreases are mainly concentrated around 40°N/S of all ocean basins. (Sub-) tropical rain rates do not show a clear sign in bias and the differences remain negligible in terms of magnitude.

These rather insignificant bias magnitudes were to be expected, as both HOAPS-4.0 and HOAPS-3.2 are based on the same neural network precipitation algorithm presented in Andersson et al. (2010b). It is therefore concluded that the observed minor differences are associated with the implementation of an updated FCDR, also utilising data from SSMIS in HOAPS-4.0 from 2005 onwards.

The climatological differences of q_a between HOAPS-4.0 and HOAPS-3.2 are negligible because the same statistical retrieval was applied in HOAPS-4.0 and in HOAPS-3.2. Nevertheless small differences exist which can be explained with the utilisation of the updated FCDR.

Figure 1 (middle right) shows the bias in freshwater budget (E-P) between HOAPS-4.0 and HOAPS-3.2, which is clearly driven by differences in the evaporation product. With the exception of locally isolated maxima off the Arabian Peninsula and west of Mexico (up to 1 mm/d), the HOAPS-4.0 freshwater budget experiences an overall reduction, owing to smaller evaporation estimates (Figure 1, top right). The reduction takes on values of up to 0.7 mm/d along the Tropical Pacific ITCZ (global man difference: -0.27 mm/d). The observed reduction in HOAPS-4.0 E-P implies that the global fresh water imbalance of 0.77 mm/d observed for HOAPS-3.2 (1992-2005: 0.73 mm/day, Andersson et al., 2011) has considerably been reduced and is now 0.50 mm/d.

We conclude that the 1D-Var retrieval underlying HOAPS-4.0 has led to an improvement of the wind speed estimates in comparison to HOAPS-3.2. Due to negligible differences in q_a, this results in a reduction in HOAPS-4.0 evaporation. This can be seen as an overall progress, specifically regarding the reduction in the freshwater imbalance from 0.77 mm/d (HOAPS-3.2) to 0.50 mm/d (HOAPS-4.0).

6 Evaluation of HOAPS parameters using gridded data

6.1 Methodology

For the following comparisons a common time period from 1988–2014 is chosen, which is covered by all data records except for the TRMM 3B43 and TMI products, which start only in 1998 and the IFREMER data, which covers the period 1993-2007.

The HOAPS land–sea and ice masks were applied to all data records to achieve a common spatial coverage of the global ice-free oceans. Apart from the differences of the mean fields for the period 1988–2014, the respective zonal means have been calculated as well as the time series of the monthly global mean values.

The IFREMER evaporation and GPCP precipitation were subtracted from each other to provide a freshwater flux product for comparison, in addition to the ERA-Interim freshwater flux product.

Here bias, RMSD and temporal stability are assessed to characterise the quality of the HOAPS-4.0 products.

The bias (or mean difference) between two estimations x_i (reference data) and y_i (estimated with the HOAPS retrieval) of the same variable is computed as:

$$\text{bias} = \frac{1}{N} \sum_{i=1}^{N} (y_i - x_i) \tag{1}$$

27

The RMSD is defined as follows:

$$\mathrm{RMSD} = \sqrt{\frac{1}{(N-1)} \sum_{i=1}^{N} (y_i - x_i)^2} \qquad (2)$$

In Eqs. 1 and 2 the sum is computed for all valid pairs, noted N. In sections 5 and 6 bias and RMSD are estimated on basis of global mean values.

The term bias describes the mean difference between HOAPS and the reference data records and also the RMSD is based on such differences. The reference data records are mature and widely used. However, it is not known whether or not the reference data records are bias free on global and regional scales. Also, it does not imply that the reference data records have superior quality relative to HOAPS-4.0.

Relative differences for every parameter of HOAPS-4.0 have been calculated as follows:

$$\mathrm{rel\_diff} = \frac{HOAPS-4.0-reference}{reference} \cdot 100$$

Only one reference has been taken for every parameter. Which reference this is, is given in Figure 26 in the Annex (section 11) together with the individual plots of the relative differences for every parameter.

Decadal stability is computed by applying linear regression analysis to the results from Eq. 1 using the function "regline_stats" from NCL (NCAR Command Language). All samples are weighted equally. In the future it is intended to weight the samples according to their individual retrieval uncertainties. The slope of the regression is the temporal change of the bias, either per month (section 7) or per collocated data pair (section 8). The probability, that the stability is smaller than a requirement is computed by integrating the Gaussian noise distribution using the 1-sigma noise level from the linear regression analysis within limits defined by the requirement. If normalised and multiplied with 100 it gives the coverage probability of the stability being within the requirement. Based on this the p-value can be computed (100-coverage probability)/100. The null hypothesis is that the stability is outside the requirement and the alternative hypothesis is that the stability is within the requirement. The null hypothesis needs to be rejected if the coverage probability >95% (or p<0.05).

The Pearson's correlation coefficient R between the variables x and y, each having N elements, is defined as follows:

$$R = \frac{N \sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{\sqrt{N \sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2} \times \sqrt{N \sum_{i=1}^{N} y_i^2 - \left(\sum_{i=1}^{N} y_i\right)^2}} \qquad (A3)$$

6.2 Near surface specific humidity

6.2.1 Results

HOAPS-4.0 climatological mean near-surface atmospheric specific humidity (q_a) and its annual cycle are shown in Figure 2. The highest values of up to 20 g/kg of the climate mean
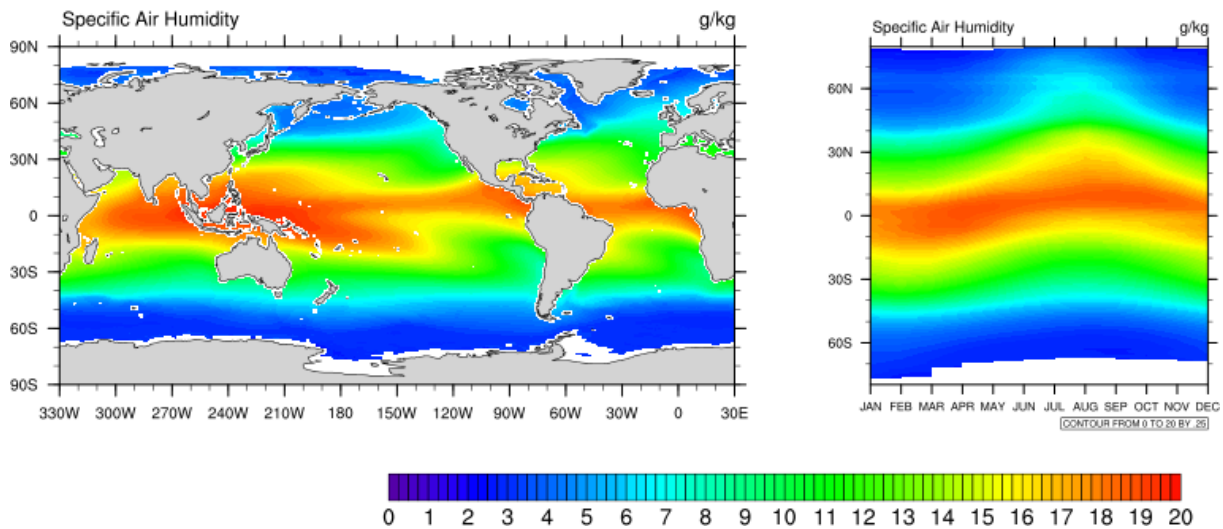
Figure 2: Climatological mean field (left) and zonal mean annual cycle (right) of HOAPS-4.0 near surface specific humidity q_a for the years 1988–2014.

q_a are found in the tropical warm pool region and the ITCZ (left panel). Towards the subpolar regions the values decrease below 2 g/kg. The annual cycle (Figure 2, right panel) clearly shows the movement of the tropical q_a maximum with the position of the sun.

The comparison of q_a between HOAPS and ERA-Interim (Figure 3, top left) exhibits the most distinct differences in the tropical and subtropical regions. Positive biases exceed 1 g/kg over the subtropical Atlantic and Pacific along 30°N/S as well as in parts of the warm pool area (<10%). Biases of up to 2.3 g/kg (20%) are located over the eastern tropical Atlantic and Pacific. By contrast, HOAPS-4.0 q_a are systematically lower than ERA-Interim along 10˚S and 10-20˚N over the Indian Ocean, the central Pacific, and eastern Atlantic. These regions comprise the negative biases up to -2.2 g/kg, which are specifically found off the Arabian Peninsula and in the Caribbean. Likewise, HOAPS-4.0 tends to underestimate q_a along the Namibian coast and subtropical coastal areas of Chile.

The comparison with NOCS (Figure 3, middle) reveals a similar pattern but is shifted toward a more negative bias. In particular over the subtropical ocean of all basins, HOAPS-4.0 q_a is partly underestimated by more than 2 g/kg (20%) in comparison to NOCS. Locally, biases of -3 g/kg are observed. Positive biases up to 2 g/kg are confined to the eastern equatorial Pacific and the tropical eastern coasts of Africa.

The deviations between HOAPS-4.0 and IFREMER (Bentamy et al, 2003) are overall small with ±0.5 g/kg in parts of the tropical regions of all ocean basins and over the western boundary currents (Figure 3, upper right). This is within 5% for the most regions, locally within 10%. In the tropical belt, the general tendency of the regional differences is similar to the previous comparisons, with IFREMER being much closer to HOAPS-4.0 than to the other products. This is also mirrored in the global mean bias, which is given by 0.03 g/kg.
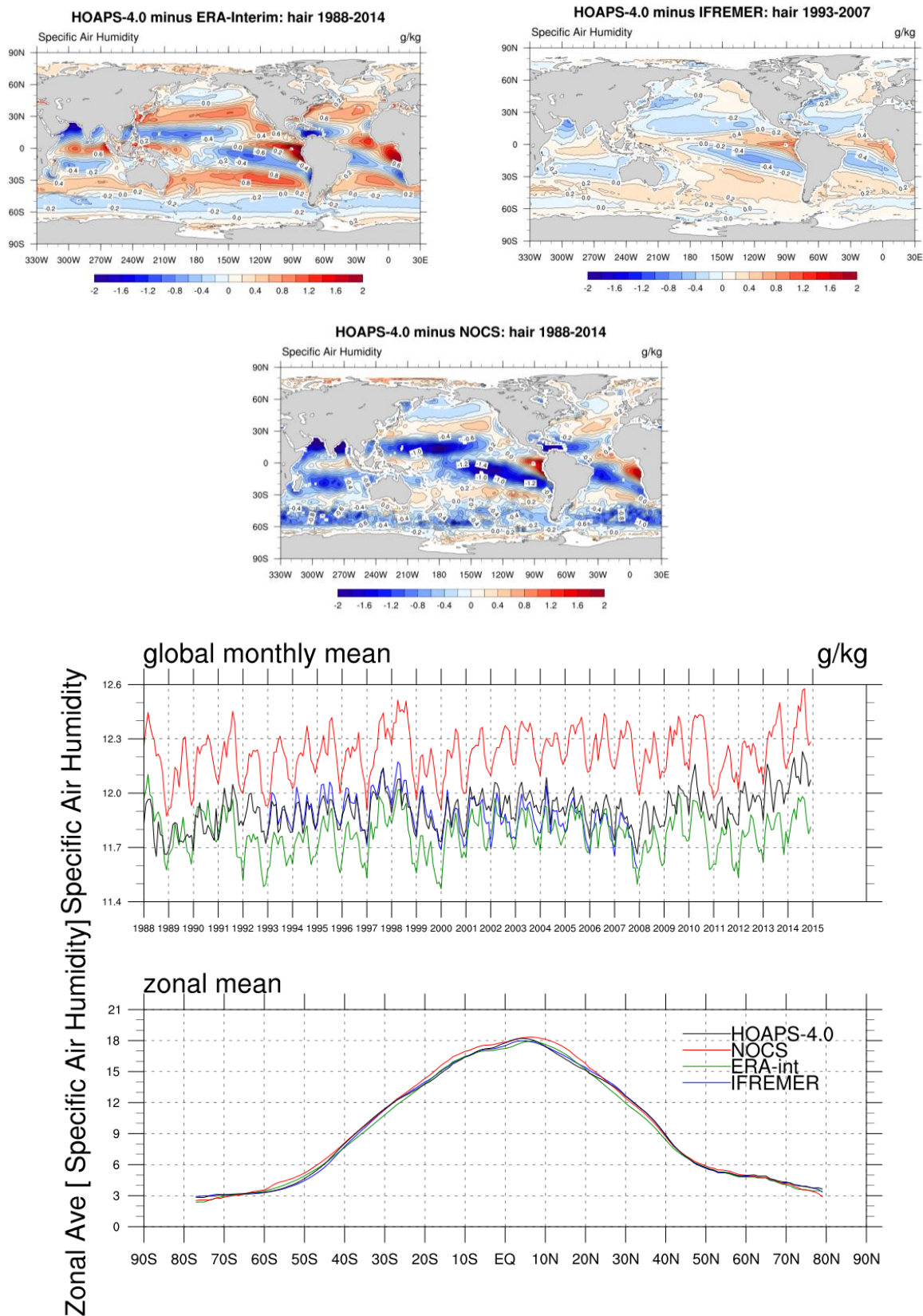
Figure 3: Difference of the 1988-2014 climatological mean HOAPS-4.0 near surface specific humidity and ERA Interim (upper left), NOCS (middle), and IFREMER (upper right, 1993-2007). The lower panels show the global monthly mean humidity time series of each data record (top) and the zonal mean humidity (bottom) for the overlapping time period 1988-2014 (1993-2007 for IFREMER).

The global mean time series of all data records (Figure 3, top of lower panels) differ in variability (global mean standard deviation of HOAPS-4.0: 0.10 g/kg; equivalent of NOCS: 0.19 g/kg), whereas HOAPS-4.0 magnitudes are slightly above the ERA-Interim values and concurrently below the NOCS estimates. At the same time, IFREMER estimates closely resemble those of HOAPS-4.0, which was to be expected. The zonal mean values of all data records (Figure 3, bottom of lower panels) show very similar characteristics. Specifically between 20°N/S, the NOCS data record is generally moister compared to both reanalysis and satellite products. Overall, the NOCS monthly global mean q_a values are on average 0.31 g/kg (3%) higher compared to HOAPS-4.0. ERA-Interim exhibits the lowest zonal mean values in the equatorial region as well as around 30˚ north and south, which results in global mean values that are on average 0.14 g/kg lower compared to HOAPS-4.0.

6.2.2 Discussion

The comparison of q_a from IFREMER and HOAPS-4.0 exhibits smallest differences because of the application of the same algorithm for microwave imager data. Hence, the deviations originate either from different sensor calibrations in the individual SSM/I and, from 2005 onwards, SSMIS brightness temperature records or from a different sampling due to the kriging technique used in the IFREMER dataset.

Larger deviations in q_a are locally observed when comparing HOAPS-4.0 with ERA-Interim and specifically NOCS. In particular, over the subtropical regions, a strong negative bias in the HOAPS-4.0 satellite retrieval compared to NOCS is evident. Jackson et al. (2009) found similar patterns in the comparison of different q_a satellite retrievals with ICOADS ship and buoy data. This dry bias in the range of 15-18 g/kg (that is, subtropics) has already been resolved for HOAPS-3.2 in Kinzel et al. (2016) and is also seen for HOAPS-4.0 (see section 7.2 (Figure 18a)). Comparing HOAPS-4.0 with ERA-Interim, this bias is less pronounced, since ERA-Interim is generally dryer in the tropical regions compared to NOCS (Andersson et al., 2011). In context of investigating LHF patterns of satellite and reanalysis data sets (HOAPS-2 and ERA-40, amongst others), Liu and Curry (2006) attribute observed LHF differences over (sub-) tropical regimes mainly to prevailing differences in q_a estimates.

Apart from the tropics, the poor sampling of the ship observations lead to larger uncertainties in the NOCS data record over the Southern Ocean (e.g., Berry and Kent, 2011). Especially during cold seasons, ground observations are sparse, leading to large deviations between the data records. Climatological averages (1988-2014) of these total uncertainties in NOCS reach 1.5-2 g/kg over the Southern Ocean (not shown). Global mean biases in NOCS show a more or less linear increase from 0.7 g/kg (1988) to 0.9 g/kg (2014). This uncertainty increase in the reference data record coincides with an increase in bias between NOCS and HOAPS-4.0 for 1996-2004 (Figure 3, top of lower panels). Over the North Atlantic and Pacific, where the sampling density is higher, this problem is not evident. However, the accuracy of satellite retrievals also depends on the representativeness of spatiotemporal variability of their a priori data used during the development of the algorithm. In presence of strong precipitation the retrieval is not applied. The associated sampling effect impacts the bias (fair weather bias) and might partly explain to the observed bias patterns in (Figure 3)

Based on the validation against the monthly mean in situ based NOCS data record it can be concluded that the HOAPS bias (RMSD) in q_a fulfils the target (optimal) requirement with a mean bias (RMSD) of -0.31 g/kg (0.21 g/kg) (Table 6-1, PRD [RD 1]).

Table 6-1: Requirements for near surface humidity product CM-12901 as given in the PRD [RD 1]. Accuracy numbers are given for global mean values. Regional larger deviations may occur.

	Threshold	Target	Optimal
Bias	1.20 g/kg	0.60 g/kg	0.30 g/kg
RMSD	2.40 g/kg	1.20 g/kg	0.50 g/kg
Decadal stability	0.20 g/kg	0.10 g/kg	0.04 g/kg

6.3 Near surface wind speed

6.3.1 Results
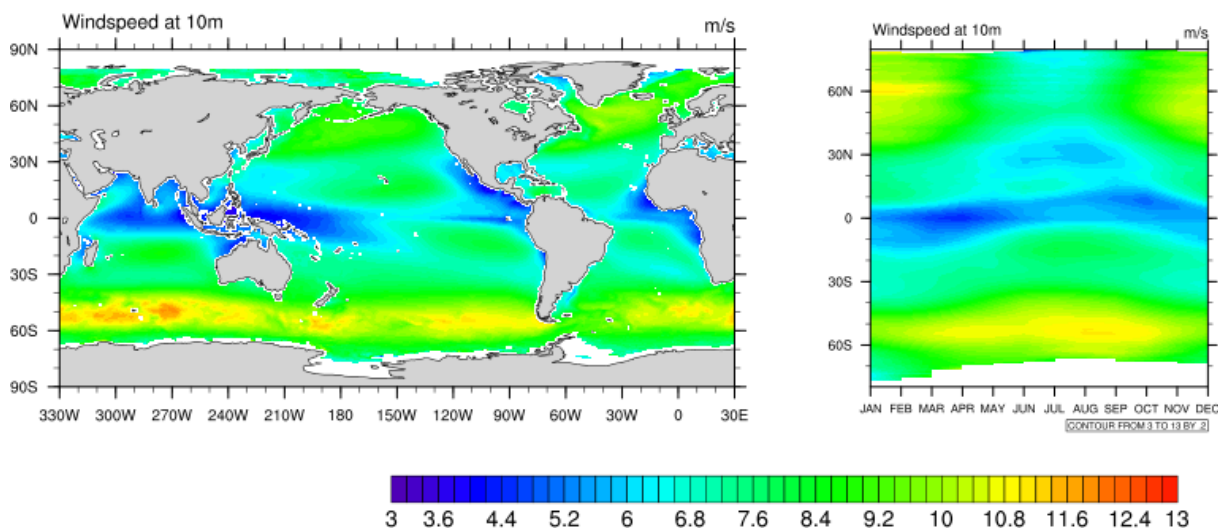
Figure 4: Climatological mean field (left) and zonal mean annual cycle (right) of HOAPS-4 near surface wind speed for the years 1988–2014.

HOAPS-4.0 climatological mean wind speed for the years 1988–2014 is shown in the left panel of Figure 4. North Atlantic and Pacific storm-track regions as well as the ''roaring forties'' and ''furious fifties'' over the Southern Ocean are characterized by maximum climate mean values of up to 13 m/s. Secondary local maxima exist in the tropical trade wind area. Moreover, the characteristic minima of the subtropical calms and the Southeast Asian warm pool region are clearly evident. The zonal mean annual cycle (Figure 4, right) highlights the wintertime maxima of wind speed in the mid- and high latitudes of both hemispheres, while only weak variability occurs in the subtropical regions.

The comparison of HOAPS-4.0 wind speed with ERA-Interim and NOCS is depicted in Figure 5. Red colours indicate regions where HOAPS-4.0 exhibits on average higher values, while in blue shaded regions HOAPS-4.0 underestimates wind speeds.

ERA-Interim wind speeds are generally lower compared to HOAPS-4.0 over the global oceans. This finds expressions in a global mean bias of 0.20 m/s. Apart from the overall bias, distinct differences occur in the intertropical convergence zone (ITCZ) as well as 30°N/S over all ocean basins, where ERA-Interim is considerably lower compared to HOAPS-4.0. Negative biases, locally as large as -2.8 m/s, are found along the western coasts of the subtropical continents and in monsoon regimes, e.g. off the Arabian and Indian Peninsula. Likewise, a general underestimation of HOAPS-4.0 wind speed dominates the southern oceanic storm track region.
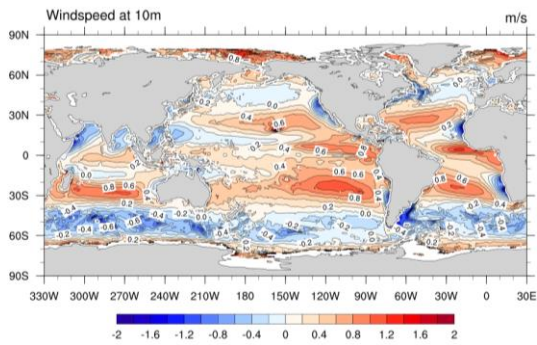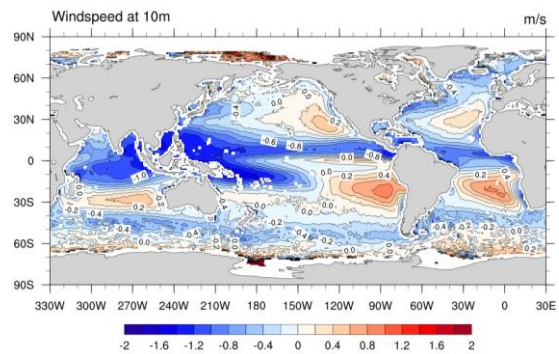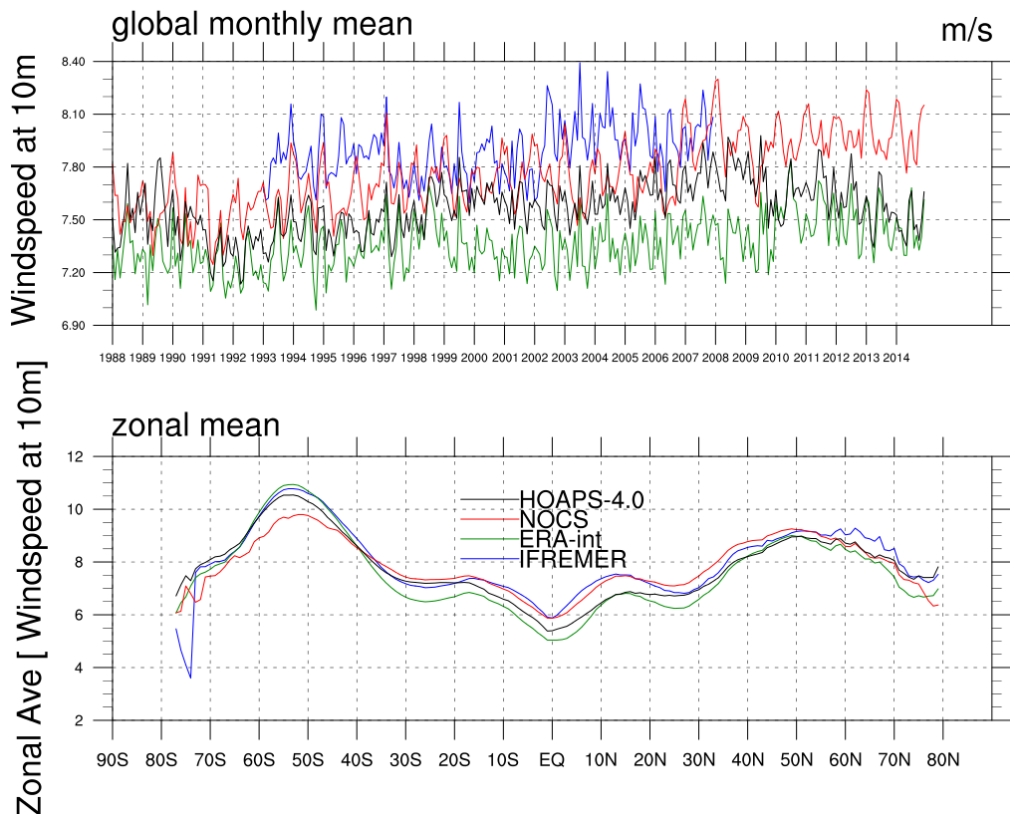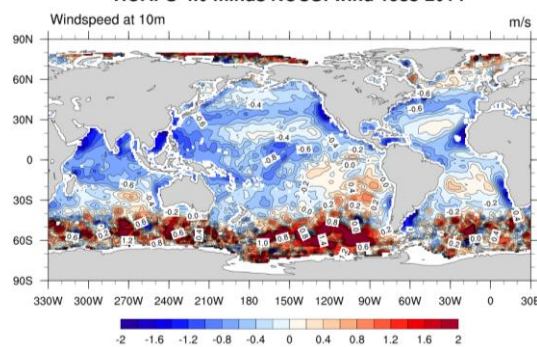
Figure 5: Difference of the 1988–2014 climate mean HOAPS-4.0 wind speed and ERA-Interim (upper left), NOCS (middle), and IFREMER (upper right, 1993-2007). The lower panels show the global monthly mean wind speed time series of each data record (top) and the zonal mean wind speed (bottom) for the overlapping time period 1988-2014 (1993-2007 for IFREMER).

Comparing HOAPS-4.0 with NOCS (Figure 5, middle) reveals largely different patterns. Negligible or (generally small) negative biases are found throughout the (sub-) tropical oceans. HOAPS-4.0 overestimations are almost exclusively confined to the extratropical storm track regions over the Southern Ocean. In this region, biases frequently exceed 2 m/s on wider ranges. Similar to the comparison to ERA-Interim HOAPS-4.0 exhibits underestimations relative to NOCS at the western boundaries of the continents. Locally, largest underestimations are in the order of 3 m/s. Despite the large positive biases over the Southern Ocean, the global mean bias is slightly negative and takes on a value of -0.19 m/s. The corresponding RMSD is given by 0.21 m/s.

A noticeable bias in the tropical western Pacific and warm pool region is evident in comparison to IFREMER (Figure 5, upper right), which coincides with low absolute wind speeds in a climatological sense. Here, HOAPS-4.0 underestimates wind speeds by 1.5-2 m/s. This also accounts for the Atlantic ITCZ regime as well as northern parts of the Indian Ocean. This pattern also coincides with maxima in precipitation. Positive biases are found over stratus regions off the west coasts of all continents. Note that the small regions of underestimation by HOAPS-4.0 at the coasts are also evident here, though less pronounced and smaller in region. The global mean deviation equals to -0.31 m/s.

Table 6-2: Requirements for near surface wind speed product CM-12911 as given in the PRD [RD 1]. Accuracy numbers are given for global mean values. Regionally, larger deviations may occur.

	Threshold	Target	Optimal
Bias	1 m/s	0.6 m/s	0.3 m/s
RMSD	1.6 m/s	0.8 m/s	0.5 m/s
Decadal stability	0.24 m/s	0.12 m/s	0.03 m/s

As described in context of Figure 5 (top left), the time series (Figure 5, top of lower pannels) indicate that ERA-Interim wind speeds are exclusively below the estimates of the remaining data records (see also Kent and Berry, 2013). This is also well represented in the zonal mean illustration (Figure 5, bottom of lower panels). Owing to the large-scale wind speed overestimation of IFREMER in the Indo-Pacific region, its global mean estimates are largest. By contrast, HOAPS-4.0 and NOCS climatological means are very similar, as the NOCS overestimations between 40°N/S and concurrent underestimations over the polar oceans almost cancel each other out. This opposed behaviour is clearly displayed in the zonal mean presentation. The standard deviations of their time series are almost identical, taking on a value of 0.21 m/s. The strong negative outlier of IFREMER wind speeds close to the Antarctic continent is likely associated with retrieval issues in the vicinity of the sea ice edge.

6.3.2 Discussion

The underestimation of ERA-Interim wind speed compared to all other data records is known from previous studies comparing satellite-retrieved and reanalysis wind speeds (e.g. Meissner et al. 2001; Kelly et al. 2001; Monahan 2006). The systematic deviations are caused by the different principles used to determine the wind speed. Inferred surface wind stress from satellite observations is often transformed to represent 10-m equivalent neutral-stability wind speed. In contrast to that, reanalysis models assimilate SSM/I and SSMIS radiances and scatterometer wind speed observations and then analyse and forecast the actual winds at 10 m. Another general source for systematic differences is that the reanalyses implement a static sea

surface, while satellite measurements are sensitive to ocean surface currents and measure the wind speed relative to the underlying sea surface (e.g. Meissner et al. 2001; Kelly et al. (2001). Additionally, regionally limited measurements acquired by radiosondes and the radiative transfer calculations underlying the satellite retrieval algorithms as well as the reanalyses lead to locally different results in the wind speed. The large differences over the monsoon regions of the Bay of Bengal and the Arabian Sea are likely to originate from lack of input data representing the specific atmospheric and sea surface properties in these regions due to atmospheric advection and oceanic upwelling (Andersson et al., 2011).

With the exception of the ITCZ region (and specifically the warm pool regime), HOAPS-4.0 and IFREMER exhibit an overall better agreement with biases mostly below 0.5 m/s. The noticeable negative biases in the tropics are assumed to be linked to the frequently strong precipitation in this region, which hampers the retrieval of wind speed using microwave radiometers, leading to gaps in the wind speed data. Filling these gaps with scatterometer-derived wind speeds as done in the IFREMER data record may explain the biases seen in Figure 5 (upper right) since the scatterometer wind speed algorithm can also be strongly affected by precipitation under certain circumstances. Under low wind speeds and when the scatter from the sea surface is low, additional volume scattering of even light precipitation leads to a spurious wind signal (e.g., Tournadre and Quilfen 2003; Wallcraft et al. 2009). An inverse effect is observed for high wind speed regimes. A similar, but considerably weaker, effect may occur in the HOAPS-4.0 data because of strong precipitation, which inhibits the retrieval of wind speed from the satellite observations (Wentz 1997; Andersson et al. 2010b). However, this affects usually only the inner cores of precipitating weather systems and hence only a limited number of satellite observations. In regions with frequent precipitation, such as the ITCZ or the Southern Ocean, this is the case for 10%−15% of the observations and might thus partly explain the bias observed in the ITCZ.

The systematic underestimation of NOCS wind speeds in high southern latitudes is mirrored in the random error estimate given in the data record, which reaches 3-4 m/s over the extratropical southern latitudes due to sparse data sampling (not shown). The limited number of observations from these regions in the NOCS data record leads to a bias toward low wind speeds since ships tend to avoid storms and high sea state related to winds, particularly during the cold season (e.g. Berry and Kent, 2009). Between 1988 and 2014 a nearly linear increase in NOCS global mean random uncertainty from 1.7 m/s to 2.4 m/s is observed (not shown). From 1988-1996 and from 2008 onwards, the biases between HOAPS-4.0 and NOCS increases (Figure 5, top of lower panels), which coincideswith the increasing random uncertainty with NOCS wind speeds. A similar, but considerably weaker, effect may occur in the HOAPS-4 data because of strong precipitation, which inhibits the retrieval of wind speed from the satellite (Wentz 1997; Andersson et al. 2010b). However, this affects usually only the inner cores of precipitating weather systems and hence only a limited number of satellite observations. In regions with frequent precipitation, such as the ITCZ or the Southern Ocean, this is the case for 10%−15% of the observations and might thus partly explain the bias observed in the ITCZ.

Comparing HOAPS-4.0 based wind speeds against the NOCS in-situ reference data, the average bias (RMSD) of -0.19 m/s (0.21 m/s), suggests that the optimal requirements defined in the PRD document are met (Table 6-2, PRD [RD 1]).

6.4 Evaporation and Latent heat flux

6.4.1 Results

The mean global ocean evaporation (Figure 6, left panel) shows the well-known climatological distributions with strong maxima over the subtropics of both hemispheres with values of up to 7 mm/d. Mid- and high latitudes exhibit generally lower values of less than 3 mm/d with the exception of the warm boundary currents of the Kuroshio Current, the Gulf Stream, and the Agulhas Current. The Gulf Stream generates the highest mean evaporation values on the globe of up to 8 mm/d. A pronounced seasonal variability can be identified in the climatological zonal mean annual cycle (Figure 6, right panel) with maximum evaporation values in the trade wind belts and secondary maxima in the mid- and high-latitude storm-track regions during the winter season of each hemisphere.

The difference in evaporation to all data records is shown in Figure 7. Generally higher differences are found in regions of large evaporation (specifically over the subtropical Pacific) and smaller values in regions with low evaporation.

In a broad band from the Kuroshio over the North Pacific to the North American west coast ERA-Interim and NOCS evaporation is systematically higher compared to HOAPS-4.0. In case of NOCS, this is associated with an underestimation of HOAPS-4.0 wind speeds (Figure 7). Regarding ERA-Interim, this is a consequence of an overrepresentation of HOAPS-4.0 q_a (Figure 3). This negative bias in evaporation continues southward along the Baja California. Over the cold tongue in the eastern equatorial Pacific and the Southeast Asian warm pool HOAPS-4.0 evaporation is also systematically lower compared to all other data records (specifically IFREMER), mainly owing to an overestimation of HOAPS-4.0 q_a.

Compared to both NOCS and ERA-Interim, an underestimation of HOAPS-4.0 evaporation is evident in the eastern tropical Atlantic with differences of more than 1 mm/d (up to 50%). This negative bias results from an overestimation of q_a along the West African coast (as over the eastern tropical Pacific, Figure 3) and an underestimation of q_s in the tropical and subtropical Atlantic. Andersson et al. (2011) hypothesized that it results from high desert aerosol loadings, which hinder the proper error correction within the SST retrieval algorithm. By contrast, q_a biases in relation to IFREMER are considerably smaller (Figure 3), as are the magnitudes of the negative evaporation biases in this region (Figure 7, upper right).

In comparison to NOCS, the southern extratropics appear exceptional (Figure 7, middle), as the bias magnitude is large compared to relatively low average evaporation values. Locally, the mean evaporation of NOCS is more than 1.5 mm/d below HOAPS-4.0 values, which corresponds to a relative difference of more than 50% in these regions. These deviations coincide with the southern hemisphere's band of strong winds between 40° and 60°S, which is considerably more pronounced in HOAPS-4.0 compared to NOCS (compare Figure 5). In relation to ERA-Interim, HOAPS-4.0 evaporation is up to 0.75 mm/d higher (i.e., more than 30 %) over the southern midlatitudinal storm tracks. This is mainly related to differences in q_s - q_a (not shown), since HOAPS-4.0 wind speeds are even slightly lower than those of ERA-Interim (Figure 5).

Over the North Atlantic and North Pacific between 40° and 80°N the comparison between the data records shows mixed results with differences that are mostly below 0.5 mm/d. While

ERA-Interim slightly overestimates evaporation over the North Atlantic, NOCS evaporation is underrepresented over higher latitudes of the North Pacific. By contrast, evaporation estimates of IFREMER are almost invariably higher than HOAPS-4.0.

Owing to the overestimation of IFREMER wind speeds over larger tropical areas centered over the warm pool, bias magnitudes of evaporation frequently exceed 1 mm/d in an absolute sense and are comparable to those associated with NOCS in these regions. In consequence, the global mean bias of HOAPS-4.0 evaporation with respect to IFREMER equals to -0.50 mm/d. Yet, apart from the Indo-Pacific region, other bias magnitudes (both positive and negative) are smaller compared to ERA-Interim and NOCS. In comparison to HOAPS-4.0 their biases yield to -0.40 mm/d and -0.28 mm/d, respectively.

In conclusion, regarding ERA-Interim and NOCS (Figure 7, top left and middle), the differences are largely determined by the humidity fields (compare Figure 3). ERA-Interim wind speeds are even larger (smaller) over the southern hemispheric storm track and off Arabia (off the tropical west coasts of Africa and South America), although corresponding evaporation estimates are smaller (larger) compared to HOAPS-4.0 (Figure 5, Figure 7). Focusing on NOCS, the footprint of wind speed biases is also mirrored in respective evaporation differences, although the signal is rather damped in comparison to the q_a biases. Near the Arabian Peninsula and the Caribbean (off Antarctica), the impact of humidity biases is specifically strong, as respective NOCS wind speeds are larger (smaller), although resulting evaporation estimates are positive (negative).

By contrast, the comparison with the IFREMER data record (Figure 7, upper right) clearly shows the footprint of the wind speed biases (Figure 5). The subtropical northern Atlantic/Pacific, along the western boundary currents, and the 60°S belt are striking. Here, IFREMER overestimates evaporation, although the q_a estimates exceed those of HOAPS-4.0. In fact, a dependency on q_a biases on evaporation differences is evident, yet considerably less apparent as seen for both ERA-Interim and NOCS.
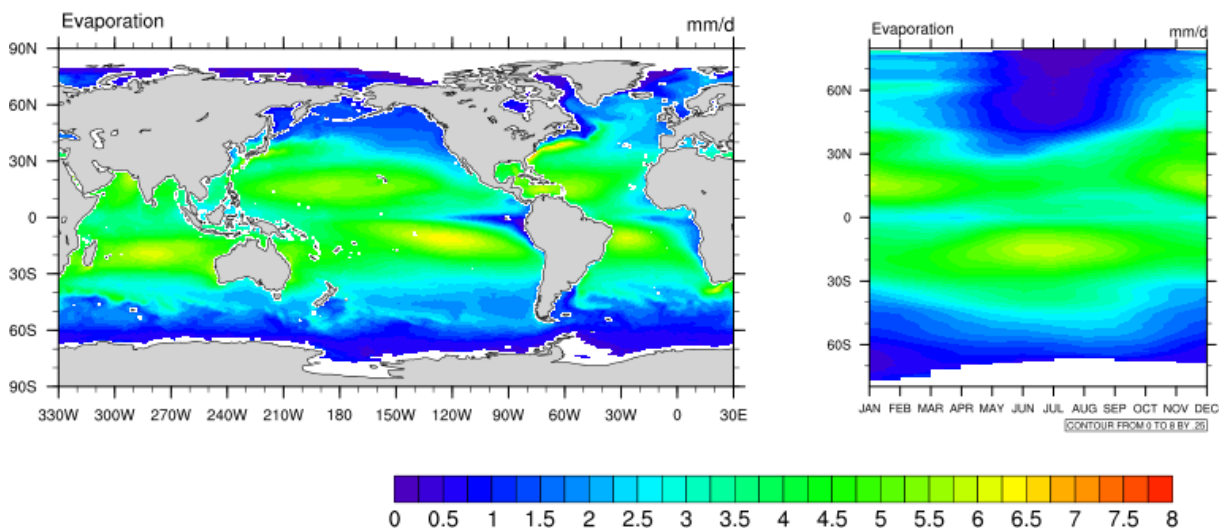
Figure 6: Climatological mean field (left) and zonal mean annual cycle (right) of HOAPS-4.0 evaporation for the years 1988–2014.

The comparison of the climatological zonal means (Figure 7, bottom of lower panels) shows an overall agreement in the location and magnitude of the maxima and minima of all data records. It clearly resolves the addressed underrepresentation of NOCS evaporation over the extratropics of the Southern Ocean (deviations of up to 50%). Over the northern hemispheric counterpart, this difference is negligible. Absolute differences of 1 mm/d between all data records are confined to the inner tropics. Specifically, HOAPS-4.0 evaporation is lowest over the inner tropics in comparison to all other data records, which is to a large extent associated with the HOAPS-4.0 underrepresentation of evaporation along the tropical eastern ocean basins. Similar deviations occur along 30°N. Between 15°S and 20°N, IFREMER zonal averages are largest in comparison to the remaining data sets, what is also mirrored in the time series (Figure 7, top of lower panels). By contrast, ERA-Interim zonal averages are largest between 20-40° of both hemispheres.

With the exception of IFREMER, the magnitudes of the global monthly mean time series (Figure 7, top of lower panels) are in close agreement for all data records. As has been discussed in context of zonal means, HOAPS-4.0 remains below ERA-Interim and NOCS (mainly from 1992-2006 and after 2011) and specifically below IFREMER from 2002-2007. The sudden increase of the IFREMER time series in 2002 is likely to be an artefact from the wind or SST input data sources. However, this does not affect the general difference patterns. The standard deviation of the monthly mean time series of HOAPS-4.0 (NOCS) equals to 0.17   mm/d   (0.15   mm/d).
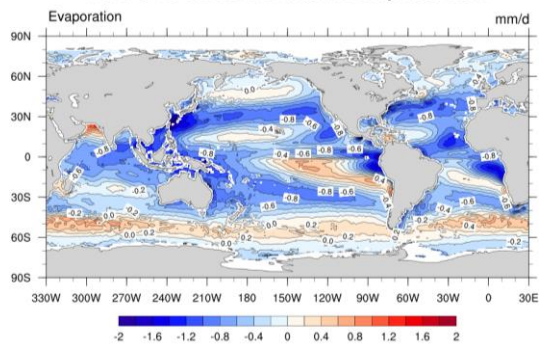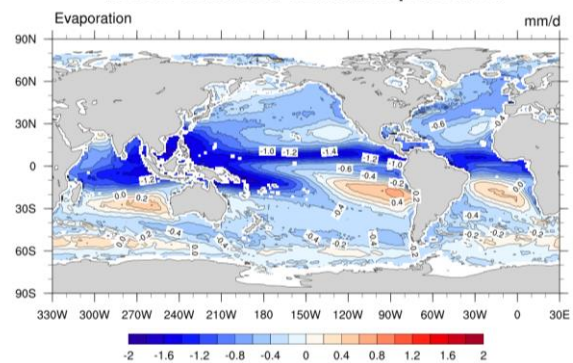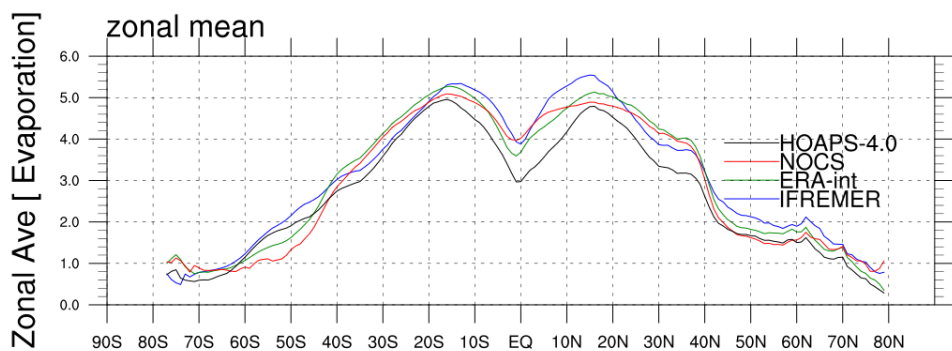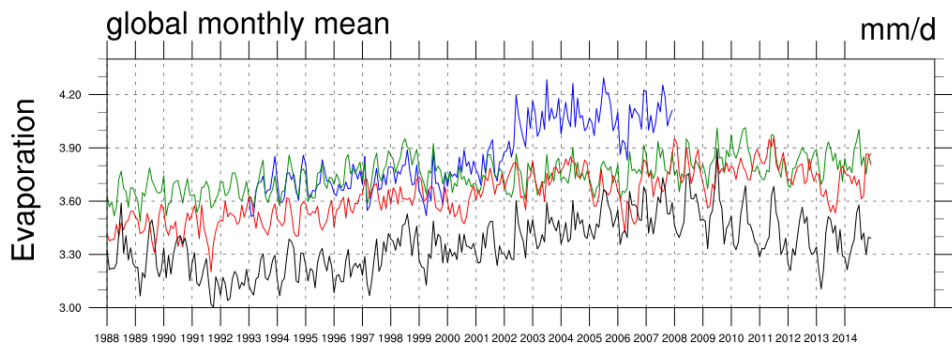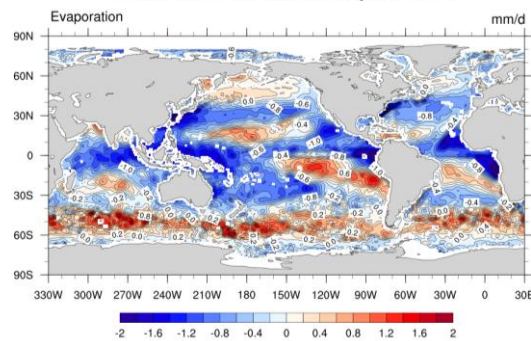
Figure 7: Difference of the 1988–2014 climate mean HOAPS-4.0 evaporation and ERA-Interim (upper left), NOCS v2.0 (middle), and IFREMER flux (upper right, 1993-2007). The global monthly mean evaporation time series of each data record (top of lower panels) and the zonal mean evaporation (bottom of lower panels) for the overlapping time period 1988–2014 (1993-2007).

6.4.2 Discussion

In the comparison with the three other data records, HOAPS-4.0 shows a overall positive bias in regions with high values of evaporation and negative bias in regions with low evaporation. When compared to the difference plots of wind speed and q_a, it appears that the large-scale deviations in evaporation in the tropical regions are primarily caused by differences of the q_a retrievals (Andersson et al., 2011). Similar conclusions have been drawn by Liu and Curry (2006) when comparing reanalysis and satellite LHF products over the (sub-) tropical regimes.

The differences in the wind speed and q_s are mostly of second-order importance, except for the strong positive bias of the IFREMER data record centred over the Indo-Pacific region. At higher latitudes, where the sea–air humidity difference is smaller, the influence of the wind speed increases. This targets the overestimation in NOCS wind speed in the extratropical North Pacific, for example, causing negative biases in evaporation. Regarding the influence of wind speeds on biases in evaporation, Brunke et al. (2011) assess the uncertainties in LHF in numerous reanalysis and satellite-derived data sets and point out that the variability in biases (compared to globally distributed ship measurements) increase in strong wind regimes.

The underestimation of HOAPS evaporation (and thus LHF) compared to the reference data sets has already been picked up in earlier studies. Specifically in lower latitudes, Brunke et al. (2002), Chou et al. (2003, 2004), and Kubota et al. (2003) have demonstrated a underestimation of HOAPS-1 LHF in comparison to remotely sensed data, reanalyses, and observational data. Bourras (2006) and Liu and Curry (2006) demonstrated good performance of HOAPS-2 flux data. Specifically, Bourras (2006) compares HOAPS-2 to four other remotely-sensed LHF data sets and concludes that HOAPS-2 is the most appropriate product to study turbulent fluxes over the world oceans. Similar reasoning is done in Liu et al. (2011) for HOAPS-3 over the Southern Ocean. HOAPS-3 is furthermore assessed by Andersson et al. (2011), who conclude that global mean time series of satellite, reanalysis, and in-situ data often agree within 10% of the individual products and HOAPS-3 tends to agree better with other satellite-derived data sets than with observational and reanalysis data.

As in Andersson et al. (2011), the validation results of HOAPS-4.0 evaporation indicate that the compared products differ less than 10-15% for most of the investigated time period from 1988 to 2014 (Figure 7, top of lower panels). This is in accordance with results for global mean estimates of Trenberth et al. (2009) who found similar differences between satellite- and model-based data records. However, the time series and zonal averages shown in Figure **7**(lower panels) suggest that this does not apply to the bias of HOAPS-4.0 to IFREMER, which has increased by 0.24 mm/d in comparison to HOAPS-3.2 (compare Figure 1).

The reasons of uncertainties in the evaporation estimates point at uncertainties associated with the retrieval of wind speed, q_a, and SST that are affected by precipitation and clouds. Depending on the methodology and sampling density this may lead to errors in the absolute values and the temporal variability in regions with persistent cloud cover and frequent precipitation. Wind speed and q_a cannot reliably be retrieved under strong precipitation. Similar to the SST, missing values may be interpolated as it is done in the IFREMER product by a kriging procedure. In HOAPS-4.0 the missing values for q_a and wind speed are not interpolated resulting in a considerably lower number of evaporation observations in regions with frequent precipitation. The strongest effect is observed over the Southern Ocean, the ITCZ, and tropical warm pool, where in $10\%-15\%$ of all SSM/I and SSMIS observations the

retrieval of wind speed, q_a and hence evaporation is not possible. Precipitation-related retrieval issues are also picked up in Andersson et al. (2011), for example.

Table 6-3: Requirements for evaporation (CM-12801) and latent heat flux (CM-12811) products as given in the PRD [RD 1]. Accuracy numbers are given for global mean values. Regionally, larger deviations may occur.

	Threshold		Target		Optimal	
Bias	0.7 mm/d	20 W/m^2	0.36 mm/d	10 W/m^2	0.09 mm/d	2.5 W/m^2
RMSD	1.24 mm/d	35 W/m^2	0.62 mm/d	17 W/m^2	0.53 mm/d	15 W/m^2
Decadal stability	0.32 mm/d	9 W/m^2	0.14 mm/d	3.9 W/m^2	0.0043 mm/d	0.12 W/m^2

Finally note, that the cold bias in the SST data record, which according to Andersson et al. (2010b) caused the minimum in HOAPS-3.2 after the eruption of Mount Pinatubo, has considerably been reduced in HOAPS-4.0 (see Figure 7, top of lower panels).

Based on the comparison of HOAPS-4.0 monthly mean evaporation (latent heat flux) against the in situ based NOCS data, it can be concluded that the requirements defined in the product requirement document are met with a mean bias of -0.28 mm/d (7%, target) and a RMSD value of 0.24 mm/d (6%, optimal). The corresponding values for latent heat flux are - 7.6 W/m^2 (target) and 6.5 W/m^2 (target), respectively (Table 6-3, PRD [RD 1]).

6.5 Precipitation

6.5.1 Results

The climatological mean precipitation from HOAPS-4.0 (Figure 8, left panel) well represents the known global distribution of precipitation. Dominant features are the overall highest rain rates in the ITCZ, exceeding 10 mm/d, and the regional maxima over the tropical Indian Ocean and the South Pacific conversion zone (SPCZ). The North Atlantic and Pacific storm tracks are also clearly identifiable with maximum values of up to 9 mm/d over the Gulf Stream and Kuroshio currents. Global precipitation minima can be observed in the so called subtropical oceanic deserts in the eastern subtropical Atlantic and Pacific.

HOAPS-4.0: rain 1988-2014
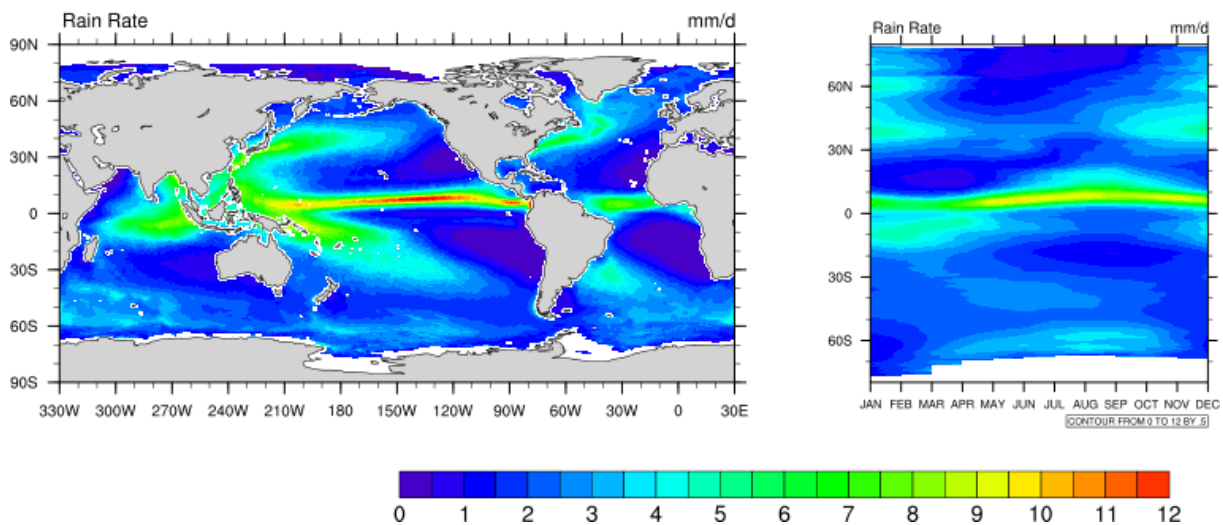
Figure 8: Climatological mean field (left) and zonal mean annual cycle (right) of HOAPS-4.0 precipitation for the years 1988–2014.

The zonal mean annual cycle (Figure 8, right panel) clearly shows the seasonal displacement of the ITCZ as well as the high precipitation values over the Northern Hemisphere storm tracks during the cold season. Similar features, although smaller in magnitude, are observed over the extratropical Southern Ocean. The development of the southern hemisphere subtropical maximum in the SPCZ between January and April is also evident.

The comparison of HOAPS-4.0 precipitation with ERA-Interim, TRMM 3B43 and GPCP is depicted in Figure 9. Red colours indicate regions where HOAPS-4.0 exhibits on average higher values, while in blue shaded regions HOAPS-4.0 underestimates precipitation. ERA-Interim precipitation is generally higher on a global scale compared to all satellite-derived products as depicted in the difference plot of HOAPS-4.0 and ERA-Interim, the zonal means and the global monthly mean time series of the data records in Figure 9 This bias originates mainly from the tropical belt, where ERA-Interim exceeds HOAPS-4.0 partly by more than 2 mm/d (which corresponds to 50% off Somalia). The issue of excessive tropical precipitation in ERA-Interim (e.g., Kim et al., 2013) is already known from the former ERA-40 reanalysis. However, the tropical moisture budget in ERA-Interim appears to be improved over ERA-40, for which this positive bias was even stronger (Simmons et al., 2007). Except for the large tropical biases, the deviations between HOAPS-4.0 and ERA-Interim are small and remain mostly below 1 mm/d (<20%). HOAPS-4.0 precipitation values are noticeably larger compared to ERA-Interim over the western boundary currents as well as over the ITCZ region of the central Pacific and the south-eastern tip of the SPCZ. Nevertheless, average HOAPS-4.0 precipitation is 0.47 mm/d below the ERA-Interim global mean.
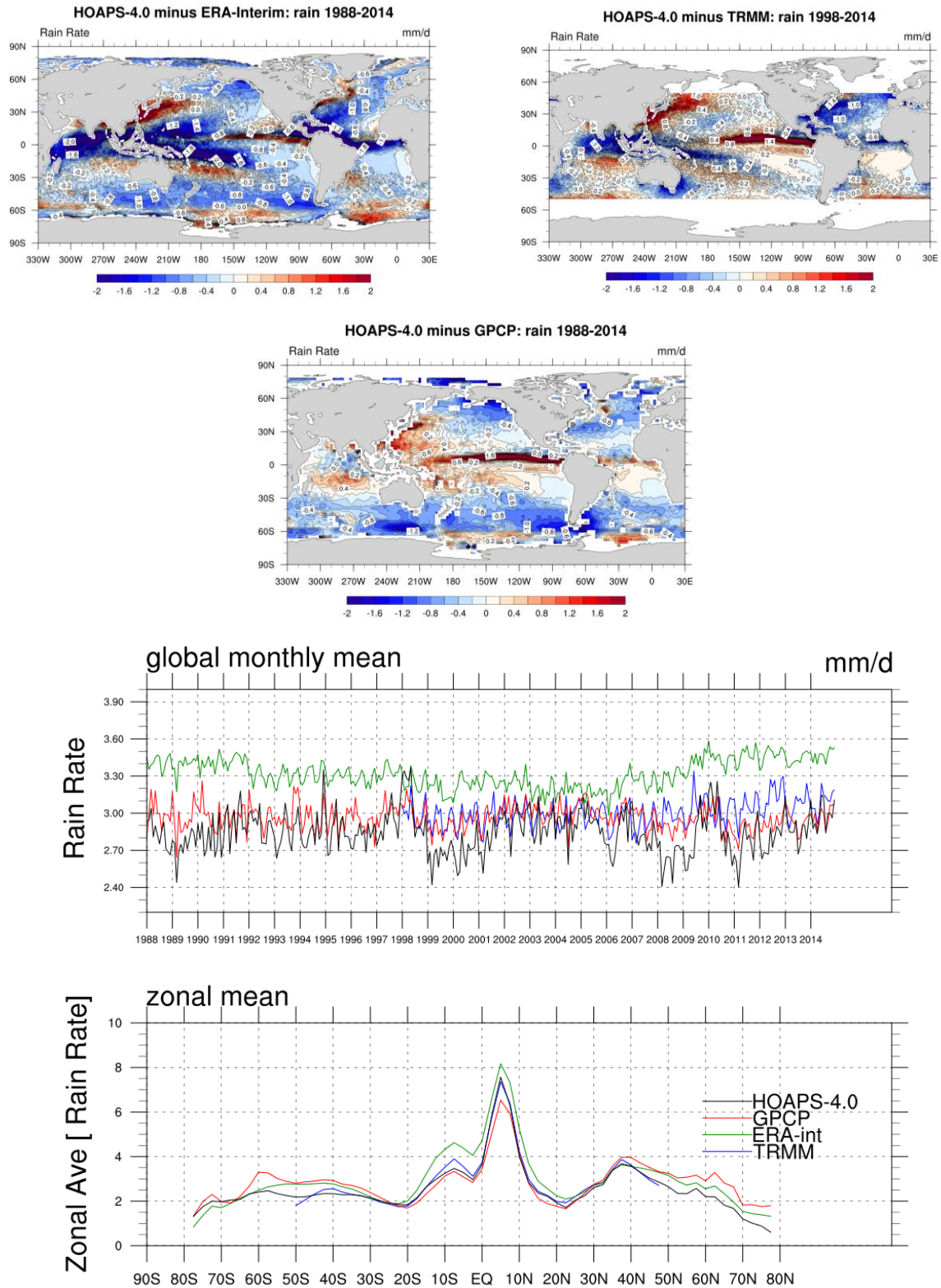
Figure 9: Difference of the 1988–2014 climate mean HOAPS-4.0 precipitation and (upper left) ERA-Interim, (upper right) TRMM 3B43 (1998–2014), and (middle) GPCP. The lower panels shows the global monthly mean precipitation time series of each data record (top, within ±40°N/S for TRMM) and the zonal mean precipitation (bottom) for the overlapping time period 1988–2014 (1998–2014 for TRMM 3B43).

Bias magnitudes between HOAPS-4.0 and TRMM 3B43 precipitation (Figure 9, upper right) often remain lower compared to the ERA-Interim differences, as deviations below 0.5 mm/d (5%−10%) occur for most regions. HOAPS-4.0 precipitation exceeds the TRMM estimates by up to 2 mm/d over the tropical central Pacific (global bias maximum) and relatively large differences are found in regions of high precipitation variability, i.e. over the western subtropical Pacific and over the central Indian Ocean. The overestimation along the SPCZ is less pronounced as in the comparison of HOAPS-4.0 and ERA-Interim. HOAPS also exhibits overestimations over the Kuroshio current. As average precipitation rates in the above-mentioned regions are comparatively large, relative deviation often remain below 20%. Over the entire North Atlantic basin, HOAPS-4.0 precipitation is systematically lower compared to the TRMM data, with the exception of the Gulf Stream region off Newfoundland. Other regions with lower precipitation in HOAPS-4.0 are found in the region south of Australia and parts of the northern Indian Ocean.

The comparison of HOAPS-4.0 with GPCP (Figure 9, middle) exhibits similar differences for the tropical belt as the comparison of HOAPS-4.0 and TRMM. Relative to TRMM the differences appear smoothed and HOAPS-4.0 values are larger over the SPCZ than GPCP by around 0.5 mm/d (10%). Similar to ERA-Interim and TRMM, the maximum deviation of about 1.5 mm/d (30%−40%) is found in the Pacific ITCZ. Over the mid–high latitudes between 40˚ and 70˚, the precipitation in GPCP is generally 10%−30% higher compared to HOAPS-4.0. Locally the values exceed 50%. Locally, however, extratropical biases are positive, e.g. south of Greenland and, as also indicated in Figure 9 (top left), close to the coast of Antarctica.

The global mean time series (Figure 9 top of lower panels) of the satellite-derived products exhibit values around 3 mm/d, while the ERA-Interim precipitation record tends to be approximately 0.4 mm/d above the satellite climatologies. The month-to-month variability of the HOAPS-4.0 record is slightly larger compared to the other data records. The standard deviation of the monthly global mean values of HOAPS-4.0 equals to 0.17 mm/d. As indicated by the time series, the corresponding GPCP variability is lower and takes on a value of 0.06 mm/d. The monthly mean global bias with respect to GPCP slightly decreases from 0.9 mm/d (1988) to approximately 0.8 mm/d (2014) (not shown). Specifically prior to 1995 and from 2005 onwards, a concurrent bias reduction between HOAPS-4.0 and GPCP occurs (Figure 9, top of lowerpanels). It is therefore suggested that the decrease in bias is linked to a reduction in the monthly mean random uncertainty of GPCP rainfall.

For the zonal means Figure 9, bottom of lower panels), the consistency among all products is best between 15-30°S and 15-40°N. Specifically equatorward of 20°S up to the northern branch of the ITCZ, ERA-Interim precipitation averages are well above the means of all other data sets, which is also mirrored in the time series as well as in the bias magnitudes shown in Figure 9 (top left). Here, latitudinal-mean biases are in the order of 1-2 mm/d. GPCC estimates are largest over the extratropical oceans of both hemispheres, followed by ERA-Interim and HOAPS-4.0 precipitation rates. Specifically in northern (sub-) polar regions, an increase in relative differences among the data sets can be observed.

6.5.2 Discussion

The comparison of HOAPS-4.0 precipitation with the ERA-Interim reanalysis and the satellite products GPCP and TRMM 3B43 exhibits considerable absolute differences in regions with high precipitation variability. Whereas variability patterns are often similar, substantial

disagreements in intensities occur, as has also been illustrated by Shin et al. (2011) on a global scale.

The largest absolute differences are found over the ITCZ, while the relative differences are largest at high latitudes, where precipitation amounts are less. This is in agreement with previous inter-comparison studies that included satellite-based as well as model-based precipitation estimates. These showed regionally large differences among the individual products that are up to 50% in regions of strong precipitation and at high latitudes (e.g., Adler et al. 2001 and Klepp et al. 2005). Similar conclusions are drawn by Tian and Peters-Lidard (2010), who demonstrate that largest uncertainties are located over high latitudes, specifically during the cold season. Particularly in the tropical regions model-based data (e.g., reanalysis products) are found to perform significantly poorer than satellite-derived fields (Trenberth and Guillemot 1998; Janowiak et al. 1998; Shinoda et al. 1999). Béranger et al. (2006) conclude that the western south Pacific is a region that exhibits large differences among a range of precipitation data records, which also include data of the ERA reanalysis, GPCP, and HOAPS-2.

HOAPS-4.0 precipitation turns out to be substantially higher in the Pacific ITCZ and the Kuroshio current compared to the other data records, while precipitation in subtropical regions often agrees better. Regarding ERA-Interim, this agrees with findings of Pfeifroth et al. (2013), who demonstrate that ERA-Interim overestimates small and medium precipitation in comparison to HOAPS-3, whereas high amounts are underestimated. Furthermore, the authors state that GPCP shows lowest values over the tropical western Pacific, which is mirrored in Figure 9 (middle).

At higher latitudes between 40˚ and 70˚ north and south, GPCP exhibits a known high bias, which has already been investigated in comparison to HOAPS-3 (Klepp et al. 2010). For these latitudinal bands, GPCP utilizes Television and Infrared Observation Satellite (TIROS) Operational Vertical Sounder (TOVS) infrared data to compensate deficiencies in the GPCP high-latitude microwave-based retrievals (Adler et al. 2003). At midlatitudes, the TOVS data are adjusted to the SSM/I and SSMIS estimates. Toward the poles the adjustment is transitioned to a bias adjustment based on rain gauges. At high latitudes from 70˚ and beyond the adjustment is done using land-based rain gauge data. In conclusion, the observed biases between HOAPS-4.0 and GPCP (both positive and negative, compare Figure 9 in the (sub-) polar oceans need to be treated with care.

Comparisons with TRMM products should give deeper insight in the quality of HOAPS-4.0 precipitation values because of the calibration of the TRMM product with the precipitation radar. In general, the HOAPS-4.0 precipitation is slightly smaller than the TRMM product. The previously observed conspicuous decrease in the TRMM 3B43 time series since 2003 is not evident anymore in this TRMM 3B43 V7. However, the TRMM time series of global mean precipitation exhibits a small increase starting in 2011 which is not present in all other data records. This contributes to the observed bias but does not fully explain the bias. The reason for the seeming increase is unclear.

Because of the lack of reliable in situ measurements, detailed quantitative comparisons for oceanic precipitation are rare and validation efforts are still mostly limited to short period regional inter-comparison studies. Moreover, the strong spatial and temporal variability of the precipitation complicates such validation efforts. In particular the availability of reliable ground data for validation studies is very limited. The only frequent measurements in the central Pacific are taken by several rain gauges on buoys of the Tropical Atmosphere Ocean (TAO) project and precipitation radars on atolls. The representativeness of measurements

from these systems is limited by their spatial restriction and the need of wind corrections for gauge under-catchment. But as these are the only available precipitation data record, several studies evaluated satellite-based precipitation products using the atoll and buoy data. The results indicate a possible systematic underestimation of inner tropical precipitation up to 15% by various satellite retrievals (Adler et al. 2001, 2003; Bowman et al. 2009; Sapiano and Arkin 2009). In contrast to that, the HOAPS-4.0 retrieval exhibits mostly higher mean precipitation values in this region, specifically over the Central and Eastern tropical Pacific.

In an absolute sense, Pfeifroth et al. (2013) show that average monthly deviations (1989-2005) between tropical Pacific rain-gauge atoll stations (PACRAIN) and ERA-Interim, GPCP, and HOAPS-3 are in the range of 20-30%. Whereas GPCP (HOAPS-3) is on average 12% (7%) below PACRAIN, precipitation estimates of ERA-Interim are 9% larger. Advances regarding the reliability of reference data have recently been achieved in the framework of the oceanic shipboard precipitation measurement network for surface validation (OceanRAIN, Klepp, 2015) and will be considered in near future. It comprises a continuously growing high-quality in-situ precipitation data record based on optical disdrometers installed on research vessels, which is suitable for validation purposes.

For the mid–high latitudes detailed case study analyses on mid latitude cyclones with intense postfrontal mesoscale convective mixed-phase precipitation were carried out by Klepp et al. (2003). Utilizing in situ voluntary observing ship data, it was shown that, in contrast to other satellite products, HOAPS-1 recognizes small-scale intensive precipitation systems in cold air outbreaks with reliable patterns and intensities. This type of precipitation is also mostly missing in a large sample of events investigated in the ECMWF numerical weather prediction and ERA-40 reanalysis data records (Klepp et al. 2005). Furthermore, Klepp et al. (2010) demonstrates the ability of HOAPS-3 to detect even light amounts of cold season snowfall with a high accuracy of 96% between point-to-area collocations of ship-based optical disdrometers and satellite data.

Also for the North Atlantic region, Andersson et al. (2010a) carried out an analysis of the HOAPS-3 precipitation variability connected to the North Atlantic Oscillation. It is shown that the response of precipitation structures to the atmospheric fluctuations is well represented in HOAPS-3 and that the HOAPS-3 precipitation fields showed convincing consistency with land-based rain gauge data in magnitude and variability. It is likely that similar conclusions may be drawn using HOAPS-4.0, as both versions include the same precipitation algorithm.

Table 6-4: Requirements for precipitation product CM-12611 as given in the PRD [RD 1]. Accuracy numbers are given for global mean values. Regional larger deviations may occur.

	Threshold	Target	Optimal
Bias	0.6 mm/d	0.3 mm/d	0.15 mm/d
RMSD	1.0 mm/d	0.5 mm/d	0.25 mm/d
Decadal stability	0.034 mm/d	0.02 mm/d	0.004 mm/d

HOAPS-4.0 exhibits minima in precipitation in 1999, 2006, 2008 and 2011. These minima temporally coincide with La Nina events and periods. They are hardly evident in the references data but were also observed in HOAPS-3.2 (1999 only due to limited temporal coverage). From a theoretical view point the global mean precipitation is expected to change

according to changes in SST (O'Gorman et al., 2012) and in atmospheric dynamics (Trenberth, 2011). Evaluating results from climate models the change in precipitation with the change in SST is between 1 %/K and 2.3 %/K (Liu and Allan, 2013). Also, the upcoming version of GPCP, i.e., version 3, will be based on the latest GPROF version. Figure 3 in Kummerow et al. (2015) exhibits a fairly strong variability in precipitation between El Nino and La Nina events, i.e., the amplitude is close to 0.4 mm/d and thus considerably stronger than the amplitude observed for GPCP as depicted in Figure 9.

Because of the lack of reliable in situ measurements the validation of HOAPS-4.0 monthly mean precipitation is done against other satellite based climatologies and reanalysis data. The well established GPCP product is taken as the reference, because the TRMM product time series is relatively short. Comparing HOAPS-4.0 precipitation against GPCP data depicts a mean bias of -0.12 mm/d and a RMSD value of 0.24 mm/d and it can be concluded that the optimal requirements defined in the product requirement document are met (Table 6-4, PRD [RD 1]).

6.6 Freshwater Flux

6.6.1 Results

The difference between evaporation and precipitation yields the oceanic freshwater flux into the atmosphere. Dominant features of either precipitation or evaporation fields determine the resulting global distribution of the freshwater flux as shown in Figure 10. A net flux into the ocean is mainly found in regions of precipitation maxima in the ITCZ, the midlatitude storm tracks, and at high latitudes. In contrast, subtropical regions generate the major part of the freshwater flux into the atmosphere. In the annual cycle, the dominant features of the input parameters are reproduced.

The difference map of the climatological mean fields of HOAPS-4.0 and ERA-Interim (Figure 11, top left) is mostly an inversion of the difference between the precipitation data records shown in Figure 9. In the inner tropics, the atmospheric freshwater deficit of ERA-Interim exceeds HOAPS-4.0 by up to 2 mm/d (central tropical Pacific, Caribbean), while the deficit in HOAPS-4.0 is larger in the eastern Pacific ITCZ and around 30˚ north and south. In the eastern

Pacific as well as in the Atlantic the difference in the freshwater flux is mostly determined by the evaporation fields. On a global average, the ERA-Interim freshwater deficit in comparison to HOAPS-4.0 equals to 0.10 mm/d.

The differences between HOAPS-4.0 and the combination of the IFREMER evaporation and GPCP precipitation (Figure 11, top right) shows the underestimation of HOAPS-4.0 along the ITCZ and the Indo-Pacific region in general, with biases in the order of 1.5-2 mm/d (in an absolute sense).

HOAPS-4.0: budg 1988-2014
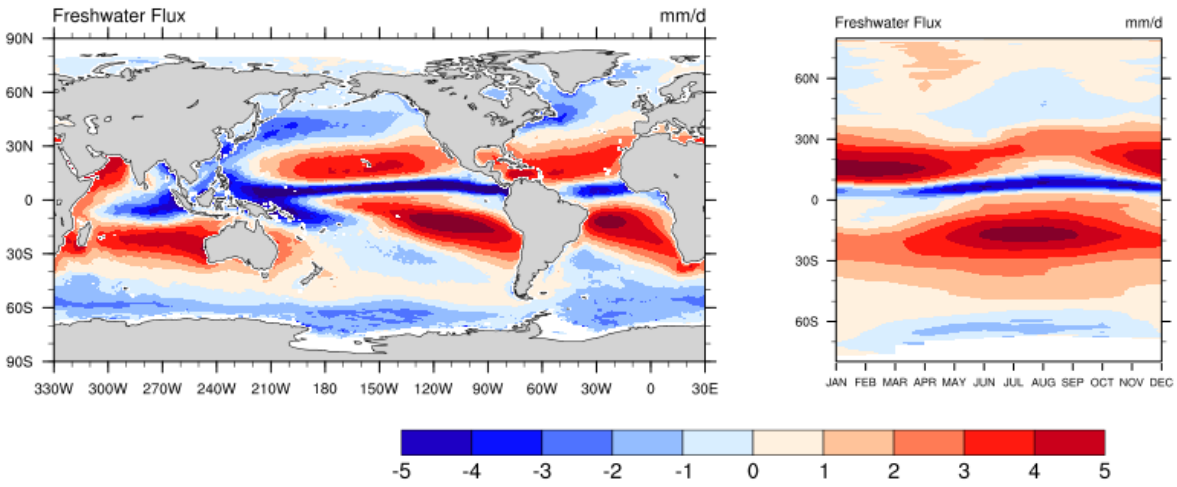
Figure 10: Climatological mean field (left) and zonal mean annual cycle (right) of HOAPS-4.0 freshwater flux for the years 1988–2014.
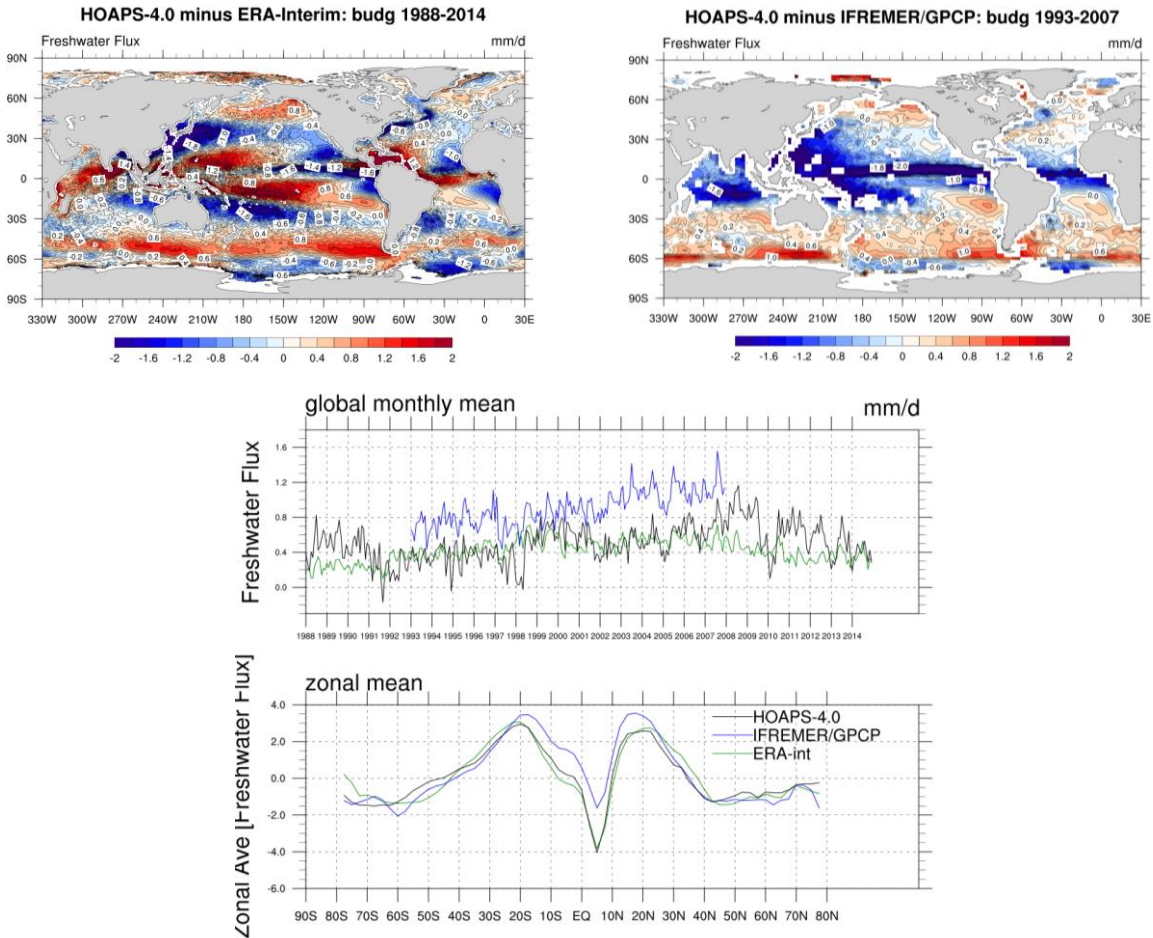
Figure 11: Difference of the 1988–2014 climate mean HOAPS-4.0 freshwater flux to (upper left) ERA-Interim and (upper right) IFREMER–GPCP (1993-2007). The lower panels show the global monthly mean freshwater flux time series of each data record (top) and the zonal mean freshwater (bottom) for the overlapping time period 1988–2014 (1993-2007 for IFREMER-GPCP).

The negative difference in evaporation between HOAPS-4.0 and IFREMER and the positive bias in precipitation between HOAPS-4.0 and GPCP add and explain this feature. At mid- and high latitudes, the positive bias in the GPCP precipitation (Figure 9, middle) leads to an enhanced freshwater flux into the ocean of IFREMER–GPCP compared to HOAPS-4.0, i.e. positive biases in freshwater flux differences. The global mean bias equals to -0.37 mm/d, indicating a freshwater deficit in HOAPS-4.0 with respect to IFREMER-GPCP.

The basic structure of the zonal means from each data record is comparable, as depicted in the bottom panels of Figure 11. Nevertheless, especially in the tropical regions distinct differences up to 2.5 mm/d are evident. Over the tropical oceans of the southern hemisphere, positive values of the satellite-derived products indicate a net freshwater flux into the atmosphere, while ERA-Interim shows negative values. This enhanced loss of freshwater from the atmosphere into the ocean in ERA-Interim leads to global monthly mean values that are generally about 0.1 mm/d lower compared to HOAPS-4.0 and even 0.46 mm/d below those of IFREMER-GPCP. The different magnitudes in monthly mean freshwater fluxes are clearly resolved in the time series (Figure 11, bottom, top). The mean globally averaged HOAPS-4.0 net ocean surface freshwater flux into the atmosphere for 1988–2014 is 0.50 mm/d (IFREMER–GPCP: 0.86 mm/d; ERA-Interim: 0.40 mm/d). The time series of both satellite-based data records exhibit very similar variability with a standard deviation of 0.20 mm/d.

6.6.2 Discussion

The mean oceanic freshwater flux in HOAPS-4.0 for the 1988–2014 period is 0.50 mm/day, which is equivalent to a liquid water volume of about 62,000 km^3/a. For a closure of the global freshwater balance, this transport of freshwater from the ocean into the atmosphere should be compensated for the most part by continental runoff. Long-term mean runoff data published and summarized by the Global Runoff Data Center (GRDC) add up to a mean value of approximately 0.34 mm/d (equivalent 42,000 km^3/a (Wilkinson et al., 2014). The uncertainties of different runoff estimates are still in the order of 10%–20% (Andersson et al., 2011).

Table 6-5: Requirements for freshwater flux product CM-12821 as given in the PRD [RD 1]. Accuracy numbers are given for global mean values. Regional larger deviations may occur.

	Threshold	Target	Optimal
Bias	1.3 mm/d	0.36 mm/d	0.09 mm/d
RMSD	1.6 mm/d	0.62 mm/d	0.25 mm/d
Decadal stability	0.35 mm/d	0.14 mm/d	0.005 mm/d

Additionally, other runoff sources, such as annual ice melt and groundwater flow into the ocean are estimated to be less than 10% of the river discharge (Burnett et al. 2001). Comparing these values to the HOAPS-4.0 global ocean freshwater flux leaves an imbalance of about 0.15 mm/d in the global freshwater balance. Following these approximations, the imbalance for the combined IFREMER/GPCP fields is in the order of 0.50 mm/d. For ERA-Interim, an imbalance of about 0.05 mm/d remains.

Because of the lack of reliable in situ measurements, the validation of HOAPS-4.0 monthly mean freshwater flux is done against a combination of other satellite based climatologies and reanalysis data. Comparing HOAPS-4.0 freshwater flux against the IFREMER-GPCP reference, it can be concluded that the threshold requirements defined in the product requirement document are met (Table 6-5, PRD [RD 1]). More precisely, the bias (RMSD) equals to -0.37 mm/d (0.44 mm/d).

6.7 Total column water vapour

6.7.1 Results

6.7.1.1 Comparison to ERA-Interim, TMI and other SSM/I and SSMIS data

HOAPS-4.0 climatological mean TCWV for the years 1988–2014 is shown in the left panel of Figure 12. The ITCZ and the warm pool regions are characterized by maximum climate mean values exceeding 50 kg/m^2. In general TCWV gradually decreases from the tropics to the poles and minimum TCWV values are present at northern and southern most regions. Secondary local minima exist over stratus regions off the continental west coasts. The zonal mean annual cycle (Figure 12, right panel) highlights the movement of the ITCZ and its split, while only weak variability occurs close to the poles.
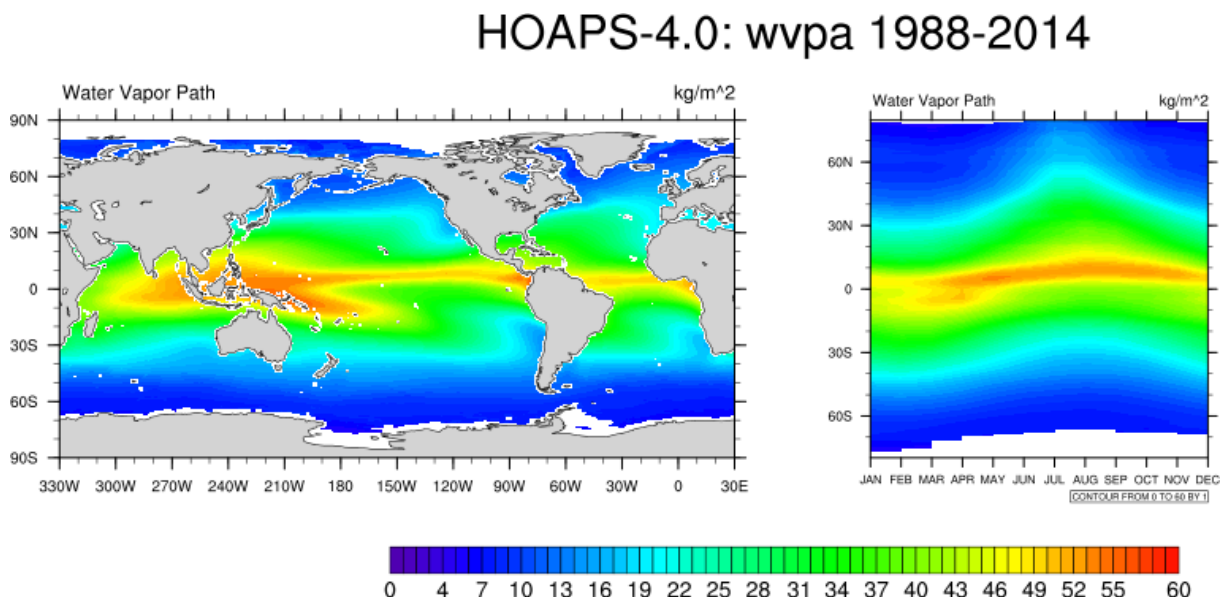
Figure 12: Climatological mean field (left) and zonal mean annual cycle (right) of HOAPS-4.0 TCWV for the years 1988–2014.

The comparison of HOAPS-4.0 TCWV with ERA-Interim, REMSS-SSM/I and REMSS-TMI is depicted Figure 13. Red colours indicate regions where HOAPS-4.0 exhibits on average higher values, while in blue shaded regions HOAPS-4.0 underestimates TCWV.
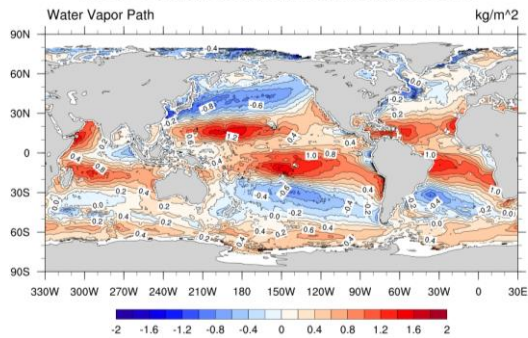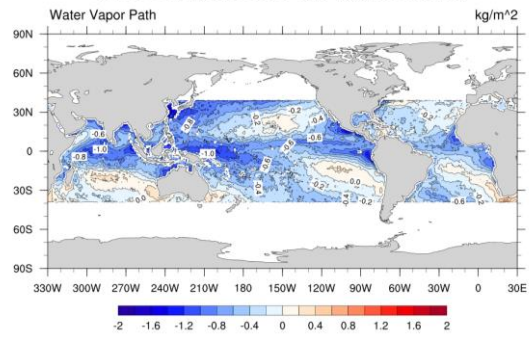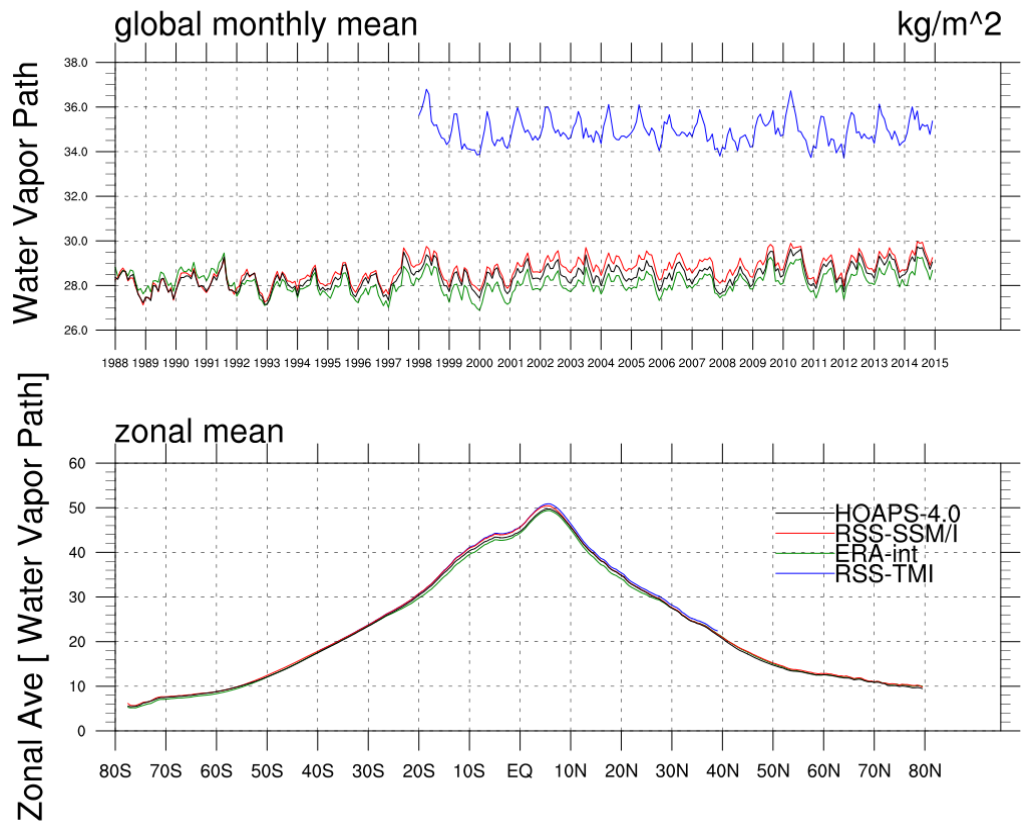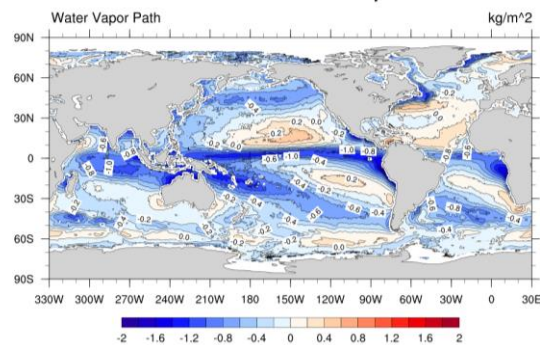
Figure 13: Difference of the 1988–2014 climate mean HOAPS-4.0 TCWV and (upper left) ERA-Interim, (upper right) REMSS-SSM/I, and (lower right) REMSS-TMI. The lower left panel shows the global monthly mean TCWV time series of each data record (TMI based on smaller spatial region) and the zonal mean TCWV for the overlapping time period 1988–2014 (1998-2014 for TMI).

ERA-Interim TCWV is generally lower compared to HOAPS-4.0 over the global oceans. This finds expression in a global mean bias of 0.3 kg/m^2. The largest differences occur over subtropical dry regions, where ERA-Interim is considerably lower compared to HOAPS-4.0. Here, the positive bias exceeds 2.5 kg/m^2. Smaller but still positive differences are present around Antarctica. Largest negative differences occur over the storm track regions. The global mean RMSD is 1.1 kg/m^2.

Comparing HOAPS-4.0 with REMSS data records (Figure 13, top right and middle) reveals a different pattern, although a few distinct features are similar to Figure 13 (top left). The global mean biases are -0.35 kg/m^2 (SSM/I) and -0.39 kg/m^2 (TMI). The most striking feature is the negative difference in the ITCZ. Here, underestimations by HOAPS-4.0 exceed values of -1.75 kg/m^2 (REMSS-SSM/I). In general, the range of differences is smaller relative to REMSS-TMI than to REMSS-SSM/I. Positive but small differences occur in subtropical dry regions, though smaller in area and slightly shifted than in the comparison to ERA-Interim. As against ERA-Interim the difference to both REMSS data records exhibits regional minima over the storm track regions. The difference relative to REMSS-SSM/I is close to 0 kg/m^2 in the southern ocean. Relative to SSM/I and TMI the global mean RMSD is 0.8 kg/m^2 and 1.0 kg/m^2, respectively.

The time series and the zonal plot confirm that the bias is generally small (Figure 13, lower panels). The zonal plot further confirms that the maximum differences occur in the tropics and subtropics though still small in magnitude. Note that the mean TMI data is computed for data within ±40°N/S only.

6.7.1.2 Comparison to GPS RO data

Within a Federated Activity CM SAF cooperates with the ROM SAF on the evaluation of GPS RO and HOAPS-4.0 data. Among others, instantaneous specific humidity profiles from COSMIC and Metop have been provided by ROM SAF for comparison to HOAPS-4.0. In order to have large numbers of collocations HOAPS-4.0 and GPS RO data are compared as zonal averages with bin size of 5°. The GPS RO data has been quality controlled following recommendations by ROM SAF (personal communication on 21 October 2016). A conservative land/sea ice mask (Schröder et al., 2016, as in section 6.7.1.3) is applied while arithmetically averaging the satellite data during zonal binning. Data from COSMIC is used for the period January 2007 to December 2014 and data from Metop for the period January 2009 to December 2014.
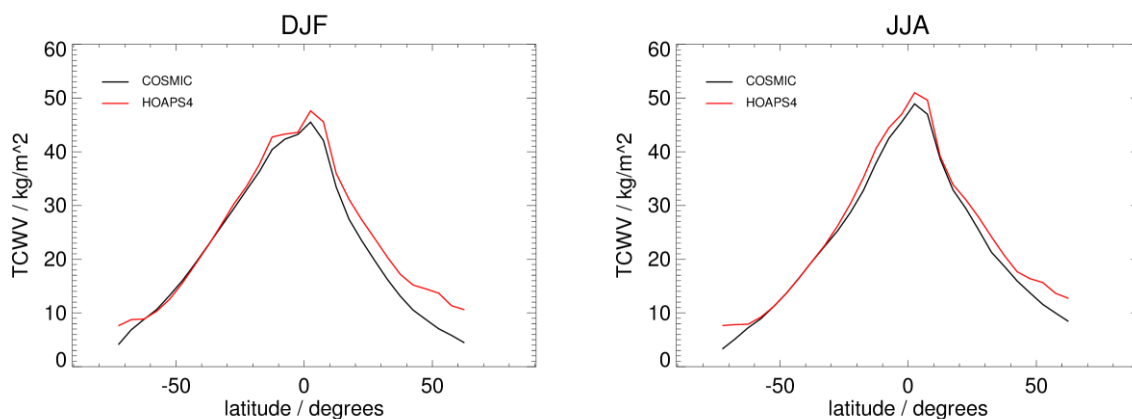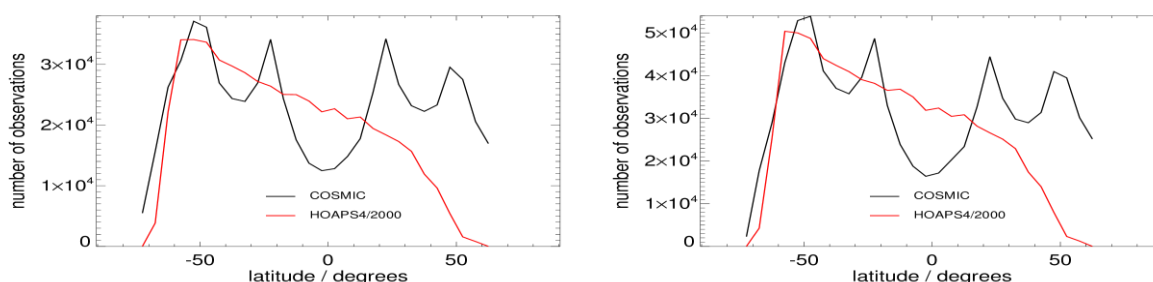
Figure 14: Zonal averages of TCWV from COSMIC and HOAPS-4.0 for winter (top left) and summer (top right) using data from the period January 2007 to December 2014. Corresponding number of observations are given in the bottom row. Note that the HOAPS-4.0 numbers have been scaled. The smallest number of observations in HOAPS-4.0 is >8000.

Figure 14 shows the zonal averages of TCWV from COSMIC and HOAPS-4.0 for summer and winter. The overall distributions are similar. However, zonal averages from HOAPS-4.0 are in the majority of zonal bins larger than those from COSMIC, with maximum differences in the extratropics of the northern hemisphere during boreal winter. Intercomparison results from the GEWEX water vapor assessment (G-VAP, see section 6.7.1.3) support the generally larger values found for HOAPS-4.0. The difference between HOAPS-4.0 and GPS RO might be caused by remaining uncertainties of GPS RO to measure humidity near the surface. However, open loop technology as implemented on Metop and COSMIC should minimize if not remove such problems (e.g., Ho et al., 2007, Lauritsen et al., 2011). Results from the comparison of HOAPS-4.0 and Metop are very similar to results presented in Figure 14 and are therefore not shown. Zonal distributions from HOAPS-4.0 are further analysed in section 6.7.1.3.

6.7.1.3   G-VAP

To date, a large variety of satellite based water vapour data records is available (see e.g. http://gewex-vap.org/?page_id=309 or http://ecv-inventory.com). Without proper background information and understanding of the limitations of available data records, these data may be incorrectly utilised or misinterpreted. The GEWEX Data and Assessments Panel (GDAP) has initiated G-VAP which has the major purpose to quantify the current state of the art in water vapour products being constructed for climate applications and to support the selection process of suitable water vapour products by GDAP for its production of globally consistent water and energy cycle products. This is achieved by inter-comparison and evaluation, and by providing reasons for differences and limitations where possible. More information on G-VAP is available at http://gewex-vap.org and results related to the analysis of long-term TCWV data records were published by Schröder et al. (2016). Among others, G-VAP assesses the quality of long-term data records (temporal coverage of more than 20 years, starting in the 1980s) and of data records from the full archive (temporal coverage of more than 10 years). The precursor version of HOAPS-4.0, namely HOAPS-3.2, was part of both analyses. The long-term analysis focused on global and regional stability and HOAPS-3.2 was used as reference to detect break points (Schröder et al., 2016). At present the final WCRP report on G-VAP is drafted and will be reviewed by GDAP at the end of 2016.
TCWV from HOAPS-4.0 has been included in the archive and the following analysis was carried out using G-VAP tools:
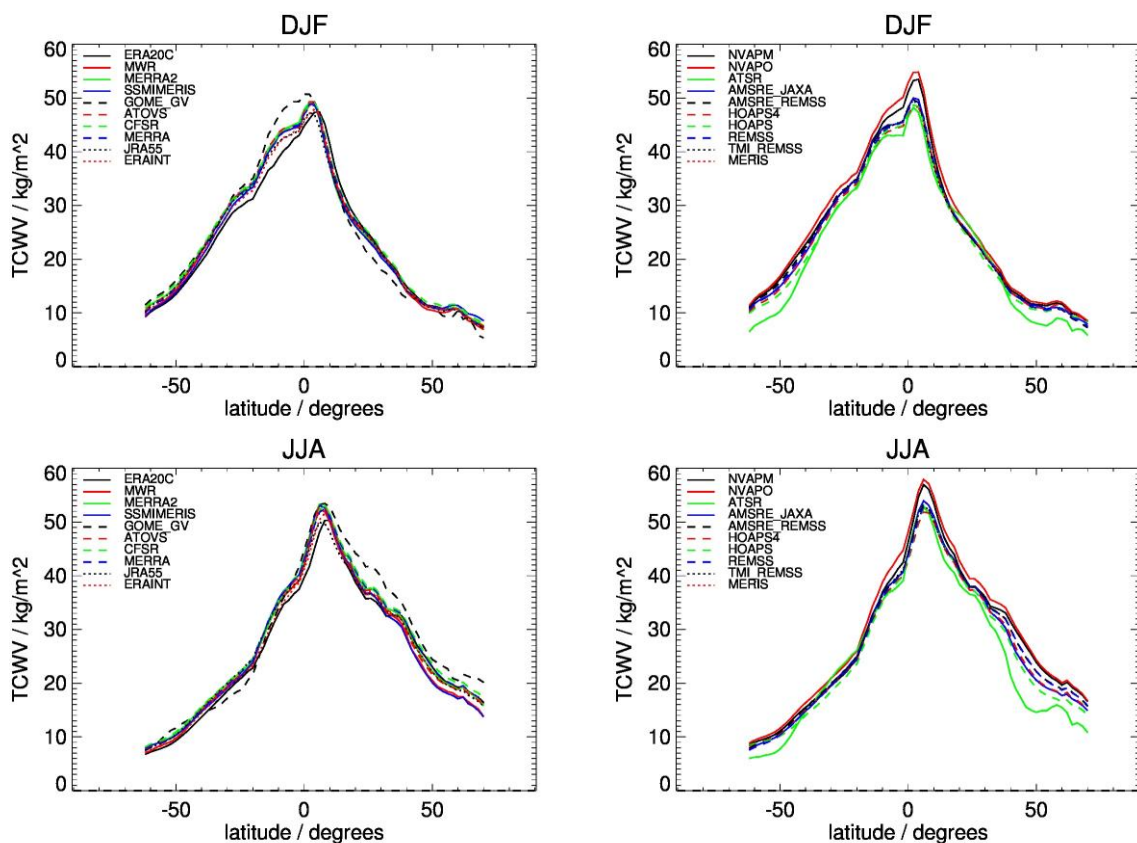
Figure 15: Zonal averages of TCWV for the period 1988-2008 using data of 20 records over global ice-free oceans. Top panels show results from summer and bottom panels show results from winter. The red dashed line in the right panels shows results from HOAPS-4.0.

- Intercomparison of zonal averages,
- Intercomparison of spatial trend maps using a subset of data records from the G-VAP long-term archive,
- Average trend and regression values considering the global ice-free ocean.

The methodologies are described in Schröder et al. (2016).

Figure 15 shows zonal averages of TCWV from 20 data records including HOAPS-4.0 and HOAPS-3.2. The data was pre-processed to cover the common period 1988-2008 on a regular longitude/latitude grid of 2° and as described in section 6.7.1 a conservative common land/sea-ice mask was applied during computation of the zonal averages. Zonal averages from HOAPS-4.0 are very similar to the majority of other microwave based TCWV data records. Relative to HOAPS-3.2 HOAPS-4.0 exhibits slightly larger values in the extratropics, which leads to values closer to the seeming ensemble median. This is in particular valid for the extratropics in the northern hemisphere.

In Figure 16 trend estimates for TCWV from HOAPS-4.0, HOAPS-3.2 and ERA-Interim are shown. While the patterns and trend values are very similar between HOAPS-3.2 and HOAPS-4.0, ERA-Interim exhibits significantly smaller trend values. The latter is caused by break points as described in Schröder et al. (2016). Over the global ice-free ocean the trend estimate from HOAPS-4.0 is 0.23 ± 0.09 kg/m^2/decade which is not significantly different from the HOAPS-3.2 trend estimate. From a theoretical point of view the expected change can be estimated. For HOAPS-4.0 a value of 7.2 ± 0.3 %/K was found what is practically identical to the corresponding HOAPS-3.2 value.
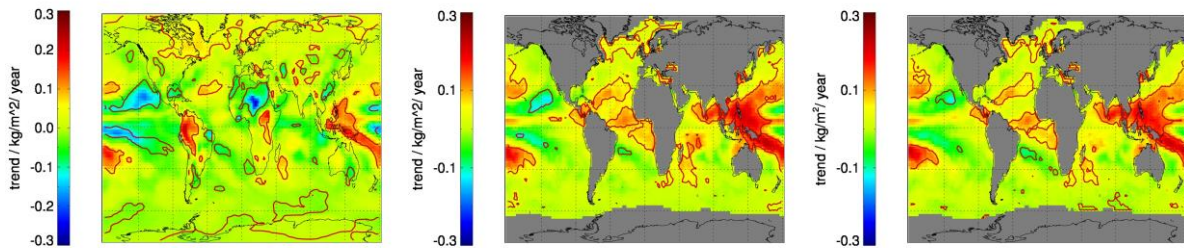
Figure 16: Trend estimates for TCWV and the period 1988-2008: ERA-Interim (left panel), HOAPS-3.2 (middle panel) and HOAPS-4.0 (right panel).

Fortunately CM SAF was funded to continue G-VAP until 2022, also with the goal to reassess the quality of updated data records. It was consensus at G-VAP workshops to continue G-VAP beyond the publication of the final report. Thus, the new HOAPS-4.0 data record will be fully included in the long-term analysis of G-VAP in the near future.

Table 6-6: Requirements for TCWV product CM-12701 as given in the PRD [RD 1]. Accuracy numbers are given for global mean values. Regional larger deviations may occur.

	Threshold	Target	Optimal
Bias	3.0 kg/m^2	1.4 kg/m^2	1.0 kg/m^2
RMSD	5.0 kg/m^2	2.0 kg/m^2	1.0 kg/m^2
Decadal stability	0.4 kg/m^2	0.2 kg/m^2	0.08 kg/m^2

6.7.2 Discussion

The small global mean bias relative to SSM/I, TMI and ERA-Interim and the reasonably small RMSD again confirm the high quality of the HOAPS TCWV data record. The precursor version, HOAPS-3.1, was also compared to those three data records (REMSS in version 6), see Schröder et al. (2013). They observed similar results, with bias values around 0.5 kg/m^2 (positive against ERA-Interim and negative against REMSS, in line with the updated results for HOAPS-4.0) and RMSD typically smaller than 1.5 kg/m^2.

It might be expected that the data records are comparable, in particular the HOAPS-4.0 and the REMSS-SSM/I: both rely on observations from the same sensor. Apart from that, all other steps, the intercalibration of the SSM/I, the retrieval, the auxiliary data and the gridding procedure are different. Thus also ERA-Interim and REMSS-TMI have been considered. However, ERA-Interim assimilates SSM/I data and TMI is also a microwave imager with very similar channel characteristics.

G-VAP found that the majority of the SSM/I based data records exhibit similar quality in terms of stability and homogeneity. In contrast, the majority of non-SSM/I data records exhibit relatively large break points which temporally coincide with changes in the observing system and which lead to trend estimates significantly different from the SSM/I based data records and not in line with theoretical expectation (e.g., Schröder et al., 2016). This is also valid for ERA-Interim. Note that the climate change analysis is not an application area of ERA-Interim and that trend estimation was used as a tool to identify issues in the data records within G-VAP.

The intercomparison of zonal averages also exhibits very good agreement among the majority of considered data records. Interestingly, the HOAPS-4.0 data record is more humid in the tropics than GPS RO data from ROM SAF while results based on REMSS-SSM/I and TMI are similar to HOAPS-4.0 in the tropics. As ERA-Interim is also slightly smaller than the microwave-based products, it might be the case that the underestimation of GPS RO in the tropics might originate from the utilised background, i.e. ERA-Interim.

Based on the comparisons of the HOAPS-4.0 monthly mean TCWV against ERA-Interim and the satellite-based REMSS SSM/I and TMI products, which show average (absolute) biases and RMSD of <0.4 kg/m^2 and 1.1 kg/m^2, respectively, it can be concluded that the optimal (bias) and target (RMSD) requirements defined in the product requirement document are met (PRD – RD 1, Table 6-6).

7 Decadal stability

In order to assess the decadal stability of the HOAPS-4.0 parameters the time-series of global monthly mean anomalies to a reference data record have been analysed. The references for each parameter are as defined below:

near surface wind speed $\Rightarrow$ NOCS
near surface specific humidity $\Rightarrow$ NOCS
evaporation/latent heat flux $\Rightarrow$ NOCS
TCWV $\Rightarrow$ SSM/I REMSS
precipitation $\Rightarrow$ GPCP
freshwater flux $\Rightarrow$ IFREMER – GPCP

Table 7-1: Results from the decadal stability analysis of global monthly mean anomalies (numbers are per decade). The values in brackets give the probability that the stability is smaller than the requirement (given here: target and optimal).

	Decadal stability		Decadal stability
Parameter	**Target**	**Optimal**	**HOAPS-4.0**
Near surface humidity CM-12901	0.10 g/kg (100%)	0.04 g/kg (99.9%)	0.02 ± 0.007 g/kg
Near surface wind speed CM-12911	0.12 m/s (98.3%)	0.03 m/s (0.0%)	-0.09 ± 0.012 m/s
Evaporation CM-12801	0.14 mm/d (100%)	0. 0043 mm/d (2.8%)	-0.02 ± 0.010 mm/d
Latent heat flux CM-12811	3.9 W/m^2 (100%)	0.12 W/m^2 (3.0%)	-0.64 ± 0.300 W/m^2
Precipitation CM-12611	0.02 mm/d (74.4%)	0.004 mm/d (10.8%)	0.01 ± 0.0090 mm/d
Freshwater flux CM-12821	0.14 mm/d (96.3%)	0.005 mm/d (0.0%)	-0.09 ± 0.028 mm/d
Vertically integrated water vapour CM-12701	0.20 kg/m² (100%)	0.08 kg/m² (100%)	0.00 ± 0.008 kg/m²
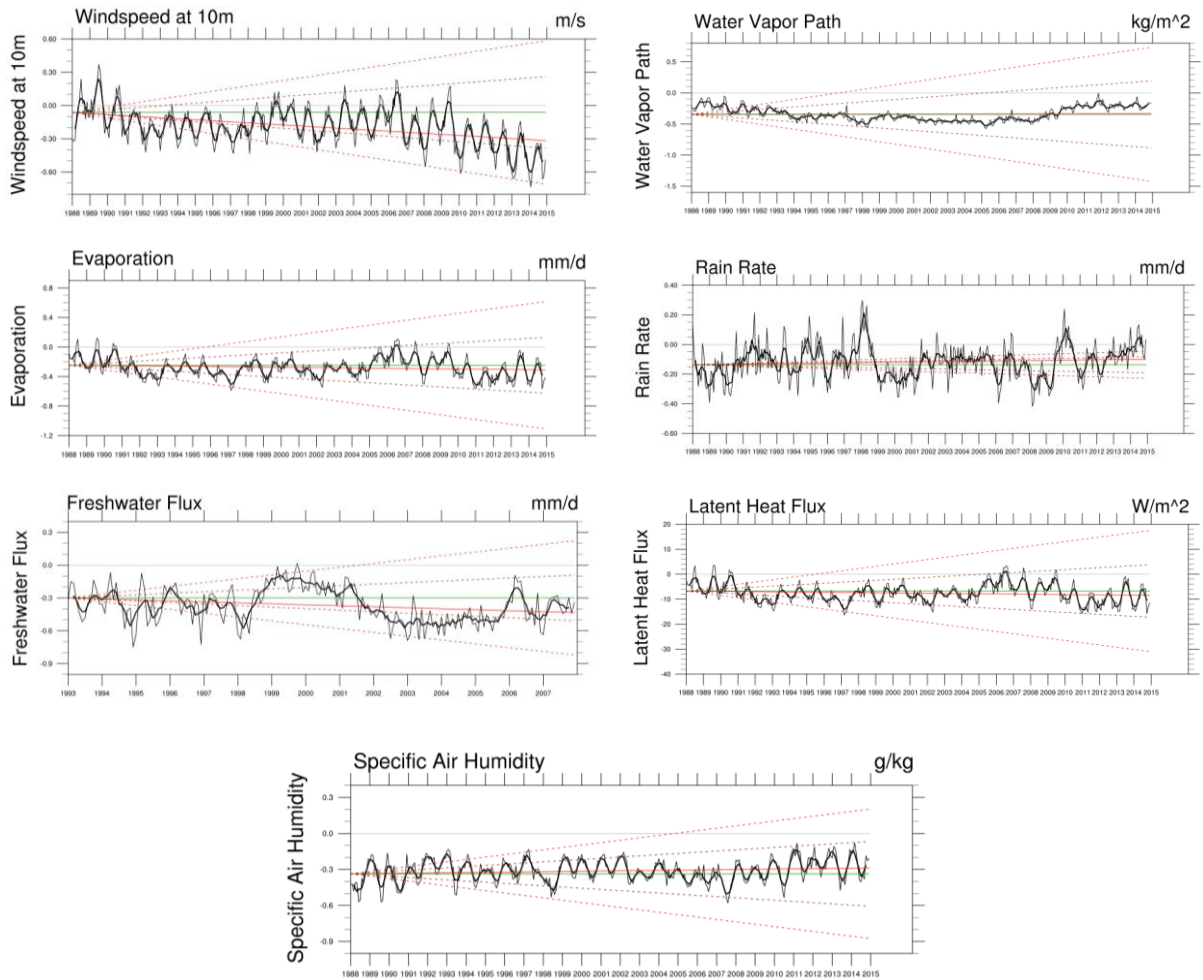
Figure 17: Time series of global monthly mean anomalies of HOAPS-4.0 parameters minus reference (thin black line) for the time period 1988-2014. The thick black lines are 5-monthly running means. The light red (dark red) dashed lines represent the threshold (target) requirements. The red line shows results from the linear regression and the green line is 0 line.

Figure 17 shows the anomalies for the overlapping time period from January 1988 to December 2014. Table 7-1 summarizes the results of the linear regression analysis and includes the probability that the stability is smaller than the target and optimum requirement (see section 6.1 for details). For all parameters the stability is significantly smaller than the target requirements, except for precipitation. The stability is even significantly smaller than the optimum requirement for near surface specific humidity and TCWV. The probability with respect to the threshold requirement is 100%, except for precipitation where it is 99.8% (not shown).

The stability was estimated by applying linear regression without taking into account uncertainties associated with the input to the regression and autocorrelation and without masking outliers. The consideration of autocorrelation would increase the uncertainty of the stability estimation and with that decrease the coverage probability. In contrast, the removal of outliers would decrease the uncertainty but would also impact the stability estimate. The consideration of uncertainties during regression would impact stability estimation and its uncertainty.

Note that the stability has been assessed on a global ice-free ocean scale but that on regional scale the actual stability might be different from the global ice-free ocean estimate. Further

note that the stability of the vertically integrated water vapour relative to the data from REMSS is likely a lower bound on stability because both the HOAPS-4 and REMSS data are based on SSM/I and SSMIS observations.

NOCS data quality is affected by the availability of in-situ data. In order to assess the impact of regions with low data density in NOCS the stability has also been estimated using data from the northern hemisphere only. In case of near surface wind speed, the absolute stability decreases and the coverage probability increases. In contrast, the absolute stability increases while the coverage probability decrease for near surface specific humidity.

Although the stability of near surface specific humidity and freshwater flux are statistically significant all stabilities are within the target or even optimal requirement.

8 Evaluation of HOAPS-4.0 parameters using instantaneous data

8.1 Methodology

To quantify both systematic and random uncertainties in HOAPS-4.0 latent heat flux (LHF) related parameters, instantaneous SSM/I data records of near-surface specific humidity (q_a), wind speed (U), and sea surface saturation specific humidity (q_s) are collocated to buoy and ship based in-situ measurements (compare section 4.8). Temporal and spatial collocation thresholds of 60 min and 50 km are chosen, respectively, following results of an autocorrelation function analysis presented in Kinzel (2013).

We follow a common approach and iteratively transfer in-situ wind speeds to 10 m above sea level after Fairall et al. (2003), under the assumption of an equivalent neutral stability background profile. Flow distortions due to the measurement platforms, which introduce biases to the wind speeds (e.g. Weller et al., 2008), have not been further assessed. The vast majority of collocated match-ups include buoy measurements, which are likely to be less affected by this source of error compared to ship records. However, e.g. Large et al. (1995) and Thomas et al. (2005) conclude that buoy winds tend to be lower than respective NWP and ship data, specifically during strong wind conditions. This behavior may contribute to the linear increase in dU shown in Figure 21b (Section 7.3).

By contrast, sensor height corrections involving q_a records are 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. A discussion regarding this approach is published in Kinzel et al. (2016). Generally, adjustments to marine humidity observations are challenging, as the required metadata are not always available (Roberts et al., 2012). Regarding ship data, Kent et al. (2007) pick up the potential bias source of sensor height corrections and demonstrate that NOCv2.0 sensor height corrections cause an increase in q_a in the order of 0.1-0.2 g/kg over the period 1971-2006. Due to continuously rising sensor heights during the last decades, an artificial drying trend will be detected, if a correction is not performed. Refraining from height corrections may thus partly explain the overall positive q_a bias (see section 7.2, Figure 20). However, the 1D-Var near-surface specific humidity in HOAPS-4.0 is defined for 2 m above sea level. As the vast majority of the applied in-situ q_a measurements are derived from buoys, whose sensors are located in similar heights, we conclude that our approach is justified. Furthermore, results in Kent et al. (2007) suggest that the impact of different measurement techniques on the bias is equally important and not trivial.

Furthermore, in-situ bulk SST measurements are converted to skin SST measurements following Donlon et al. (2002) for wind speeds above 2 m/s. Collocated match-ups during local day time and/or light winds conditions (< 2 m/s) are discarded from further analysis to prevent additional uncertainties associated with the diurnal warming layer. Sensor depth corrections are not performed, as the night-time temperature gradient within the surface layer during significant wind speeds is assumed to be negligible.

It is likely that biases in the HOAPS-4.0 LHF-related parameters do not solely depend on the respective parameter itself, but rather on the ambient atmospheric conditions. We argue that a combination of instantaneous HOAPS-4.0 q_a ('hair'), U ('wind'), sea surface temperature ('asst'), and water vapour path ('wvpa') are suitable for describing the prevalent atmospheric state. In consequence, the biases (HOAPS-4.0 minus in-situ), that is dq_a, dU, and dq_s as a function of the four state variables, can be separately investigated in one-dimensional space. Results of these evaluations are shown for dq_a, dU, and dLHF (derived via bulk equation) in form of scatter density plots for 2001-2004 (Figure 18, Figure 21, Figure 24). By contrast, their spatial bias distributions (Figure 19, Figure 22) as well as time series of bias and total RMSD (Figure 20, Figure 23, Figure 25) are presented independently of the state variables.

Due to the collocation procedure, the random uncertainty of each match-up does not only include the satellite retrieval error (E_{tot}), but also contributions of in-situ measurement noise (E_{ins}) and random collocation errors (E_C). To isolate E_{tot} of q_a, U, and q_s from these random uncertainties, an error decomposition of the random component via multiple triple collocation (MTC) analysis is carried out for q_a, U, and q_s. A bias correction with respect to the in-situ ground reference is a prerequisite for applying this MTC analysis. More details regarding this multi-dimensional approach (it is a function of the atmospheric state, i.e., depends on hair, wind, wvpa, and asst) is given in Kinzel et al. (2016), for example for q_a.

Both E_{ins} and E_C result from the MTC analyses (compare Table 8-1). Given their magnitudes, the monthly-mean RMSD estimates derived in section 7.2-7.4 are corrected (that is, reduced) in order to quantify the random uncertainty associated with the satellite retrieval (E_{tot}).

Associated results and discussions are given per parameter in the next sections.

Table 8-1: Average random uncertainty associated with the in-situ measurement (E_{ins}), the collocation procedure (E_C), and the satellite retrieval (E_{tot}), as derived from MTC analyses. Compare Kinzel et al. (2016) for details regarding the method.

	E_C	E_{ins}	E_{tot}
Hair	0.5 g/kg	0.5 g/kg	1.0 g/kg
Wind	1.2 m/s	1.9 m/s	1.3 g/kg

8.2   Near surface specific humidity

With the exception of a linear wind speed dependency, the average bias in q_a is small and not a function of hair, wvpa, and asst, as is indicated by the linear fits, respectively (Figure 18a, c, d). The characteristic pattern of dq_a(hair) (Figure 18a), that is the HOAPS-4.0 overestimation (underestimation) of q_a in midlatitudes (tropics/subtropics), has been extensively discussed in Kinzel et al. (2016), specifically regarding the maximum in random uncertainty for q_a

between 12-15 g/kg. To some degree, the bias distribution resembles that of the SST-dependency (Figure 18d). HOAPS-4.0 tends to overestimate q_a in low-wind regimes (that is (inner) tropics), whereas the bias becomes negative in strong wind regimes (Figure 18b), yet with a smaller random component.

Figure 19 shows the global distribution of dq_a illustrated in Figure 18a-d. Large areas within the subtropics and tropics exhibit negative biases, what is also reflected in Figure 18a for q_a in the range of 15-18 g/kg and Figure 18d (SST = 26-28°C). The distinct minimum southeast of the Arabian Peninsula is remarkable, as average wind speeds are very low and q_a is in the order of 15 g/kg, both of which would imply a positive bias according to Figure 18a-b. This underestimation of HOAPS-4.0 q_a may be related to retrieval issues owing to high aerosol loads. The effect of scattering on the retrieval schemes by aerosols has been assessed by Ge et al. (2008), for example. The underestimation of HOAPS-4.0 q_a in the extratropical storm track regions (Figure 19) can be traced back to negative dq_a for strong wind speeds, as is also indicated in Figure 19b. Positive biases occur more sporadically, although hotspots are seen along the (sub-) tropical western margins of the African and South American continent. These regions coincide with global wind speed minima (on a climatological scale) and SST in the vicinity of 20°C. Respective positive biases are well resolved in Figure 18b and Figure 18d.
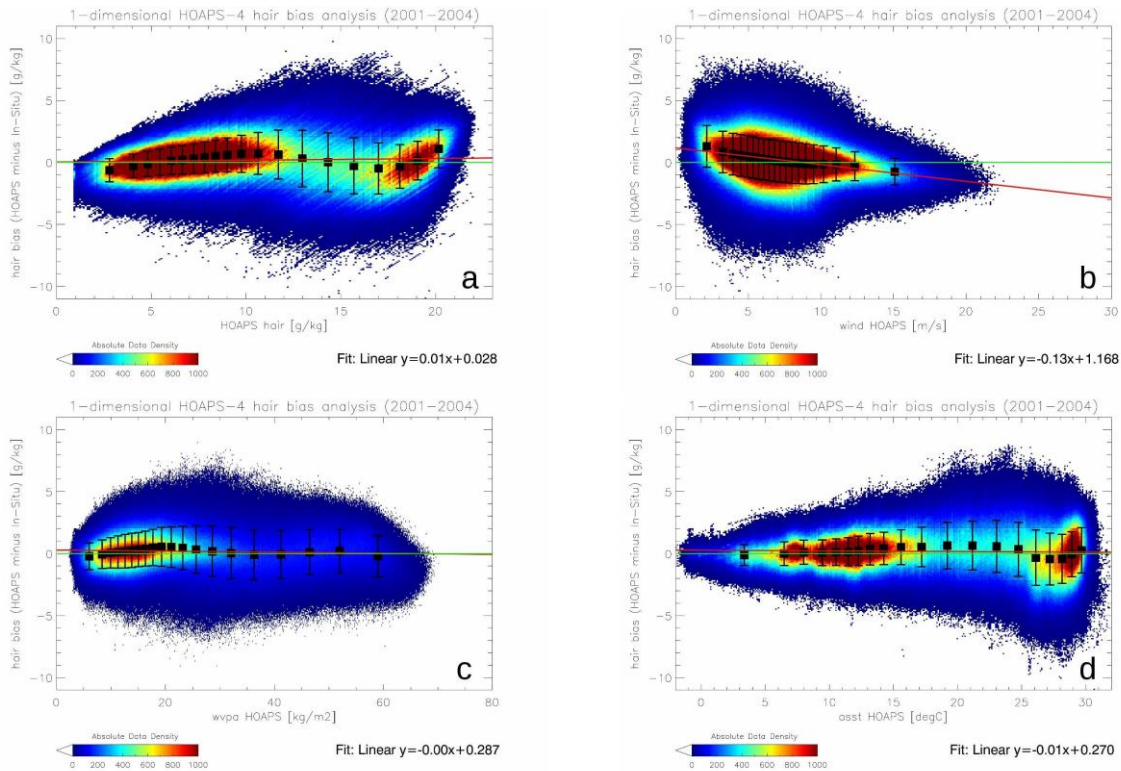
Figure 18: Scatter density plot of dq_a [g/kg] as a function of a) q_a ('hair', top left), b) U ('wind', top right), c) water vapour path ('wvpa', bottom left), and d) SST ('asst', bottom right). For the time period 2001-2004 the one-dimensional bias analyses are illustrated which are based on double collocations. Black (transparent) squares indicate significant (insignificant) bin biases (at the 95% level). Their standard deviations are given by the black bars. Each of the 20 bins includes 5% of the overall match-up data. Approximately 7.2 million match-ups contribute to these figures.

They are characterized by upwelling of cold waters, which are likely not captured correctly by the retrieval (e.g. Jackson et al., 2009). The frequent occurrence of stratus clouds may also

introduce uncertainties, as is proposed by Smith et al. (2011) for coastal regions off Peru. The latter in combination with the beam filling effect might explain the observed regional maxima in bias because the beam filling effect leads to larger observed brightness temperatures at 19 and 22 GHz (Bremen et al., 2002) and thus to potentially larger q_a.

The global average q_a bias is given by -0.1 g/kg, implying that HOAPS-4.0 slightly underestimates q_a. The weighted global average equals to 0.18 g/kg, where each bin is weighted by the amount of contributing match-ups. This slight positive bias is also represented in Figure 20 for 2001-2004.

Figure 20 shows the monthly time series of HOAPS-4.0 bias in q_a (thin black line) along with its bias-corrected RMSD (gray bars), for the time period of 1991-2008. The time series are restricted to the SSM/I time period up to 2008, as no DWD-ICOADS reference data are available after 2008. 5-month running means of bias and RMSD are shown as thick black and gray lines, respectively. The average bias tends to be slightly negative from 1991-1999, whereas it is slightly greater than zero during 2000-2005. The long-term average dq_a (RMSD) is given by 0.06 ± 0.18 g/kg (1.65 ± 0.14 g/kg). This bias is insignificant at the 95% level (t-test). The linear trend (red line) indicates an increase in bias over time, where the decadal trend is in the order of 0.15 ± 0.02 g/(kg decade). Performing the same stability analysis approach as in section 7 indicates that the decadal stability of q_a is significantly smaller than the threshold requirement, yet larger than the target requirement.
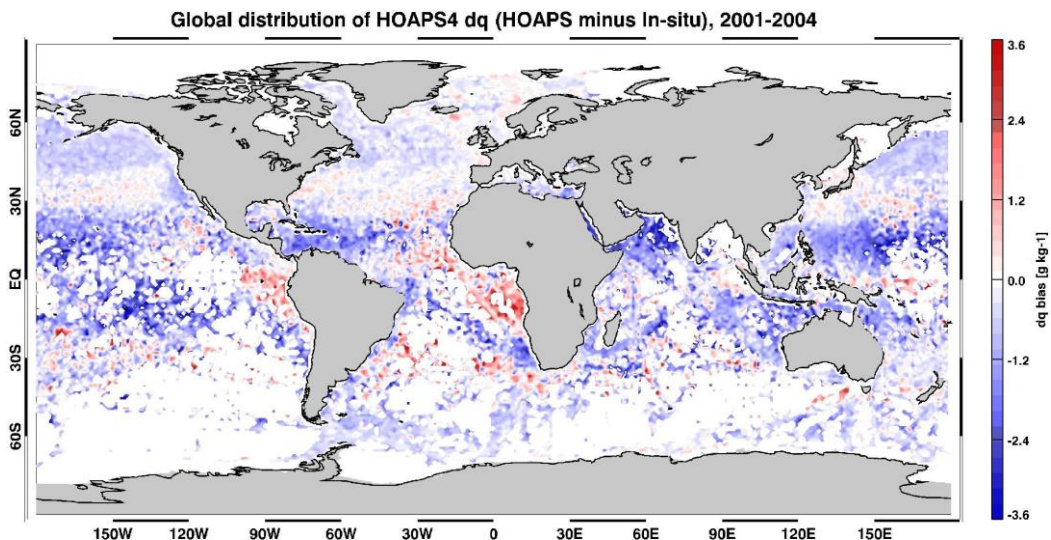
Figure 19: Bias map dq_a (HOAPS-4.0 minus in-situ [g/kg]), showing the distribution of the dq_a match-ups illustrated in Figure 18. Whereas red shading indicates an overestimation of HOAPS-4.0 q_a, blue shading corresponds to an underestimation in HOAPS-4.0 q_a.
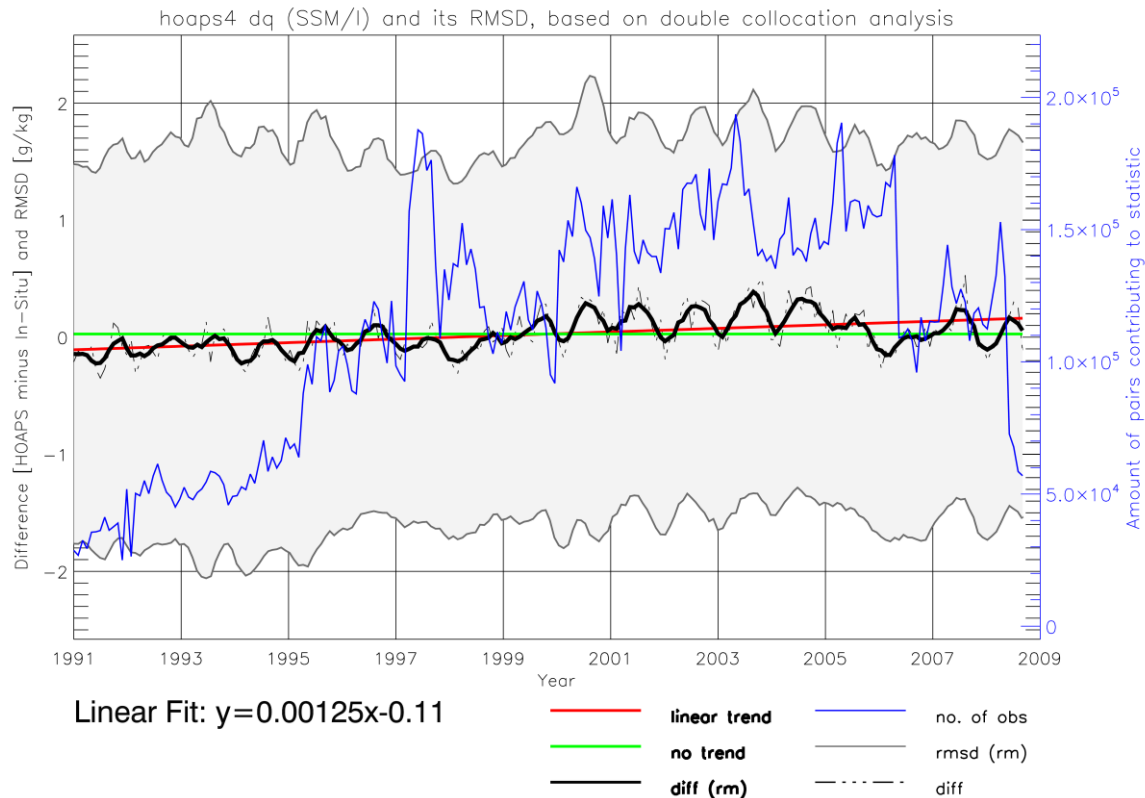
Figure 20: Time series of monthly dq_a [g/kg] (thin black line) (HOAPS-4.0 minus in-situ) and the corresponding bias-corrected RMSD (gray shading) for 1991-2008. The 5-month running mean bias is shown as the black bold line, respectively. The red line indicates the linear trend, whereas the green line represents the no-trend line. The blue graph illustrates the amount of contributing match-ups per month..

The 5-month running mean suggests an annual cycle of the dq_a, where bias minima tend to occur during boreal winter and maxima during boreal summer. Mean q_a resulting from the collocation analysis are larger during boreal summer (10-12 g/kg) than during boreal winter (8-9 g/kg), which coincides with shift towards larger dq_a during summer (Figure 18a). Strongest underestimations of HOAPS-4.0 q_a occur during boreal winter 1997/1998, which coincides with a strong El Niño signal.

The average RMSD is affected by contributions of random in-situ measurements errors (E_{ins}) as well as random uncertainty owing to the collocation approach (E_c). In this regard, Kinzel et al. (2016) showed that the average contribution of E_{tot} to the overall RMSD (that is, the bars in Figure 20) is about 23%. This suggests that the approximate average random retrieval uncertainty (E_{tot}, compare section 7.1) is in the order of 1.25 g/kg.

The blue line represents the amount of contributing match-ups per month. The correlation coefficient between monthly dq_a and number of match ups equals to 0.43, which is significant at the 95% level. The change to smaller bias values between 1991 and 2005 might be associated with the sampling of the diurnal cycle, owing to additional operations of the NOAA F10 and F15 platforms [RD 3, Figure 2]. The match-up data density maximizes during 1997, 2003, and 2005. In 1997, this can be explained with up to four simultaneously operating SSM/I instruments. Also, a mean increase in match-up density up to 2006 can be traced back to an increasing amount of available in-situ data.

8.3 Near surface wind speed

Regarding the bias in U, Figure 21 indicates that the dependencies on hair, wvpa, and asst on average result in only small biases and exhibit no significant dependence on these variables (Figure 21a, c, d). The larger hair, which often goes along with an increase in wvpa and asst, the smaller is the random uncertainty signal. The wind speed bias as a function of wind itself (Figure 21b) indicates a small overestimation of HOAPS-4.0 U in strong wind regimes, whereas regions subject to low- and medium range wind speeds exhibit negligible biases with respect to the in-situ ground reference. The increasing systematic uncertainty for higher wind speeds goes along with a continuously rising random error from 1.5 m/s (tropics) to 3.5 m/s (extratropical storm tracks).
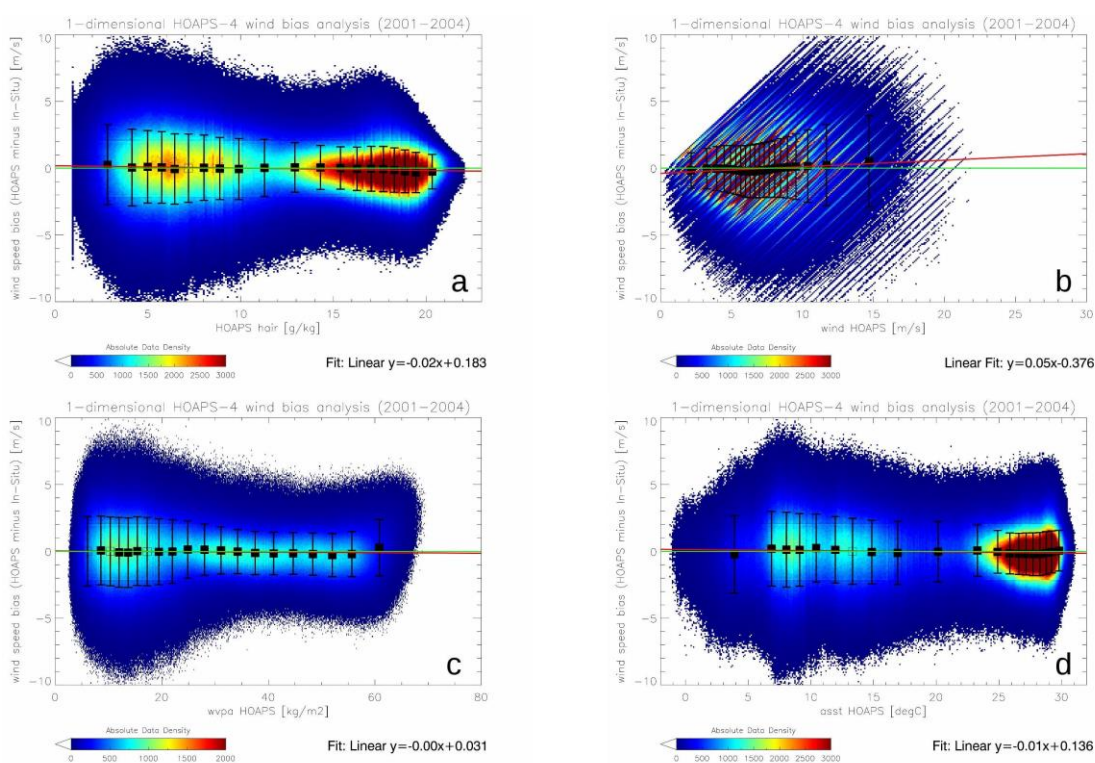
Figure 21: As in Figure 18, but for HOAPS-4.0 dU [m/s]. Around 17.9 million collocated pairs contribute to these figures.

Analogously to Figure 19 for dq_a, Figure 22 presents the global distribution of dU. In comparison to dq_a, the wind speed bias distribution resolves more variability (as is also illustrated by the random uncertainty magnitudes in Figure 21) and thus no clear signals are evident as to the mean sign of the bias. With the exception of the bias dependency on U itself, both Figure 21 and Figure 22 do not reveal any distinct systematic error dependencies. The global average U bias is given by 0.01 m/s. The corresponding weighted global average equals to -0.03 m/s, where each bin is weighted by the amount of contributing match-ups.

Figure 23 shows the time series of dU and associated RMSD. The overall bias (RMSD) is -0.05 ± 0.17 m/s (2.30 ± 0.34 m/s). This bias is insignificant at the 95% level. According to Kinzel et al. (2016), the contributing fraction of E_{tot} to the overall RMSD is on average as 50%. This implies that the RMSD associated with the satellite retrieval is in the order of 1.15 m/s.

	Validation Report **HOAPS version 4.0**	Doc.No.:SAF/CM/DWD/VAL/HOAPS
		Issue: 1.1
		Date: 31.01.2017

Prior to 1995, all monthly biases are negative, that is HOAPS-4.0 underestimates U. Conversely, most monthly biases become positive during the second half of the time series. This results in an overall positive trend in dU of 0.17 ± 0.02 m/(s decade). As for q_a in section 8.2, the decadal stability of U is significantly smaller than the threshold requirement, yet larger than the target requirement.

As for q_a, the seasonality in dU is evident, where bias maxima are frequently found during boreal winter. The correlation between dU and match-up density is 0.49, which is statistically significant at the 95% level. A possible reason for this strong relationship is analogous to that of Figure 20. The spread clearly maximizes during 2000. As none of the three contributing SSM/I instruments stand out regarding striking offsets (not shown), the reason for this strong variability is likely to originate in spurious in-situ data records. Similar analyses regarding HOAPS-3.2 have also revealed RMSD maxima in 2000, which supports this hypothesis.
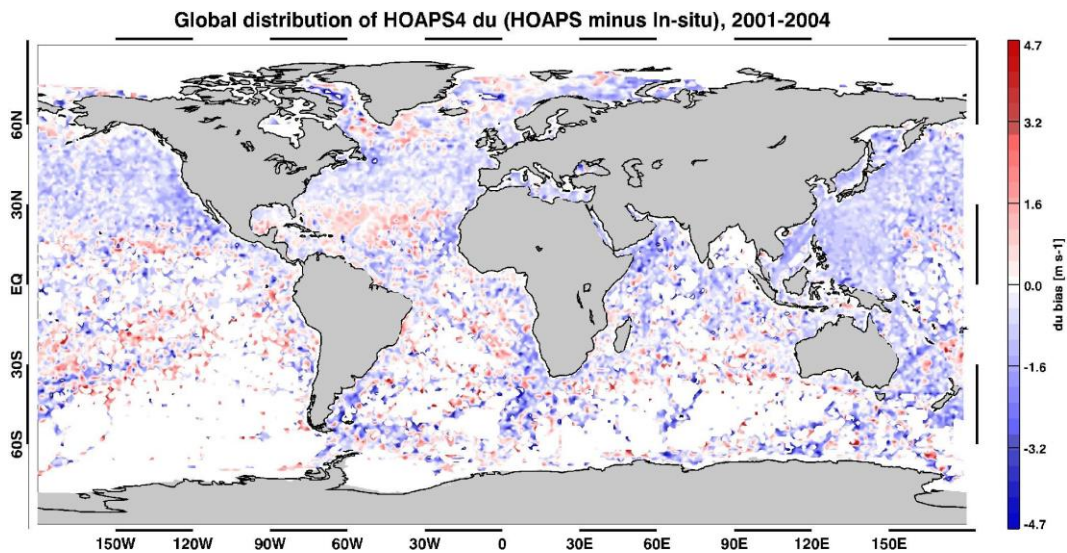
Figure 22: As in Figure 19, but for dU [m/s].
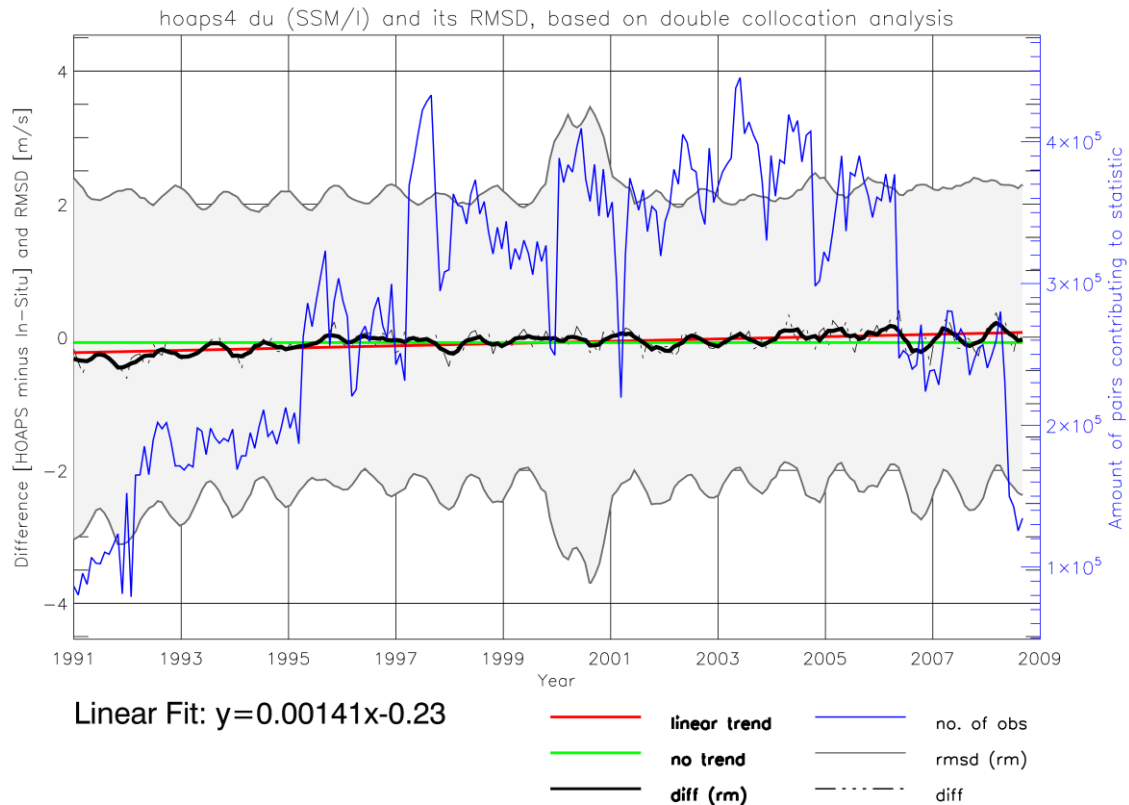
Figure 23: As in Figure 20, but for dU [m/s].

Next to a dependency on match-up data density, the positive trend in dU may generally be linked to an average increase in ship anemometer heights (compare Kent et al., 2007, Figure 3 therein). This height increase implies a correction of in-situ wind speeds to 10 m above sea level, which may introduce additional systematic uncertainty if the equivalent neutral stability assumption is violated. Also, anemometer heights are sometimes not known (see Kent et al., 2007, Figure 9 therein). In the framework of the present bias analysis, this necessitated sensor height estimations as a function of thermometer heights or ship lengths (Kent et al., 2007, Table 4 therein), which is also subject to uncertainty. Both hypotheses will only have a marginal impact on the instantaneous bias analysis, as a vast majority of match-ups includes buoy and not ship records but might impact stability considerations.

8.4   Latent heat flux and evaporation

Analogously to Figure 18 and Figure 21, Figure 24 shows the resulting one-dimensional bias analyses regarding HOAPS-4.0 LHF ('late'). In comparison to q_a and U, the amount of match-ups is 75% and 90% lower, respectively, as the required input parameters to the iterative flux algorithm were often not available concurrently. For 2001-2004, Figure 24a indicates that HOAPS-4.0 is underestimating LHF by 10-20 W/m^2 ($\triangleq$ 0.35-0.7 mm/d evaporation) for mid-latitudinal and inner-tropical q_a regimes, whereas it overestimates in-situ records by similar magnitudes for very low as well as subtropical q_a.
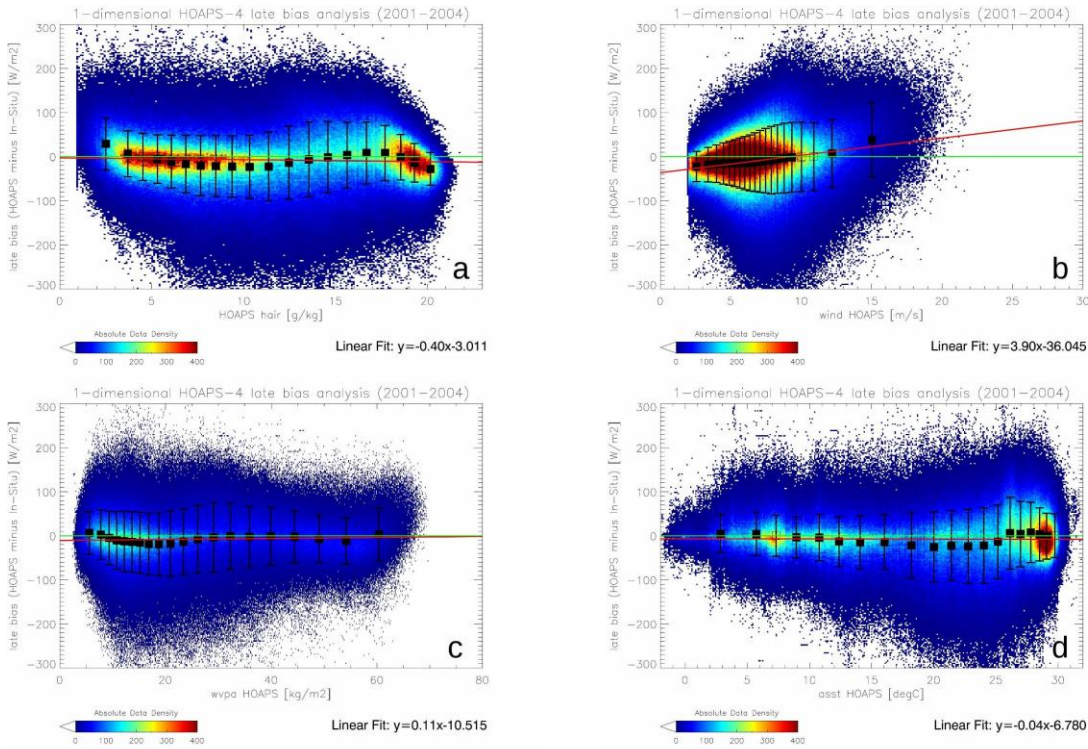
Figure 24: As in Figure 18, but for HOAPS-4.0 LHF [W/m^2]. Around 1.8 million collocated pairs contribute to these figures.

The bias distribution opposes that shown in Figure 18a, as higher q_a imply lower LHF, given constant U and q_s. Similar conclusions regarding bias distributions can be drawn when comparing Figure 24c-d with Figure 18c-d. Figure 24b indicates HOAPS-4.0 LHF underestimations in low and medium wind regimes (on average by 20 W/m^2) ($\triangleq$ 0.7 mm/d evaporation), whereas regions of strong winds show a LHF overestimation of up to 10-30 W/m^2 ($\triangleq$ 0.35-1.05 mm/d evaporation). This is in line with Figure 19b, as U is proportional to LHF, given constant q_a and q_s. Apart from wind speed, a linear dependency of the LHF bias on the dependent variables is not observed.

Figure 25 shows the monthly bias time series and its RMSD of HOAPS-4.0 LHF for 1991-2008. The overall bias average is -5.9 ± 8.0 W/m^2 ($\triangleq$ -0.21 ± 0.28 mm/d evaporation), indicating that HOAPS-4.0 slightly underestimates LHF (evaporation) in comparison to the in-situ ground reference. This is a direct consequence of the slight overestimation in HOAPS-4.0 q_a (Figure 20) and the concurrent underestimation in U (Figure 23), which in combination results in smaller LHF. This bias is insignificant at the 95% level. It shows a clear seasonality, where maxima, in an absolute sense, tend to occur during boreal winter months. This is in line with largest concurrent dU and smallest dq_a. There seems to be a transition from negative to rather positive biases between 2005 and 2006, which is also resolved in Figure 22 in form of a decrease in dq_a. Two striking maxima in dLHF between 15-30 W/m^2 ($\triangleq$ 0.53-1.05 mm/d evaporation) occur during boreal summers of 1991 and 2006, which coincide with the overall largest HOAPS-4.0 underestimation in q_a (1991) and strongest HOAPS-4.0 overestimation of U (2006), respectively.
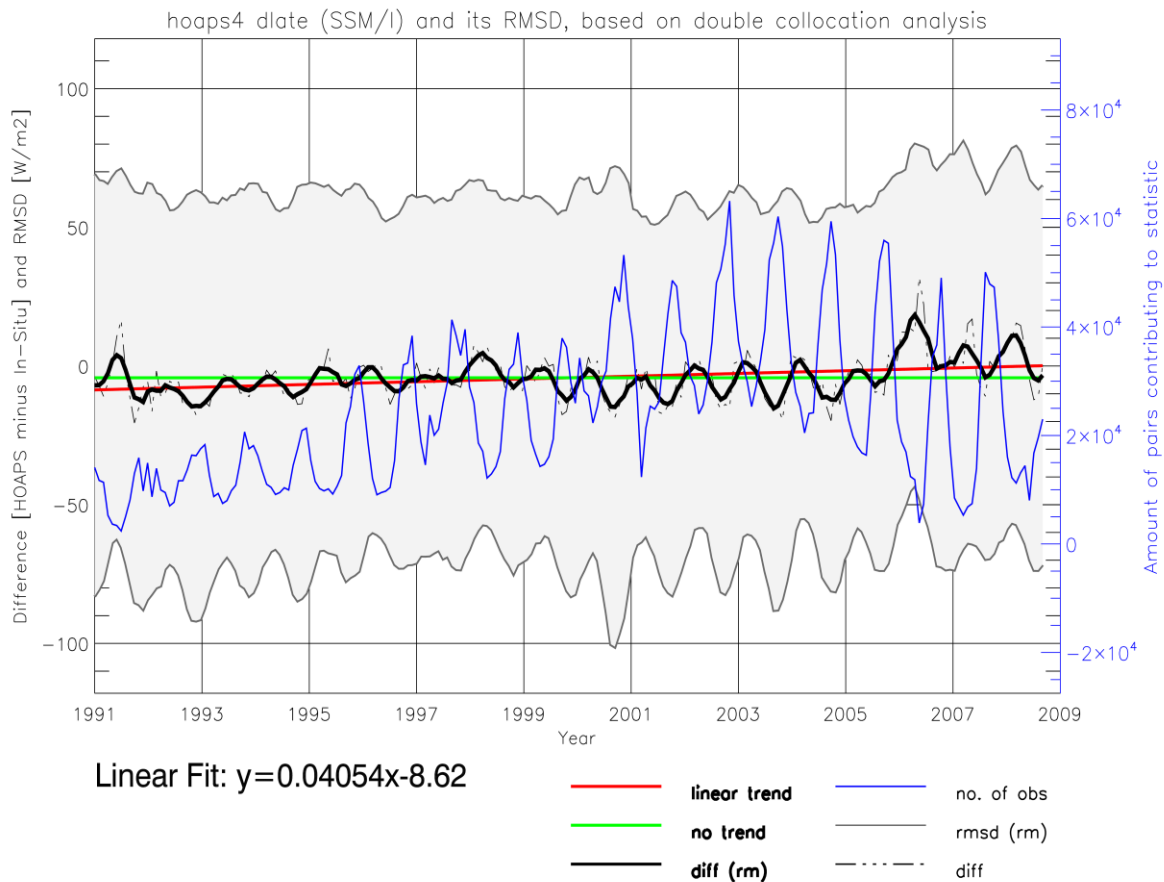
Figure 25: As in Figure 20, but for dLHF [W/m^2].

The decadal trend in dLHF is positive and takes on a value of 4.9 ± 1.0 W/(m^2 decade) (≙ 0.17 ± 0.04 mm/(day decade) evaporation). As for q_a and U in the preceding subsections, the decadal stabilities of LHF and evaporation are significantly smaller than respective threshold requirements, yet larger than the target requirements.

The average RMSD of 68.2 ± 8.0 W/m^2 (≙ 2.38 ± 0.28 mm/d evaporation) corresponds to 72% of the average match-up LHF (evaporation) of 93.5 W/m^2 (≙ 3.26 mm/d evaporation). To quantify the retrieval-related RMSD, we derive a mean bulk-equation based LHF, based on the match-up satellite averages of U, q_s, and q_a. The average density of moist air, ρ_a, is estimated by assuming a sea level pressure of 1013.25 hPa and a near-surface air temperature 1.25°C below the average match-up SST (compare Bentamy et al., 2003). Based on this mean SST, the latent heat of vaporization (lhv) is calculated. We furthermore estimate a mean transfer coefficient C_E, following the approach shown in Bentamy et al. (2003). For simplicity, we assume that variabilities in lhv, C_E, and ρ_a are negligible regarding the total satellite-related RMSD in LHF. Assuming independent variables, we perform error propagation analysis and yield a retrieval-related RMSD estimate of HOAPS-4.0 LHF of 28 W/m^2 (≙ 0.98 mm/d evaporation).

It should be mentioned that large monthly biases need to be treated with care, as the amount of contributing match-ups to the statistic varies considerably and is specifically low during these months (only several 1000 match-ups, see blue line). The correlation between dLHF and

match-up density is -0.36 and is statistically significant at the 95% level. Thus, for sparse collocation density, the illustrated biases and possible trends need to be treated with caution.

8.5 Concluding Remarks

In summary, the results of the instantaneous validation analyses suggest that all HOAPS-4.0 evaporation-related parameters meet the requirements that have been defined in the Product Requirements Document [RD 1].

Regarding q_a, the bias (0.05 ± 0.18 g/kg, insignificant at 95% level) meets the optimal -, retrieval-related RMSD (1.25 g/kg) and stability (0.15 ± 0.02 g/(kg decade)) the threshold criterion, as derived from Figure 20. This is approximately in line with results of identical analyses based on HOAPS-3.2. To be specific, a reduction regarding the decadal stability from 0.19 g/(kg decade) (HOAPS-3.2) to 0.15 g/(kg decade) (HOAPS-4.0, significantly smaller than the threshold requirement) has been achieved.

The bias (0.06 ± 0.18 m/s, insignificant at 95% level) for U meets the optimal -, whereas the retrieval-related RMSD (1.15 m/s) and the decadal stability (0.17 ± 0.02 m/(s decade), significantly smaller than the threshold requirement) satisfy the threshold requirement. In comparison to HOAPS-3.2, the trend has been reduced by 0.08 g/(kg decade) and an absolute bias reduction of 79% is registered.

HOAPS-4.0 LHF meets the target requirement regarding bias (-5.9 ± 8.0 W/m^2, insignificant at 95% level) and the threshold criteria regarding retrieval-related RMSD (28 W/m^2) and temporal stability (4.9 W/(m^2 decade), significantly smaller than the threshold requirement). Magnitudes of both bias and RMSD resemble that of HOAPS-3.2. Regarding the decadal trend, a reduction by 3.8 W/(m^2 decade) has been achieved in HOAPS-4.0, compared to the former version. Similar conclusions are drawn for the evaporation parameter.

Note that the evaluation of the instantaneous data is undertaken by referring to requirements, which have been defined for the HOAPS-G product [RD 1] and that different reference data records were used. Also the decadal trends presented here (Figure 20, Figure 23, Figure 25) rely only on the available match-ups and are therefore representative for the regions covered by the buoys and ships. Respective biases need to be considered with care, as they seem to depend on the collocated data density. This is likely the cause for discrepancies in decadal stabilities when comparing section 7 (often meeting target or optimal requirements) to section 8 (meeting threshold requirements). We therefore recommend treating the pixel-level stability analyses with caution.

Results on pixel-level resolution have been transferred to the L3 products, in as much as both systematic and random LHF-related uncertainty estimates are available within the gridded HOAPS-4.0 climatologies (not shown in section 6). Future work aims at gaining more insights in these uncertainties in terms of magnitude, distribution and reliability. This will eventually also expand the evaluation shown in section 6 regarding the interpretation and classification of the observed deviations, e.g., to assess consistency using uncertainty estimates after Immler et al. (2010).

9 Conclusions

The HOAPS-4.0 data record has first been compared to the precursor version HOAPS-3.2 and a comprehensive comparison to readily available products was carried out. These products are derived from different data sources and represent model-based estimates from a reanalysis data record, in situ measurements from ships, and different satellite-based data records that include sensors not utilized in HOAPS-4.0.

Since a true reference for the validation of the HOAPS-4.0 products is not available, reference data records have been selected advisedly in requirements review, neither admitting that the references might not be globally bias free nor implying superior quality over HOAPS-4.0. For the evaporation and related parameters, these are the ERA-Interim reanalysis, the ship-measurement based NOCS V2 data record, and the satellite-data based IFREMER V2 flux data record. The HOAPS-4.0 precipitation product is compared to ERA-Interim and the two satellite-based products, GPCP V2.2 and TRMM 3B43, version 7. HOAPS-4.0 TCWV products are compared to the TMI-based product from REMSS and to the combined microwave imager product from REMSS (both in version 7), ERA-Interim and COSMIC and Metop GPS RO data (beta-version from ROM SAF). A recently combined buoy and ship data record from DWD and ICOADS is used as reference for comparison of meteorological and flux parameters using instantaneous data. It is concluded that the resulting comparisons are an advance over previous studies in terms of temporal coverage, in terms of number of reference data records and in terms of approaches.

While the general patterns are reproduced by all data records and global mean time series often agree within a range of 10% of the individual products, locally larger deviations occur for all parameters. Specifically regarding near-surface humidity and rainfall estimates, the satellite derived data records agree better with HOAPS than with the reanalyses and/or in-situ data. . However, the compared satellite data records are not fully independent, as the satellite input data may be of the same origin and/or similar algorithms or parameterizations are used in retrieval procedures. This also accounts to some extent for ERA-Interim, which assimilates a wide range of satellite data.

The comparison of HOAPS-4.0 to buoy and ship observations, using instantaneous data, exhibits similar or improved quality relative to HOAPS-3.2. In particular, the quality of the wind retrieval could be improved. Noticeable is that this improvement is seen in regions where local maxima in bias relative to other references occur.

The results for evaporation show that the different estimates of evaporation and freshwater flux strongly depend on the individual input parameters. The most potential for improvement of the evaporation parameter appears to be in the humidity (q_a) retrievals. Particularly in the (sub-) tropical regions, the resulting evaporation difference patterns are still strongly determined by the deviations in the q_a fields. Apart from the comparison to IFREMER, deviations in wind speeds are of secondary importance, although their overall impact increases with latitude. The implementation of a 1D-Var retrieval for HOAPS-4.0 q_a in the future may be a step towards an improvement of the evaporation estimates with respect to reference data sets. Regarding wind speed, the comparisons to HOAPS-3.2 as well as on an instantaneous level show a considerably bias reduction. Generally, more detailed validation efforts are needed to specify the biases against independent in situ data. Further detailed regional analysis of all parameters required to derive the evaporation product is envisaged within the SEAFLUX Project of the World Climate Research Programme (WCRP) Global Energy and Water Experiment (GEWEX) Data Assessment Panel (GDAP).

In regions with high aerosol load or persistent cloudiness, deficiencies in the SST data records can cause biases in the q_s fields, affecting the sea–air humidity difference of all products. For example, a low bias in the SST of the eastern and central tropical Atlantic due to desert aerosols causes an underestimation of q_s and thus of the evaporation in HOAPS-4.0. Along the African west coast this effects coincides with an overestimation of q_a in HOAPS-4.0, which enhances the low bias in evaporation. Note that the cold bias in the SST data record, which according to Andersson et al. (2010b) caused the minimum in HOAPS-3.2 after the eruption of Mount Pinatubo, has considerably been reduced in HOAPS-4.0.

Table 9-1: Contingency of achieved accuracy with the corresponding requirements from RD 1. The stability is given per decade.

	Bias	RMSD	stability
q_a (g/kg)	-0.31 (target)	0.21 (optimal)	0.02 (optimal)
Wind (m/s)	-0.19 optimal	0.218 (optimal)	-0.09 (optimal)
E (mm/d)	-0.28 (target)	0.24 (optimal)	-0.02 (target)
LHF (W/m²)	-7.60 (target)	6.51 (optimal)	-0.64 (target)
P (mm/d)	-0.12 (optimal)	0.24 (optimal)	0.01 (target)
E-P (mm/d)	-0.37 (threshold)	0.44 (target)	-0.09 (target)
TCWV (kg/m²)	-0.34 (optimal)	0.30 (optimal)	0.00 (optimal)

Differences in precipitation from HOAPS-4.0 and HOAPS-3.2 are small because the same algorithm has been used. However, differences between HOAPS-4.0 and the references exhibit relatively large maxima on regional scale. Relative to GPCP largest positive (negative) biases are found at ITCZ (storm track regions). HOAPS-4.0 is known to perform better than other comparable satellite retrievals at mid and high latitudes (Klepp et al. 2010), where the mixed SSM/I and TOVS retrieval from GPCP exhibits a systematic high bias. Global and regional aspects of precipitation validation are carried out and are planned within the framework of the International Precipitation Working Group (IPWG).

The resulting freshwater flux estimates exhibit distinct differences in terms of global averages as well as regional biases. The observed reduction in HOAPS-4.0 E-P implies that the global freshwater imbalance of 0.67 mm/d observed for HOAPS-3.2 has considerably been reduced and now is 0.40 mm/d. For the combined IFREMER–GPCP fields the imbalance is slightly larger (0.5 mm/d). Because precipitation changed only marginally between HOAPS-4.0 and HOAPS-3.2 the observed change in the budget is caused by changes in evaporation.

Based on the comparisons presented in this report, we conclude that the HOAPS-4.0 data record is within target accuracies or better, as summarised in Table 9-1 and provides consistent fields of evaporation, precipitation, and the resulting freshwater flux as well as TCWV, near surface specific humidity and near surface wind speed that are well suited for further studies on climatological and regional scale. Overall, the imbalance in the global ocean surface freshwater flux is reduced compared to previous versions of HOAPS-4.0. However, more detailed validation efforts are needed to explain and, if adequate, remove the remaining biases between the different data records. Finally, it is emphasised that there is a

need for independent, high quality reference data enabling in the first place a profound validation of climate data records (CDRs) such as the HOAPS products.

10 References

Adler, R. F., C. Kidd, G. Petty, M. Morissey, and H. Goodman, 2001: Intercomparison of global precipitation products: The third Precipitation Intercomparison Project (PIP-3). Bull. Amer. Meteor. Soc., **82**, 1377–1396, doi: 10.1175/1520-0477(2001)082<1377:IOGPPT>2.3.CO;2.

——, G. J. Huffman, A. Chang, R. Ferraro, P.-P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, A. Gruber, J. Susskind, P. Arkin, and E. Nelkin, 2003: The Version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). J. Hydrometeor., **4**, 1147–1167, doi: 10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2.

Andersson, A., S. Bakan, K. Fennig, H. Graßl, C. Klepp, and J. Schulz, 2007: Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data—HOAPS-3—Monthly mean. Electronic Publication, World Data Center for Climate, doi: 10.1594/WDCC/HOAPS3_MONTHLY.

——, ——, and H. Graßl, 2010a: Satellite derived precipitation and freshwater flux variability and its dependence on the North Atlantic Oscillation. Tellus, **62**, 453–468, doi: 10.1111/j.1600-0870.2010.00458.x.

——, K. Fennig, C. Klepp, S. Bakan, H. Graßl, and J. Schulz, 2010b: The Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data—HOAPS-3. Earth Syst. Sci. Data, **2**, 215-234, doi: 10.5194/essd-2-215-2010, 2010.

——, K. Fennig, and M. Schröder, 2011: Validation report SSM/I products – HOAPS release 3.2. EUMETSAT CM SAF Validation Report, SAF/CM/DWD/VAL/HOAPS, issue 1.1, 25 March 2011.

Ashouri H., K. Hsu, S. Sorooshian, D. K. Braithwaite, K. R. Knapp, L. D. Cecil, B. R. Nelson, and O. P. Prat, 2015: PERSIANN - CDR: Daily Precipitation Climate Data Record from Multi-Satellite Observations for Hydrological and Climate Studies. Bull. Am. Meteorol. Soc., **96**, 69-83, doi: 10.1175/BAMS-D-13-00068.1.

Bentamy, A., K. B. Katsaros, A. M. Mestas-Nunez, 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.

——, L. H. Ayina, W. Drennan, K. Katsaros, A. M. Mestas-Nunez, and R. T. Pinker, 2008: 15 years of ocean surface momentum and heat fluxes from remotely sensed observations. FLUXNEWS, Vol. 5,WCRP Working Group on Surface Fluxes,World Climate Research Programme, Geneva, Switzerland, 14–16. [Available online at http://sail.msk.ru/newsletter/fluxnews_5_final.pdf.]

——, Grodsky S. A., K. B. Katsaros, A. M. Mestas-Nunez, B. Blanke, F. Desbiolles, 2013: Improvements of air-sea fluxes derived from satellite observations. Int. J. Remote Sens. **34** (14): 5243–5261, doi: 10.1080/01431161.2013.787502.

Béranger, K., B. Barnier, S. Gulev, and M. Crépon, 2006: Comparing 20 years of precipitation estimates from different sources over the world ocean. Ocean Dyn., **56**, 104–138, doi: 10.1007/s10236-006-0065-2.

Berry, D. I., and E. C. Kent, 2009: A new air - sea interaction gridded dataset from ICOADS with uncertainty estimates. Bulletin of the American Meteorological Society, **90**, 645 - 656. DOI: 10.1175/2008BAMS2639.1.

Berry, D. I., and E. C. Kent, 2011: Air-sea fluxes from ICOADS: the construction of a new gridded data record with uncertainty estimates. Int. J. Climatol. **31** (7): 987–1001. doi: 10.1002/joc.2059.

Boukabara, S.-A., K. Garrett, W. Chen, F. Iturbide-Sanchez, C. Grassotti, C. Kongoli, R. Chen, Q. Liu, B. Yan, F. Weng, R. Ferraro, T. Kleespies, and H. Meng, 2011: MiRS: An All-Weather 1D-Var Satellite Data Assimilation & Retrieval System, IEEE Trans. Geosci. Remote Sens., **49**, 3249-3272, doi:10.1109/TGRS.2011.2158438.

Bourras, D., L. Eymard, and W. T. Liu, 2002: A neural network to estimate the latent heat flux over oceans from satellite observations. Int. J. Remote Sens., **23**, 2405–2423, doi: 10.1080/01431160110070825.

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.

Bowman, K. P., C. R. Homeyer, and D. G. Stone, 2009: A comparison of oceanic precipitation estimates in the tropics and subtropics. J. Appl. Meteor. Climatol., **48**, 1335–1344, doi: 10.1175/2009JAMC2149.1.

Bremen, L. V., E. Ruprecht, and A. Macke, 2002: Errors in liquid water path retrieval arising from cloud inhomogeneities: The Beam-Filling Effect. Meteorol. Z. **11** (1), 13-19, doi: 10.1127/0941-2948/2002/0011-0013.

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

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

Burnett, W. C., M. Taniguchi, and J. Oberdorfer, 2001: Measurement and significance of the direct discharge of groundwater into the coastal zone. J. Sea Res., **46**, 109–116, doi: 10.1016/S1385-1101(01)00075-2.

Chou, S. H., E. Nelkin, J. Ardizzone, R. M. Atlas, and C. L. Shie, 2003: Surface turbulent heat and momentum fluxes over global oceans based on the Goddard satellite retrievals, version 2 (GSSTF2). J. Climate, **16**, 3256–3273, doi: 10.1175/1520-0442(2003)016<3256:STHAMF>2.0.CO;2.

——,——,——, and——,2004: A comparison of latent heat fluxes over global oceans for four flux products. J. Climate, **17**, 3973– 3989. 10.1175/1520-0442(2004)017<3973:ACOLHF>2.0.CO;2.

Clayson, C. A., R. J. Brent, and A. S. Bogdanoff, 2012: The SeaFlux turbulent flux data record version 1.0 documentation. SEAFLUX, pp. 1-5. [Available online at: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20120004203.pdf.]

Dee, D. P., and 35 co-authors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Quart. J. R. Meteorol. Soc., **137**, 553-597, doi: 10.1002/qj.828.

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

Fairall, C. W., 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, and M. Schröder, 2013: Fundamental Climate Data Record of SSM/I Brightness Temperatures. Satellite Application Facility on Climate Monitoring, doi: 10.5676/EUM_SAF_CM/FCDR_SSMI/V001.

Ge J., J. Huang, F. Weng, and W. Sun, 2008: Effects of dust storms on microwave radiation based on satellite observation and model simulation over the Taklamakan Desert. Atmos. Chem. Phys., **8**, 4903-4909, doi:10.5194/acp-8-4903-2008.

Hilburn, K. A., and F. J. Wentz, 2008: Intercalibrated passive microwave rain products from the Unified Microwave Ocean Retrieval Algorithm (UMORA). J. Appl. Meteor. Climatol., **47**, 778–794, doi: 10.1175/2007JAMC1635.1.

Hilburn, K. A., 2009: The passive microwave water cycle product. Remote Sensing Systems Tech. Rep. 072409, 30 pp. [Available at ftp://ftp.remss.com/water_cycle/Hilburn_water_cycle_REMSS_TR_072409.pdf.]

Ho, S.-P., Kuo, Y.-H., and Sokolovskiy, S., 2007: Improvement of the temperature and moisture retrievals in the lower troposphere using AIRS and GPS radio occultation measurements, J. Atmos. Ocean. Tech., **24**, 1726–1737, doi:10.1175/JTECH2071.1.

Hsu, K., X. Gao, S. Sorooshian, and H. Gupta, 1997: Precipitation estimation from remotely sensed information using artificial neural networks. J. Appl. Meteor., **36**, 1176–1190, doi: 10.1175/1520-0450(1997)036<1176:PEFRSI>2.0.CO;2.

Huffman, G. J., R. F. Adler, P. Arkin, A. Chang, R. Ferraro, A. Gruber, J. Janowiak, A. McNab, B. Rudolf, and U. Schneider, 1997: The Global Precipitation Climatology Project (GPCP) combined precipitation data record. Bull. Amer. Meteor. Soc., **78**, 5–20, doi: 10.1175/1520-0477(1997)078<0005:TGPCPG>2.0.CO;2.

——, ——, D. T. Bolvin, G. Gu, E. J. Nelkin, K. P. Bowman, Y. Hong, E. F. Stocker, and D. B. Wolff, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. J. Hydrometeor., **8**, 38–55, doi: 10.1175/JHM560.1.

Huffman, G. J., R. F. Adler, D. T. Bolvin, G. Gu, 2009: Improving the Global Precipitation Record: GPCP Version 2.1. Geophys. Res. Lett., 36,L17808, doi:10.1029/2009GL040000.

Huffman, G. J., D. T. Bolvin: TRMM and Other Data Precipitation Data Set Documentation. 25 May 2014, available at ftp://meso-a.gsfc.nasa.gov/pub/trmmdocs/3B42_3B43_doc.pdf.

Hyland, R. W. and A. Wexler, Formulations for the Thermodynamic Properties of the saturated Phases of H2O from 173.15K to 473.15K, ASHRAE Trans, **89(2A)**, 500-519, 1983.

Immler, F. J., J. Dykema, T. Gardiner, D. N. Whiteman, P. W. Thorne, and H. Vömel, 2010: Reference quality upper-air measurements: Guidance for developing GRUAN data products. Atmos. Meas. Tech., **3**, 1217–1231, doi:10.5194/amt-3-1217-2010.

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**, D16303, doi: 10.1029/2008JD011341.

Janowiak, J. E., A. Gruber, C. R. Kondragunta, R. E. Livezey, and G. J. Huffman, 1998: A comparison of the NCEP–NCAR reanalysis precipitation and the GPCP rain gauge–satellite combined data record with observational error considerations. J. Climate, **11**, 2960–2979, doi: 10.1175/1520-0442(1998)011<2960:ACOTNN>2.0.CO;2.

Jones, C., P. Peterson, and C. Gautier, 1999: A new method for deriving ocean surface specific humidity and air temperature: An artificial neural network approach. J. Appl. Meteor., **38**,1229–1246, doi: 10.1175/1520-0450(1999)038<1229:ANMFDO>2.0.CO;2.

Kelly, K. A., S. Dickinson, M. J. McPhaden, and G. C. Johnson, 2001: Ocean currents evident in satellite wind data. Geophys. Res. Lett., **28**, 2469–2472, doi: 10.1029/2000GL012610.

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.

Kent, E. C., S. Fangohr and D. I. Berry, 2013: A comparative assessment of monthly mean wind speed products over the global ocean, Internat. J. Climatol., **33**: 2520-2541, doi: 10.1002/joc.3606.

Kim, J.-E. and M. J. Alexander, 2013: Tropical Precipitation Variability and Convectively Coupled Equatorial Waves on Submonthly Time Scales in Reanalyses and TRMM, Journal of Climate, **26**, 3013-3030, doi: 10.1175/JCLI-D-12-00353.1.

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

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. Journal of Atmospheric and Oceanic Technology, **33**, 1455-1471, doi: http://dx.doi.org/10.1175/JTECH-D-15-0122.1.

Klepp, C. P., S. Bakan, and H. Graßl, 2003: Improvements of satellite-derived cyclonic rainfall over the North Atlantic. J. Climate, **16**, 657–669, doi: 10.1175/1520-0442(2003)016<0657:IOSDCR>2.0.CO;2.

——, ——, and ——, 2005: Missing North Atlantic cyclonic precipitation in ECMWF numerical weather prediction and ERA- 40 data detected through the satellite climatology HOAPS II. Meteor. Z., **14**, 809–821, doi: 10.1127/0941-2948/2005/0088.

Klepp, C., K. Bumke, S. Bakan, and P. Bauer, 2010: Ground validation of oceanic snowfall detection in satellite climatologies during LOFZY. Tellus, **62**, 469–480, doi: 10.1111/j.1600-0870.2010.00459.

Klepp, C., 2015: The oceanic shipboard precipitation measurement network for surface validation – OceanRAIN, Atmospheric Research, **163**, 74-90, doi: 10.1016/j.atmosres.2014.12.014.

Kobayashi, S., Y. Ota, Y. Harada, A. Ebita, M. Moriya, H. Onoda, K. Onogi, H. Kamahori, C. Kobayashi, H. Endo, K. Miyaoka, and K. Takahashi, 2015: The JRA-55 Reanalysis: General Specifications and Basic Characteristics. J. Met. Soc. Jap., **93** (1), 5-48, doi: 10.2151/jmsj.2015-001.

Kubota, M., N. Iwasaka, S. Kizu, M. Kondo, 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., A. Kano, H. Muramatsu, and H. Tomita, 2003: Intercomparison of Various Surface Latent Heat Flux Fields, J. Climate, **16**, 670-678m, doi: 10.1175/1520-0442(2003)016<0670:IOVSLH>2.0.CO;2

Kubota, M., and H. Tomita, 2007: Introduction of J-OFURO latent heat flux version 2. Proc. Joint 2007 EUMETSAT Meteorological Satellite Conf. and 15th Satellite Meteorology and Oceanography Conf., Amsterdam, Netherlands, EUMETSAT and Amer. Meteor. Soc.

Kubota, T., S. Shige, H. Hashizume, K. Aonashi, N. Takahashi, S. Seto, M. Hirose, Y. N. Takayabu, T. Ushio, K. Nakagawa, K. Iwanami, M. Kachi, and K. Okamoto, 2007: Global precipitation map using satellite-borne microwave radiometers by the GSMaP project: Production and validation. IEEE Trans. Geosci. Remote Sens., **45**, 2259–2275, doi: 10.1109/TGRS.2007.895337.

Kummerow, C., D. Randel, M. Kulie, N. Wang, R. Ferraro, S. Munchak, and V. Petkovic, 2015: The Evolution of the Goddard Profiling Algorithm to a Fully Parametric Scheme. J. Atmos. Oceanic Technol., **32**, 2265-2280, doi:10.1175/JTECH-D-15-0039.1.

Large, W. G., J. Morzel, and G. B. Crawford, 1995: Accounting for surface wave distortion of the marine wind profile in low-level ocean storm wind measurements, J. Phys. Oceanogr., **25**, 2959-2971, doi: 10.1175/1520-0485(1995)025<2959:AFSWDO>2.0.CO;2

Lauritsen, K. B., Syndergaard, S., Gleisner, H., Gorbunov, M. E., Rubek, F., Sørensen, M. B., and Wilhelmsen, H., 2011: Processing and validation of refractivity from GRAS radio occultation data, Atmos. Meas. Tech., **4**, 2065-2071, doi:10.5194/amt-4-2065-2011.

Liu, C. and R. P. Allan, 2013: Observed and simulated precipitation responses in wet and dry regions 1850-2100 , Environmental Research Letters, **8**, 034002, doi:10.1088/1748-9326/8/3/034002

Liu, J. P., and J. A. Curry, 2006: Variability of the tropical and subtropical ocean surface latent heat flux during 1989–2000. Geophys. Res. Lett., **33**, L05706, doi: 10.1029/2005GL024809.

Meissner, T., D. Smith, and F. Wentz, 2001: A 10 year intercomparison between collocated Special Sensor Microwave Imager oceanic surface wind speed retrievals and global analyses. J. Geophys. Res., **106**, 11 731–11 742, doi: 10.1029/1999JC000098.

Monahan, A. H., 2006: The probability distribution of sea surface wind speeds. Part II: Data record intercomparison and seasonal variability. J. Climate, **19**, 521–534, doi: 10.1175/JCLI3641.1.

O'Gorman, P. A., R. P. Allan, M. P. Byrne and M. Previdi, 2012: Energetic constraints on precipitation under climate change, Surv. Geophys., **33**, 585-608, doi: 10.1007/s10712-011-9159-6.

Pfeifroth, U., R. Mueller, and B. Ahrens, 2013: Evaluation of Satellite-Based and Reanalysis Precipitation Data in the Tropical Pacific, Journal of Applied Meteorology and Climatology, **52**, 634-644, doi: 10.1175/JAMC-D-12-049.1.

Rienecker, M. M., and 28 co-authors, 2011: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. J. Climate, **24**, 3624–3648, doi: 10.1175/JCLI-D-11-00015.1.

Roberts, J. B., F. R. Robertson, C. A. Clayson, and M. G. Bosilovich, 2012: Characterization of turbulent latent and sensible heat flux exchange between the atmosphere and ocean in MERRA. J. Climate, **25**, 821-838, doi: 10.1175/JCLI-D-11-00029.1.

Saha, S., and 51 co-authors, 2010: The NCEP Climate Forecast System Reanalysis. Bull. Amer. Meteor. Soc., **91**, 1015–1057, doi: 10.1175/2010BAMS3001.1.

Sapiano, M. R. P., and P. A. Arkin, 2009: An intercomparison and validation of high-resolution satellite precipitation estimates with 3-hourly gauge data. J. Hydrometeor., **10**, 149–166, doi: 10.1175/2008JHM1052.1.

Schlosser, C. A., and P. R. Houser, 2007: Assessing a satellite-era perspective of the global water cycle. J. Climate, **20**, 1316– 1338, doi: 10.1175/JCLI4057.1.

Schröder, M., Jonas, M., Lindau, R., Schulz, J., and Fennig, K., 2013: The CM SAF SSM/I-based total column water vapour climate data record: methods and evaluation against re-analyses and satellite. Atmos. Meas. Tech., 6, 765–775, doi:10.5194/amt-6-765-2013.

Schröder, M., M. Lockhoff, J. Forsythe, H. Cronk, T. H. Vonder Haar, R. Bennartz, 2016: The GEWEX water vapor assessment (G-VAP) – results from the trend and homogeneity analysis. J. Applied Meteor. Clim., **55** (7), 1633-1649, doi: /10.1175/JAMC-D-15-0304.1.

Shie C.-L., K. Hilburn, L. S. Chiu, R. Adler, I.-I. Lin, E. Nelkin, J. Ardizzone, and S. Gao, 2012: Goddard Satellite-Based Surface Turbulent Fluxes, Daily Grid, Version 3, Savtchenko A (ed). Greenbelt, MD: Goddard Earth Science Data and Information Services Center (GES DISC), doi: 10.5067/MEASURES/GSSTF/DATA301.

Shin, D.-B., J.-H. Kim, and H.-J. Park, 2011: Agreement between monthly precipitation estimates from TRMM satellite, NCEP reanalysis, and merged gauge-satellite analysis, J. Geophys. Res., **116**, D16105, doi:10.1029/2010JD015483.

Shinoda, T., H. H. Hendon, and J. Glick, 1999: Intraseasonal surface fluxes in the tropical western pacific and Indian oceans from NCEP reanalyses. Mon. Wea. Rev., **127**, 678–693, doi: 10.1175/1520-0493(1999)127<0678:ISFITT>2.0.CO;2.

Simmons, A., S. Uppala, D. Dee, and S. Kobayashi, 2007: ERA– Interim: New ECMWF reanalysis products from 1989 onwards. ECMWF Newsletter,No. **110**, ECMWF, Reading, United Kingdom, 25–35.

Smith, S. D., 1980: Wind stress and heat flux over the ocean in gale force winds. J. Phys. Oceanogr., **10**, 709–726, doi: 10.1175/1520-0485(1980)010<0709:WSAHFO>2.0.CO;2.

——, 1988: Coefficients for sea surface wind stress, heat flux, and wind profiles as a function of wind speed and temperature. J. Geophys. Res., **93** (C12), 15.467–15.472, doi: 10.1029/JC093iC12p15467

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.

Swift, C. T., L. S. Fedor, and R. O. Ramseier, 1985: An algorithm to measure sea ice concentration with microwave radiometers. J. Geophys. Res., **90**, 1087–1099, doi: 10.1029/JC090iC01p01087.

Tian, Y. and C. D. Peters-Lidard, 2010: A global map of uncertainties in satellite-based precipitation measurements, Geophys. Res. Letters, **37**, L24407, doi: 10.1029/2010GL046008.

Tournadre, J., and Y. Quilfen, 2003: Impact of rain cell on scatterometer data: 1. Theory and modeling. J. Geophys. Res., **108**, 3225, doi: 10.1029/2002JC001428.

Trenberth, K. E., and C. J. Guillemot, 1998: Evaluation of the atmospheric moisture and hydrological cycle in the NCEP/NCAR reanalyses. Climate Dyn., **14**, 213–231, doi: 10.1007/s003820050219.

——,J. T. Fasullo, and J. Kiehl, 2009: Earth's global energy budget. Bull. Amer. Meteor. Soc., **90**, 311–323, doi: 10.1175/2008BAMS2634.1.

——, 2011: Changes in precipitation with climate change. Clim. Res., **47** (1), 123, doi:10.3354/cr00953.

Wallcraft, A., A. Kara, C. Barron, E. Metzger, R. Pauley, and M. Bourassa, 2009: Comparisons of monthly mean 10 m wind speeds from satellites and NWP products over the global ocean. J. Geophys. Res., **114**, D16109, doi: 10.1029/2008JD011696.

Weller, R. A., E. F. Bradley, J. B. Edson, C. W. Fairall, I. Brooks, M. J. Yelland, and R. W. Pascal, 2008: Sensors for physical fluxes at the sea surface: Energy, heat, water salt. Ocean Sci., **4**, 247-263, doi: 10.5194/osd-5-327-2008.

Wentz, F. J., 1997: A well-calibrated ocean algorithm for Special Sensor Microwave/Imager. J. Geophys. Res., **102** (C4), 8703–8718, doi: 10.1029/96JC01751.

Wentz, F., and R. W. Spencer, 1998: SSM/I rain retrievals within a unified all-weather ocean algorithm. J. Atmos. Sci., **55**, 1613–1627, doi:10.1175/1520-0469(1998)055,1613:SIRRWA.2.0.CO;2.

Wilkinson, K., M. von Zabern, and J. Scherzer, 2014: Global Freshwater Fluxes into the World Oceans: Technical Report prepared for the GRDC. - Koblenz, Federal Institute of Hydrology (BfG). - (GRDC Report; 44). doi: 10.5675/GRDC_Report_44. [Available online at http://www.bafg.de/GRDC/EN/02_srvcs/24_rprtsrs/ report_44.pdf? __blob=publicationFile.]

Willmott, C. J. and K. Matsuura, 2005: Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research, **30**, 79-82.

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 Trans. Geosci. Remote Sens., **48**, 338–348, doi: 10.1109/TGRS.2009.2023982.

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

Worley, S., S. Woodruff, R. Reynolds, S. Lubker, and N. Lott, 2005: ICOADS release 2.1 data and products. Int. J. Climatol., **25**, 823–842, doi: 10.1002/joc.1166.

Xie, P., and P. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. Bull. Amer. Meteor. Soc., 78, 2539–2558, doi: 10.1175/1520-0477(1997)078<2539:GPAYMA>2.0.CO;2.

Yu, L. S., and R. A. Weller, 2007: Objectively analyzed air–sea heat fluxes for the global ice-free oceans (1981–2005). Bull. Amer. Meteor. Soc., **88**, 527–539, doi: 10.1175/BAMS-88-4-527.

——, X. Jin, and R. A. Weller, 2008: Multidecade global flux data records from the objectively analyzed air–sea fluxes (OAFlux) project: Latent and sensible heat fluxes, ocean evaporation, and related surface meteorological variables. Tech. Rep. OA- 2008-01, Woods Hole Oceanographic Institution, OAFlux Project, 64 pp.

11 Annex

For completeness the relative differences between data from HOAPS-4.0 and from a reference are shown in Figure 26. Budget is not shown as the denominator is distributed around 0 mm/d.
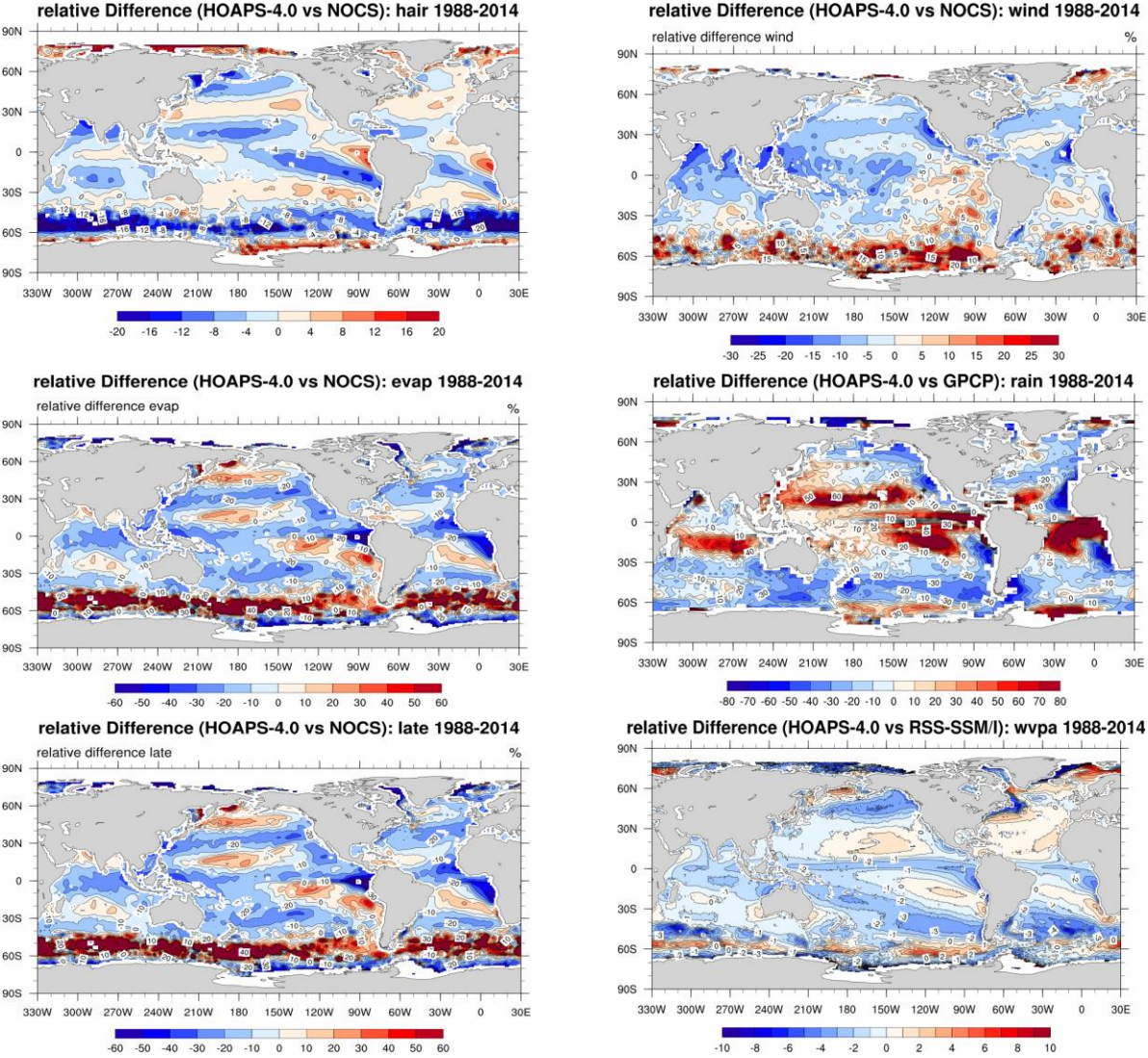
Figure 26: Relative differences of near surface specific humidity (hair), wind speed (wind), evaporation (evap), rain rate (rain), latent heat flux (late) and vertically integrated water vapour (wvpa). Note the differences in scale.