Consistency of satellite climate data records for Earth system monitoring

Consistency of satellite climate data records for Earth system monitoring

Thomas Popp[1], Michaela I. Hegglin[2], Rainer Hollmann[3], Fabrice Ardhuin[4], Annett Bartsch[5], Ana Bastos[6], Victoria Bennett[7], Jacqueline Boutin[8], Carsten Brockmann[9], Michael Buchwitz[10], Emilio Chuvieco[11], Philippe Ciais[12], Wouter Dorigo[13], Darren Ghent[14], Richard Jones[15], Thomas Lavergne[16], Christopher J. Merchant[2,17], Benoit Meyssignac[18], Frank Paul[19], Shaun Quegan[20], Shubha Sathyendranath[21], Tracy Scanlon[13], Marc Schröder[3], Stefan G. H. Simis[21], Ulrika Willén[22]

[1] *German Aerospace Center (DLR), Wessling, Germany*

[2] *University of Reading, Reading, UK*

[3] *Deutscher Wetterdienst (DWD), Offenbach, Germany*

[4] *Lab. of Ocean Physics and Satellite oceanography, Ifremer, Plouzané, France*

[5] *b.geos GmbH, Korneuburg, Austria*

[6] *Dept. Of Geography, Ludwig-Maximilians-Universität, München, Germany*

[7] *Centre for Environmental Data Analysis, STFC Rutherford Appleton Laboratory, Harwell, UK and National Centre for Earth Observation, UK*

[8] *Sorbonne Université, CNRS, IRD, MNHN, Laboratoire d'Océanographie et du Climat: Expérimentations et Approches Numériques (LOCEAN), Paris, France*

[9] *Brockmann Consult GmbH, Hamburg, Germany*

[10] *Institute of Environmental Physics (IUP), University of Bremen, Bremen, Germany*

[11] *Environmental Remote Sensing, University of Alcalá, Alcalá de Henares, Spain*

[12] *IPSL – LSCE, Gif sur Yvette, France*

[13] *TU Wien, Department of Geodesy and Geoinformation, Vienna, Austria*

[14] *National Centre for Earth Observation, Department of Physics & Astronomy, University of Leicester, Leicester, UK*

Early Online Release: This preliminary version has been accepted for publication in *Bulletin of the American Meteorological Society*, may be fully cited, and has been assigned DOI 10.1175/BAMS-D-19-0127.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

[15] *Met Office Hadley Centre, Exeter, UK*

[16] *Norwegian Meteorological Institute, Oslo, Norway*

[17] *National Centre for Earth Observation, University of Reading, Reading, UK*

[18] *LEGOS CNES, CNRS, IRD, Université de Toulouse, Toulouse, France*

[19] *Department of Geography, University of Zurich, Zurich, Switzerland*

[20] *School of Mathematics and Statistics, University of Sheffield, Sheffield, UK*

[21] *Plymouth Marine Laboratory, Plymouth, UK*

[22] *Swedish Hydrological and Meteorological Institute, Norrköping, Sweden*

Corresponding author: Thomas Popp, German Aerospace Center (DLR), Münchner Str. 20, 82234 Wessling, Germany, Phone: +49 / 8153 28 1382, E-mail: thomas.popp@dlr.de

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

Abstract

Climate Data Records (CDRs) of Essential Climate Variables (ECVs) as defined by the Global Climate Observing System (GCOS) derived from satellite instruments help to characterize the main components of the Earth system, to identify the state and evolution of its processes, and to constrain the budgets of key cycles of water, carbon and energy. The Climate Change Initiative (CCI) of the European Space Agency (ESA) coordinates the derivation of CDRs for 21 GCOS ECVs. The combined use of multiple ECVs for Earth system science applications requires consistency between and across their respective CDRs. As a comprehensive definition for multi-ECV consistency is missing so far, this study proposes defining consistency on three levels: (1) consistency in format and metadata to facilitate their synergetic use (technical level); (2) consistency in assumptions and auxiliary datasets to minimize incompatibilities among datasets (retrieval level); and (3) consistency between combined or multiple CDRs within their estimated uncertainties or physical constraints (scientific level).

Analysing consistency between CDRs of multiple quantities is a challenging task and requires coordination between different observational communities, which is facilitated by the CCI program. The inter-dependencies of the satellite-based CDRs derived within the CCI program are analysed to identify where consistency considerations are most important. The study also summarizes measures taken in CCI to ensure consistency on the technical level, and develops a concept for assessing consistency on the retrieval and scientific levels in the light of underlying physical knowledge. Finally, this study presents the current status of consistency between the CCI CDRs and future efforts needed to further improve it.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

Capsule

In this study, the ESA Climate Change Initiative (CCI) introduces a three-level definition of consistency between multiple satellite-based Climate Data Records (CDRs) of Essential Climate Variables (ECVs), discusses consistency status and requirements and develops a concept for assessing inter and across ECV consistency.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

1. Introduction

The Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5) and the three Special Reports of the AR6 cycle state that mankind and the biosphere face great threats due to the rapidly changing climate (IPCC, 2013, 2018, 2019a, 2019b). To support political decisions on climate change mitigation and adaptation, and to quantify the implications for economic and non-economic loss and damage, the United Nations Framework Convention on Climate Change (UNFCCC) requires systematic monitoring of the global climate system (e.g. Doherty et al., 2009; UNFCCC Art. 4 and Art. 5, 1992; Paris Agreement 7.7c, Adaptation). In particular, systematic monitoring is important in assessing progress on the aims of the Paris Agreement (e.g. for the global stocktake). The main tools at hand to determine the extent and impacts of climate change on local to global scales and understand its causes are a combination of global and regional climate and Earth system models, reanalysis data, and systematic observations. The latter are indispensable for all Earth system domains (atmospheric, terrestrial, oceanic) to increase the understanding of and quantify processes, budgets and reservoirs within the global Earth cycles (carbon, energy, and water).

To promote systematic climate monitoring, the World Meteorological Organization (WMO), Intergovernmental Oceanographic Commission (IOC), United Nations Environment Program (UNEP), and International Science Council (ISC), established in 1992 the Global Climate Observing System (GCOS). GCOS aims at sustained "provision of reliable physical, chemical and bio-chemical observations and data records for the total climate system – across the atmospheric, oceanic and terrestrial domains, including hydrological and carbon cycles and the cryosphere" (GCOS, 2016). GCOS defined a set of currently 54 "Essential Climate Variables" or ECVs (Bojinski et al., 2014) which must be observed in a sustained and

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

consistent manner to enable detection of climate trends and provide data suitable for climate model evaluation and climate change attribution.

Complementary to relatively sparse airborne and ground-based measurements and inventory data, satellite observations are of ever-growing importance for evaluating, initializing and parameterizing Earth system processes represented in models. This growing importance is due to the increasing satellite global coverage and resolution (in space and time), their improved calibration accuracy and the increasing diversity of relevant observables provided by advances in satellite sensor technologies. Satellite observations can provide a significant contribution for 21 out of the 54 GCOS ECVs. Some of these are exclusively derived from satellite measurements (e.g. the Earth Radiation Budget), whereas for others dedicated spaceborne sensors provide better coverage but lower accuracy or resolution than in situ measurements (e.g. above-ground biomass, column atmospheric concentration of CO_2 and CH_4).

Studies of the Earth system require combined analysis of datasets of many variables. Since these are derived from different sources (satellite-, ground-, air- and model-based) and processing systems, one underlying precondition of any such analysis is that the datasets are **consistent**. However, despite the importance of consistency, many open questions remain, ranging from a clear definition of consistency for multiple quantities, to systematically assessing consistency between the many data records used.

Possible reasons for inconsistencies include the use of different auxiliary datasets, simplifications in corrections and retrieval algorithms, calibration uncertainties and differences in sampling and gridding. For example, a time series of a single variable built from data records obtained from different sensors may exhibit "jumps" where they are merged with each other, which may hinder any trend analysis. Another example is assigning different land cover classes (e.g. glacier, water, rock or vegetation) to the same pixel by using

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

different glacier masks, which may lead to highly variable budget calculations of related exchange processes.

Consistency as an issue in creating satellite-based data records was first met by operational entities like NOAA, EUMETSAT or NASA within their near-real time (NRT) processing chains across different satellite missions. This includes aspects such as common input datasets, gridding methodology, cloud and land/sea masking, aerosol and water vapour corrections, and the land cover map used. The measures taken are typically documented in Algorithm Theoretical Baseline Documents (e.g. consistent OMI-MODIS cloud products: Siddans, 2016 or merged TROPOMI-VIIRS cloud product: NASA, 2014). However, the need for consistency across different variables, domains and processing systems is inherent in climate studies and thus much broader than in the often independent NRT applications.

Over the past ten years, space agencies (including ESA, EUMETSAT, NASA, and NOAA) have emphasised the generation and delivery of satellite-based CDRs. Hollmann et al. (2013) describe the efforts of ESA through its Climate Change Initiative (CCI). CCI leverages and harvests the long-term satellite archives available from European and other satellites, and enhances or expands these records with observations from other space agencies to obtain global coverage. In addition, CCI is extending its newly established CDRs with the most recent satellite instruments to guarantee continuation into the future using operational missions (e.g. Sentinel). During its first six years (2011-2017), CCI implemented 14 projects, each targeting one (or two) ECVs; in 2018, CCI was expanded to include nine additional ECVs, as shown in Figure 1. It should be noted that most of the ECVs consist of several quantities, so-called products (detailed information on the products of each CCI ECV for which CDRs have been processed in CCI is available at http://cci.esa.int). Of course, products within a particular ECV have to be consistent. A particular element within the CCI program is

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

independent analysis of the quality of its CDRs and particularly their consistency (between different ECVs and products) in a climate modelling context by the CCI Climate Model User Group (CMUG) and several budget closure study projects.

Together with the Copernicus Climate Change Service (C3S) and contributions from EUMETSAT through its Satellite Application Facilities (SAFs) such as the Climate Monitoring SAF (Schulz et al., 2009), the NOAA Climate Data Record program (https://www.ncdc.noaa.gov/cdr, Bates et al. 2016), and the NASA Measures program (https://earthdata.nasa.gov/measures), about 1000 different satellite-based CDRs for GCOS ECV products and further variables are available or will become available in the near future. An overview of these CDRs is given in the ECV inventory (https://climatemonitoring.info/ecvinventory), recently established by the joint CEOS-CGMS Working Group on Climate, which conducts regular gap analysis to define future satellite development needs.

This study introduces a concept developed in CCI to define and assess consistency between multiple satellite-based ECV products. It is shown that such an assessment allows remaining inconsistencies to be identified and quantified in the light of given CDR uncertainties and relevant physical principles. A key application of assessing and ensuring consistency is in closure studies where multiple CDRs are used together. A selection of topics for such closure studies is briefly discussed in this paper to illustrate the concept.

Section 2 discusses different kinds of inconsistencies and develops a definition of consistency, followed by a brief analysis of ECVs covered by CDRs from CCI and consistency needs in Earth system monitoring in Section 3. Section 4 develops a concept for assessing the different levels of consistency and illustrates it with examples from different ECV products in CCI. Section 5 presents a discussion of the main findings and identifies remaining consistency gaps.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

2. Consistency in Earth system monitoring

Whilst ”consistency” (e.g. between two datasets) is a concept frequently referred to in the observation community, there is, to our knowledge, no comprehensive definition specific to observation datasets of different variables. This may reflect the complexity of relations between the large set of ECVs. This study proposes such a comprehensive definition and an assessment concept for consistency. The focus is on consistency between datasets of different variables, as needed for climate studies, but also single-variable cases are included.

According to the common definition of the word “consistency” (Oxford dictionary), it is “the quality of always behaving in the same way or of having the same opinions or standards; the quality of being consistent, i.e., 1/ in agreement with something; not contradicting something, 2/ happening in the same way and continuing for a period of time, 3/ consistent with something in agreement with something, not contradicting something, 4/ having different parts that all agree with each other”. In the observation scientific community, consistency is usually understood as “agreement”, “compatibility” or “no contradiction”. When considering CDRs, “consistency” goes beyond “agreement” and rather refers to “compatibility”. Firstly, agreement per se can only be tested between datasets of the same variable. A mature terminology and a comprehensive set of mathematical tools for this purpose exists, which forms the basis of most calibration, validation and model evaluation activities. Secondly, there can even be cases where two datasets of the same variable agree (their bias is smaller than their combined uncertainties) but are inconsistent (for example if only one of two datasets shows a distinct diurnal or seasonal cycle). In contrast, regionally averaged time series of one variable can disagree (have regional biases larger than the combined uncertainties), but be consistent in their temporal behaviour, as shown for multi-sensor AOD records (Sogacheva, et al., 2020).

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

In a physical sense, consistency can be understood as fulfilling a conservation balance equation (of mass or energy) or exhibiting a correlation in time or space between two data records as expected by a physical theory. In CDR production, also simple category inconsistencies occur (e.g. for one pixel land cover assigns bare soil, while biomass gives a non-zero carbon mass to it).

Immler et al. (2010) defined consistency between measurements of the GCOS Reference Upper Air Network (GRUAN) as “when the independent measurements agree to within their individual uncertainties”, which requires knowledge of their (combined) uncertainties. This definition applies to different measurements of the same variable, but in the wider context of Earth system monitoring, a definition of consistency across multiple ECVs is also needed.

Several kinds of inconsistency between different data records of the same quantity or of different quantities can be recognised:

- Inconsistencies due to differences in auxiliary datasets;
- Temporal inhomogeneities in time series (e.g. due to calibration biases, degradation in data obtained from a sequence of different input data records, or sampling differences in terms of measurement time, frequency, or geographical coverage during gridding);
- Spatial inhomogeneities due to combining fields from different datasets (e.g. with different observing geometry or different sampling, e.g. all-sky versus clear-sky sampling).

Many of these inconsistencies are linked to the statistical properties of the raw data used to create a CDR, when for practical reasons simplifications and aggregations cannot be avoided.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

To cover the wide range of aspects of consistency, it is convenient to structure it on three complementary levels:

(1) Consistency on the technical level: Harmonised data format and metadata description to ease acquisition and combined usage of multiple CDRs;
(2) Consistency on the retrieval level: Use of the same auxiliary datasets in retrievals to minimize contradictions in outputs linked to common information (e.g. a water mask);
(3) Consistency on the scientific level: Compatibility of the relevant characteristics of two or more CDRs (e.g. patterns, variability, trends, …) with a reference (represented by a physical equation, a model or a fiducial reference) within their combined uncertainties.

While consistency on a **technical level** is easy to define and needs limited scientific insight, it is often a resource-consuming barrier hindering data use. Thus the Earth observation community has focused on this area in recent years (e.g. by adopting common metadata standards). In particular, the CCI program has adopted existing solutions (and when needed developed new ones) that facilitate combined satellite-based CDR use. This includes a harmonized data format (netCDF, with a few exceptions where a different standard is needed for a particular community, e.g. shapefiles for glaciers) and a common metadata convention (CCI data standards: ESA, 2019), which follow the CF convention (http://cfconventions.org). It covers additional cross-ECV standardized metadata attributes, using common vocabularies for index terms and harmonized variable names, as well as a harmonized / interoperable data access portal with common catalogue and data services to simplify multi-quantity data search and download within the CCI portfolio (http://cci.esa.int/data). This common vocabulary also helps to reduce inconsistent nomenclature, such as labelling slightly different variables as the same retrieved quantity

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

(e.g. due to wavelength dependencies of retrieved information). Furthermore, the underlying documentation of algorithms and datasets in CCI has been harmonized to some extent, as in other initiatives such as the SAF network or NOAA CDR program. This information helps users to quickly understand each dataset and its strengths, weaknesses and limitations. A good example of the benefit of such harmonised climate data records on the technical level is given by the CCI toolbox (https://climatetoolbox.io), which can be used for harmonized data pre-processing, analysis and visualisation of the multiple CDRs in a standardized way.

On the **retrieval level**, consistency aims at using the same (or a similar) observation strategy (same or similar satellite sensors, frequencies, etc.), and similar auxiliary datasets for the same variable in different retrieval algorithms. Those auxiliary datasets are either categorical datasets, so-called "masks", or continuous datasets of physical variables. Typical masks used in many retrieval algorithms include, for example, a particular land cover (vegetated areas), land-water, sea ice, snow cover and glacier masks, since many retrieval algorithms behave differently over different surface types. Other masks commonly needed across many variables are cloud masks, since many retrievals in the visible to thermal spectral range need to avoid contamination by clouds. Frequently used continuous auxiliary data fields include meteorological fields (e.g. from reanalysis) and climatologies of atmospheric variables (e.g. water vapour, aerosols, ozone) to conduct atmospheric corrections of visible bands used to retrieve land and ocean ECVs.

There is no sharp boundary between retrieval and scientific consistency. Ultimately, **scientific consistency** deals with the compatibility in CDR properties relevant for climate processes. All data records of a single ECV product, if obtained from different sources, need to be consistent within their uncertainties and within sampling differences. One aspect is consistency across borders in space (horizontally and vertically) and in time. Most

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

importantly, systematic biases between datasets need to be avoided as they may lead to errors when evaluating model performance (e.g. Waugh and Eyring, 2008). This applies to different combinations such as one variable based on multiple sensors, one sensor but using multiple algorithms, or combined satellite, model and in situ data. Finally, when several datasets of different variables are included in a physical model or budget equation, multi-variable consistency needs to distinguish uncertainties of calculated closure budgets due to propagated input uncertainties from real physical process imbalances or net effects.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

3. Consistency needs for CCI Earth System Climate Data Records

In this section, linkages on the retrieval and scientific level between the different CCI ECVs (Figure 1) are analysed. This analysis remains at the high level of the GCOS ECVs while it is well understood that most ECVs consist of several different quantities, or so-called products (e.g. the glacier ECV in CCI consists of the three products glacier outlines, elevation change and velocity). In most of the analysis in this study the primary product of an ECV is considered (e.g. aerosol optical depth for aerosol properties) and the most common methodology used to retrieve it. This means that for using a specific CDR of one ECV there may be a need to assess in more detail its linkages if, for example, a new retrieval technique in another spectral range is considered or if another product of this ECV is assessed. Detailed information on the products of each CCI ECV for which CDRs have been processed is available at http://cci.esa.int.

As a first step, the needs for consistency between ECVs on the **retrieval level** are assessed. Retrievals of Earth system variables from satellite observations aim to produce high quality CDRs by constraining the (often under-determined) inversion equations as good as possible. Typically, the measurements are chosen to have high sensitivity to the target variable, but they are usually subject to perturbations from other variables. In such cases, the inversion needs to either co-retrieve these additional variables or use auxiliary datasets to describe their spatio-temporal distributions. Moreover, different retrieval algorithms are often optimal for use over different surface types as their reflectance or spectral characteristics are highly variable (e.g. over dark water or over bright land). The use of different approaches for obtaining the same variable in different retrieval algorithms is one possible source of inconsistency between CDRs.

After processing, all CDRs have to pass validation against external reference datasets (e.g. from ground-based stations) to quantify their accuracy. Furthermore, CCI insists for

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

CDRs to be accompanied by proper uncertainty characterisation (using error propagation or uncertainty characterization during validation) within their data files (Merchant et al., 2017), so that uncertainties can be assessed when using the datasets. However, since reference data can have temporal or spatial representativeness issues and different validation methods also have their inconsistencies, unexplored uncertainties may remain (for the retrieved values themselves and for the estimated uncertainties). Validation and error propagation implicitly quantify inconsistencies from using imperfect auxiliary datasets and retrieval simplifications to within uncertainties. However, proof of consistency needs to explicitly test together the CDRs considered.

The part of Table 1 that is above the diagonal summarizes links between ECVs generated and analysed by CCI with regard to their retrieval consistency. A need for retrieval consistency is identified where either one or both retrievals rely on consistent co-retrieved or auxiliary variables of the other ECV (links only within CCI are considered, but there are other products, algorithms or sensors for which these may not apply).

The part of Table 1 below the diagonal summarizes the need for consistency on the **scientific level** based on our knowledge of how two variables are linked by Earth system processes or cycles in more detail. For this, the relevance of CCI ECVs for the energy, water and carbon cycles is briefly recalled. Figure 2 lists available or upcoming ECVs for which ESA CCI generates CDRs that contribute to the characterisation of these three main cycles. For simplicity, each ECV is only attributed to the cycle in which it plays the most important role. Practically all ECVs contribute to the energy cycle, either directly through radiation interaction or through mass-attached energy transport in the water or carbon cycle. Studies of sub-elements of these main cycles may also be relevant (e.g. physical processes such as emission, transport, deposition or radiation interactions, chemical transformations; also regional limitations, such as ice-free conditions) which may only require consistency among a

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

reduced set of ECVs. Some further details on the CCI CDRs for the three cycles are provided in the following.

Carbon cycle: CCI CDRs help quantifying the dynamics of the amount of carbon stored in the atmosphere, oceans and terrestrial biosphere and the fluxes between these reservoirs (see overview about the carbon cycle in Le Quére et al., 2018). CO_2 in the atmosphere is a key measure of the anthropogenic perturbation to the carbon cycle. The air-sea CO_2 flux is strongly affected by sea-surface temperature (SST) and ocean photosynthetic activity (monitored using ocean colour observations). The CCI CDRs also help constraining carbon fluxes from the land biosphere (e.g. Reuter, et al., 2017) including land use change and biomass burning emissions, together with direct estimates of above-ground biomass and burned area (Chuvieco et al., 2019). Other CCI CDRs of importance to the carbon cycle are snow cover (which affects the duration and start of photosynthetic processes in boreal forests; Pulliainen et al., 2017), similar to the impact of sea ice on marine photosynthesis in high latitudes, soil moisture (which affects land-atmosphere CO2 fluxes), permafrost (which contains frozen carbon stores with about twice the mass of atmospheric carbon), and sea surface salinity, which, together with SST, determines CO_2 solubility, with important impacts in rainy regions and serves as a proxy for sea water alkalinity (Vinogradova et al., 2019).

Water cycle: CCI helps quantifying the global water cycle over land and ocean (see overview in e.g. Levizzani and Cattani, 2019) by providing CDRs related to the reservoirs within the water cycle (lake levels, sea level, sea ice, ice sheets, glaciers, soil moisture, and snow), atmospheric water vapour content (water vapour) and clouds. From these, processes such as precipitation and runoff that transfer water between the various reservoirs may be inferred. CCI delivers additional relevant parameters such as sea surface salinity (related to precipitation, evaporation and runoff), SST and LST (determining evaporation), land cover and biomass (both linked to evapotranspiration).

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-19-0127.1.

Energy cycle*:* CCI also helps constraining the global energy cycle (for an overview see Allan, 2012) by providing CDRs for SST and LST, land and sea ice, as well as snow cover, sea level (which is affected among others by the ocean heat content and land ice melt), sea state, clouds, water vapour, ozone, greenhouse gases and aerosols that help determine the vertical temperature structure of the atmosphere. Finally, the biosphere (biomass) may also be considered a part of the energy cycle since it converts solar energy into chemically-stored energy (organic matter). In the oceans, a significant portion of the organic matter sinks out of the surface layers, exporting the energy to the deep ocean (with the photosynthesis activity being observed indirectly through ocean colour).

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

4. Concept for assessing consistency on different levels

Due to the complexity of different consistency aspects no single method can be used for assessing consistency of CDRs on various levels. Therefore, a concept employing a range of appropriate methods was developed in CCI, which is summarized here and then illustrated with short examples.

4.1 Overview: Methods to assess consistency

All methods for assessing consistency contain several key elements. Firstly, any method needs to be based on physical background knowledge to understand the relevance of any disagreement or incompatibility. Such background knowledge can be a simple principle (e.g. if the land cover is bare soil and the biomass product provides a high biomass value, there is an obvious inconsistency) or knowledge of the sensitivity of a target variable toward an auxiliary dataset, or a more complex physical equation or "model". Secondly, any assessment needs to select an appropriate characteristic (patterns, time series, masks) tailored towards the relevant process (or cycle) and choose a suitable mathematical tool (metric). Finally, this metric needs to be evaluated against the relevant physical background knowledge while the threshold on the chosen metric for judging consistency depends on the considered process or cycle and the datasets. In order to make any assessment of consistency objective, a study needs to specify the threshold used. This is shown for the following examples for various metrics.

In essence, consistency then means that several datasets have been evaluated against the underlying physical background knowledge and were found "fit for purpose" for a specific application domain. This leads to cases where seemingly small values of a chosen metric (compared to its uncertainty) can mean inconsistency, whilst in other cases apparently large deviations mean consistency, as will be shown in the examples of this section. Table 2 lists a

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

variety of related basic principles and methods to assess consistency on different levels used in the following examples.

4.2 Methods to assess retrieval level consistency

As a principle, retrieval level inconsistencies become significant if the difference of the auxiliary data used in two independent processing systems multiplied by the sensitivity of the target variable to the respective auxiliary variable is larger than the target uncertainty. This means that testing retrieval level consistency needs to assess auxiliary dataset differences in the light of target variable sensitivities or incompatibilities.

Consistency of categorical auxiliary datasets ("masks")

A first approach to assess consistency of masks used in independent retrievals lies in visual inspection of combined maps of datasets, as for example, of surface temperature composed from four independent CDRs for land (LST), sea surface (SST), ice (IST) and lake surface water (LSWT) temperatures against required pixel-level agreement of the masks. In CCI the four retrievals use a common land-sea mask (and sea-ice mask), but apply different cloud mask algorithms optimized over land, sea ice and water surfaces. As shown in Figure 3, the reader can visually confirm the absence of any obvious scatter near the land-sea borders, which indicates that the land-sea masks used in the different processing systems are consistent. Additionally, the application of different optimal cloud masking in the retrievals for LST and SST has led to obvious discontinuities in the sampling with temperature observations at the land-sea border, which may be judged as second-order inconsistencies. Such visual inspection of a set of typical scenes can be employed for most ECVs to get an understanding of their physical consistency within one variable across borders of the same mask used in different retrievals. Additionally, Figure 3 shows a case where a contrast in the

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-19-0127.1.

values in the ECVs between neighbouring pixels (surface temperature of ocean and water) does not mean inconsistency, but reflects physical differences arising from the different heat capacities of water and land.

The retrieval of many ECVs needs a cloud mask to avoid cloud contamination. Also cloud properties need a cloud mask to ensure that a pixel truly represents cloud (Poulson, et al., 2012). When, for example, aerosol and cloud property retrievals for the same sensor are implemented as separate algorithms (as is usually the case), individual pixels need to be analysed either as cloud or as aerosol; analysis of the same pixel as aerosol and as cloud under the wrong assumption (cloud-free or aerosol-free) could severely degrade the retrievals and must be minimized (e.g. Sogacheva et al., 2017; Li et al., 2009). To assess if this principle is fulfilled, independent AATSR cloud masks used in the aerosol and cloud products were analysed for four days in September 2008 (covering difficult scenes with high aerosol loads or complicated mixtures of aerosol and clouds). Figure 4 shows a map of different combinations of cloud / no cloud assignment by the two cloud masks and a contingency matrix of those class combinations. The matrix shows, that while 21% of observations are not used for aerosol or cloud retrievals at all (losing sampling coverage but not leading to inconsistency), only 0.3% of them were found to violate the physical principle (i.e., no pixel must be double-analysed as clouds and as aerosols). Even if a very stringent threshold for this fraction of 1% is set (since cloud mask errors lead to very large AOD errors) the two cloud masks are fully consistent. The map also shows that the inconsistent cases (yellow pixels) occur only over land but in all climate zones. Together with the underlying physical principle one can use such a contingency matrix / mapping of class combinations to assess the contingency of masks and to understand where / when inconsistencies mostly occur and need to be corrected.

Another typical aspect of multi-quantity spatial consistency is the agreement of

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

locations between the outlines of physically related quantities (different products within one ECV, between different ECVs). For example, glacier outlines are derived from high-resolution satellite imagery or aerial photography using semi-automated mapping techniques or manual on-screen digitization (Paul et al., 2015). Due to their higher spatial resolution, the location of glaciers can be used for land cover as an independent validation source for its "permanent ice and snow" classes. Furthermore, glacier maps serve as an important auxiliary dataset for clouds and LST (to choose the correct retrieval algorithm), and lakes as a reciprocal mask (these can only occur in places not covered by glaciers) for sea ice, ice sheets and permafrost. Again, contingency matrices between glacier or lake location and the other variables can be used to assess consistency in the light of the expected compatible combinations; the threshold for the acceptable fraction of inconsistent pixels needs to be set depending on the potential harm of misclassifications. A limitation for the assessment of categorical auxiliary datasets lies in the fact that mixed cases often exist, in particular for coarser spatial resolutions.

Consistency of continuous auxiliary datasets of the same quantity

Often the retrieval of a land / ocean CDR is affected by perturbations in the measured bands due to atmospheric absorption or scattering, so an atmospheric correction needs to be applied. Examples of necessary atmospheric corrections include visible or thermal retrievals impacted by aerosol, water vapour, ozone or other trace gases (e.g. Popp 1995). A first step in algorithm development would be to assess the sensitivities of the measured reflectances to the various absorbing trace gases and to aerosol particles (e.g. Holzer-Popp, et al., 2002). This provides the basis for deciding which corrections can be neglected or made with a simple parameterization, and which need more precise corrections using an auxiliary dataset of distributions of the responsible agents influencing the signal. When the auxiliary

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

datasets come from the same sensor as the target CDR, accurate spatio-temporal matching (pixel colocation) would be possible. However, in cases where the auxiliary data come from different sensors, it may be necessary to deal with spatial and temporal mismatches, introducing a requirement for assessment of the associated additional uncertainties. Figure 5 shows a gridded map of differences of aerosol optical depth between the by-products of the ocean-colour atmospheric correction of MERIS data (processed using a NASA algorithm) and the corresponding CCI aerosol ECV product from AATSR, both for 865 nm (both sensors were on-board the same platform ENVISAT and thus exhibit zero time difference). The global average difference of AOD between both products of 0.03 is acceptable for the purpose of aerosol corrections, but the variability is larger for the aerosol ECV product than for the ocean colour product (0.10±0.11 cf. 0.13±0.04 respectively), which has higher AOD values over the open ocean, but lower ones closer to land. Given the importance of AOD in ocean-colour atmospheric correction (IOCCG, 2010), the aerosol-corrected ocean colour ECV can be regarded as consistent with the aerosol ECV in its global average, but not regionally. These results merit further investigations to identify the sources of the discrepancies and to assess the potential to improve the MERIS ocean-colour atmospheric correction algorithm by using concurrent auxiliary information on AOD from the main ECV aerosol product obtained from AATSR.

<u>4.3 Methods to assess scientific consistency</u>

Scientific consistency includes self-consistency within one quantity (when independently retrieved pieces are integrated into a longer time series or a larger map) and mutual consistency between different quantities (different products of one ECV or multiple ECV CDRs) as a consequence of all types of retrieval inconsistencies, limitations of the

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

retrieval algorithms or sensor calibrations, as well as sampling differences between aggregated datasets.

Self-consistency of a single quantity

One major problem of satellite-based CDRs is that satellite instruments typically survive in orbit only for a limited time, so that a long-term record needs to be constructed from combining data from a time series of similar sensors. Plotting regional or global data records of the related parts of a time series often allows visual inspection of their consistency, where “jumps” or “breakpoints” are obvious against background knowledge of any known or absent true discontinuities. As example, column-averaged dry-air mole fractions (“vertical columns” XCO_2) of carbon dioxide (Buchwitz et al., 2015) from the greenhouse gas (GHG) ECV are selected. Those CDRs serve as input data for inverse modelling schemes to improve the knowledge on natural and anthropogenic sources and sinks (e.g. Reuter et al., 2017). In creating a multi-sensor CDR covering a longer time period, a merging algorithm (EMMA, Reuter et al., 2013, 2020) corrects potential remaining offsets of individual datasets to avoid jumps in the merged time series. In EMMA, the ensemble members have been bias corrected and brought to common a priori CO_2 profiles before being combined to obtain the merged product. Figure 6 shows at the top the resulting multi-sensor, multi-algorithm monthly mean XCO_2 merged record for 2003-2018 for northern mid-latitudes (30°N-60°N) with the known nearly linear increase in time and seasonal cycle and no remaining biases, while in the bottom panel differences between individual ensemble members and the merged product before the corrections are shown to be larger than the required XCO_2 uncertainties of 0.5 ppm. In this case, this threshold for the target uncertainties of the gap-corrected merged dataset is defined by the user requirement for the application of XCO_2 trend analysis.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

Similarly, spatial inconsistencies in one variable can often be assessed visually by looking at maps combined from independent pieces (different sensors, different overpass times of the same sensor with different observing geometries, different algorithms). In this case, inconsistencies are visible as artificial border lines or gradients that are larger than the noise in the image. Again, physical understanding is needed to decide whether a discontinuity at a physical border is real or erroneous (e.g. surface temperature often shows true differences between land and sea as shown in Fig. 3, while a dust plume should be continuous). Another example for spatial inconsistencies revealed by data overlay are glacier outlines derived from satellite images that have been orthorectified with different digital elevation models (DEMs). In steep and/or high topography geolocation shifts of several pixels (about 30 - 90 m) can occur, making any change assessment (trend analysis) or joint use of sensors nearly impossible (Kääb et al., 2016).

Another way of testing the consistency of independently retrieved CDRs for one variable is by comparing estimates of a derived quantity such as their trend with a physical equation. For example, within the GEWEX Water Vapour Assessment (G-VAP, see http://gewex-vap.org/ for details), inter-comparisons of total column water vapour (TCWV) trend estimates from different CDRs were made and it was concluded that the trend estimates are generally significantly different. It was then shown that several data records disagree with the physical expectation from the Clausius-Clapeyron equation using data over the global ice-free ocean (Schröder et al. 2016, 2019). After homogenisation, a new analysis was applied to the trend estimates and associated results are shown in Figure 7. While the diversity in original trend estimates (-0.15 to +0.12 kg/m^2/year) is several times higher than individual uncertainties, it is largely reduced after homogenisation (-0.02 to +0.04 kg/m^2/year), but still slightly larger than the individual trend uncertainties (up to ± 0.01 kg/m^2/year). As a consequence, after homogenization there was a significant increase in the

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

fraction of datasets that can be seen as consistent as indicated by agreement of trends within twice their combined uncertainties.

Mutual consistency between different quantities

In testing multiple quantity consistency, the role of the underlying background knowledge becomes stronger since the physical processes connecting different ECVs need to be taken into account. One method to test the consistency of two ECVs is by looking at their correlations. For example, in the lower stratosphere, the strong physical dependency of lower stratospheric water vapour on tropical tropopause temperatures can be exploited to test the consistency between climate data records of temperature and stratospheric water vapour as highlighted by Hegglin et al. (2014). This study proposed a new merging method that uses a chemistry-climate model as a transfer function between different satellite instrument records to create a CDR. The methodology allows the bias between instruments to be determined throughout the instrument's lifetime and not only for the overlap period (when old instruments may show first signs of degradation), hence improving characterization of systematic differences (or biases) between datasets. By using the correlation between the newly merged stratospheric water vapour record and the zonal mean temperature from ERA-interim, visual inspection indicated that the new merging method led to physically more consistent results than the traditional one based on bias-correction of instruments during overlap periods. Figure 8 shows the visually well correlated time series of a prototype version of the stratospheric water vapour CDR merged using the new methodology in comparison with zonal mean temperatures at 100 hPa from ERA5 in the tropical region (left panel). We set the threshold for correlations to accept consistency with medium (high) confidence to 0.5 (0.7) since the co-variability of time series of two different variables may also be influenced by other processes which reduce the correlation. In this

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

case (right panel), a correlation of 0.58 suggests that the two variables are physically consistent with medium confidence; if only assessing the last 15 years (not shown) with better data quality, the correlation increases to 0.69 (consistency with high confidence).

Alternatively, differences of multi-year trend maps of one variable can be used to assess the consistency of two different ECV CDRs. The example here is the inter-comparison between wave height (measured by satellite altimetry) and sea ice concentrations assessed in Stopa et al. (2016). Daily sea ice concentrations produced from the Special Sensor Microwave Imager (SSM/I) by IFREMER (Ezraty et al., 2007) are used to define open ocean versus sea ice conditions with a 15% concentration threshold at 12.5 km resolution within the Arctic Ocean. For the period 1992-2014, the SSM/I ice concentrations are used along with wind vectors from the Climate Forecast System Reanalysis to reproduce the wave field through the numerical wave model, WAVEWATCH3 (WW3, Tolman et al., 2014), which includes wave-ice interaction through an under-ice parameterization of wave dissipation. Figure 9 shows a comparison of the trends of the significant wave height (H_s) directly measured with altimetry (denoted ALT, Queffeulou and Croize-Fillon, 2015) and from the co-located model data from WW3 (denoted WW3 CoLoc) in which SSM/I ice concentrations have been used. Qualitatively, the regional patterns match between the two datasets, despite stronger trends in the altimeters (of up to 1 cm / y). At present the confidence interval for trends in wave heights is not known. Therefore the quantitative discrepancies between modeled and measured trends here could be due to both systematic time-varying biases in the wave height ECV, which are expected to be only a function of time and sensor, or to a trend error in the surface wind reanalysis used to drive the wave model. However, the wind trends are also constrained by sea level pressure data and sea ice drift (e.g. Spreen et al. 2011). In the future, a wider range of ECVs, combined with in situ data and models, may be used for a quantitative refinement of sea state trends.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

An example of testing the (anti-)correlation of multiple regional ECV CDRs as predicted by physical theory, is the use of the El Niño Southern Oscillation (ENSO) index for ECV anomalies in the tropical Pacific Ocean Niño3.4 region (5°S-5°N, 190°E-240°E). This natural phenomenon is an ideal candidate for investigating multiple ECV consistency due to its relatively short timescale, large amplitude and multiple ECVs affected by it. This first attempt focusses on the main ENSO signatures at large scale. Physical or biological processes leading to spatio-temporal lags of a few months or a few degrees longitude between some variables have been neglected. This could be refined in future studies. ENSO variability is compared in several ocean (SST, SL, SSS, Chlor_a), atmosphere (CFChigh, TCWV, AOD550) and land (SM, burned area / fire) ECV products - see Table A1 for the acronyms, more detailed information on the datasets and their correlations. All variables were interpolated to a common 1 by 1° grid, de-seasonalised by removing the corresponding monthly mean value and normalised by dividing by the standard deviations for their respective available time period. Figure 10 shows the index variability across the tropical Pacific Ocean for the ECVs in time-longitude anomaly cross-sections. The ocean and most atmosphere ECV time series show consistent spatio-temporal co-variability, as expected. Whereas SST and SL have their largest variability in the Niño3.4 region, CFChigh and TCWV variability peak further west (~180°E), except for the strong El Niño years 1982/83, 1997/98 and 2015/16. Moreover, SSS and Chlor_a are anti-correlated with SST, as expected from a reduced upwelling. For ECVs affected indirectly by El Niño from dry conditions and wild fires over Indonesia (fire, aerosol and soil moisture), the highest correlations occur in their Indonesian time series (10°S-10°N, 100°E-150°E). For certain El Niño3.4 years, e.g. 1997, 2007 and 2015 there are clear indicators of co-variability of them and SST (Fig. 10g). Here again the use of a correlation threshold of 0.5 (0.7) for medium (high) confidence on consistency is adopted. In conclusion, by quantifying (anti-) correlations between these nine independently derived satellite ECVs

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

versus the scientific understanding of the ENSO phenomenon, a medium (high) confidence in their consistency can be shown for eight (four) of them.

4.4 State of consistency assessments for CCI ECVs

Several examples of closure / budget studies of partial Earth system cycles demonstrate the usefulness of CCI (and several other) CDRs that are consistent at all three levels. For example, closure of the carbon budget is still an outstanding scientific challenge (Le Quéré et al. 2018) that is impacted by CDR inconsistencies. Different CCI products provide direct and indirect constraints on carbon fluxes that help to improve the consistency of carbon budgets. For example, CCI greenhouse gas products are used to inform atmospheric inversions. Top down inversion results can be complemented by other ECVs to attribute diagnosed fluxes to different components such as biomass carbon changes (biomass CCI product), fire emissions (CCI products on burned area and fire size) and land use change emissions (land cover CCI products).

Another example is the regional closure of the water budget. Based on multiple satellite ECVs it has been demonstrated that the water budget can be closed within less than 10% uncertainty at a continental annual time scale, while, at monthly time scales, its residuals and uncertainty estimates are larger (about 20%; Rodell et al., 2015). These uncertainties in the water budget closure can be reduced by introducing additional constraints, e.g. by using multiple CDRs with different uncertainties of a single quantity or by additionally forcing closure of the atmosphere and ocean terms. Uncertainties in existing CDRs need to be further reduced and new CDRs of other key variables (most importantly, river discharge and irrigation water use) need to be included or developed to reach the 5% closure error targeted by GCOS (GCOS, 2016).

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

The global mean sea level budget closure has also been assessed within the CCI program by comparing the sum of changes in ocean thermal expansion, land ice melt and liquid water storage on continents with the total observed sea level change. The latter can be estimated globally from satellite altimetry with an accuracy of about 10% on different time scales (e.g. WCRP sea level budget group, 2018). These observations enable closure of the trend in sea level budget with an uncertainty of ±0.3 mm/year over the last 25 years. The sea level budget involves additional variables from the global water budget (through land ice and liquid water components) and from the global energy budget (through thermal expansion directly related to global ocean heat content; Meyssignac et al. 2017) and thus connects the energy and water budgets. At regional scale, uncertainties in the observed components of the sea level budget are considerably larger (few tens of percent) and need to be further reduced to reach the regional GCOS target.

Finally, an assessment of the current state of affairs regarding consistency between the CDRs of the CCI program was made based on the combined scientific expertise of the CCI community; it is not meant to be exhaustive but intended as initial guidance for the use of multiple ECV CDRs or for defining priorities in further consistency analysis. Table 3 provides for each pair of CDRs the consistency status as either: "no evident need to consider consistency", "further studies needed", "consistency explicitly ensured by shared processing or co-retrieving", or "studies already performed", referenced to Table A2 with the underlying publication or technical report (characterized as "theoretical", "exemplary / partial" or "comprehensive"). As can be seen from Table 3, quite some work remains to be done where the definition and concept presented in this paper can be applied and further refined.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

5. Summary and conclusions

Climate Data Records of Essential Climate Variables derived from satellite instruments provide essential information to monitor the state of the Earth system and its changing climate. A key requirement for these CDRs to be useful for Earth system science applications is that the CDRs are internally and mutually consistent. The ESA CCI program provides a set of CDRs for 21 GCOS ECVs in a common framework, and from the outset has invested heavily in establishing their consistency, as presented in this study. To our knowledge no comprehensive definition of CDR consistency exists. Therefore a three-level definition of consistency applicable to single- and multiple-variable cases is proposed and a concept for assessing if two or more CDRs are consistent with each other and possibly with reference data is presented. On the technical level, straightforward data access and usage, including availability of comprehensive documentation and product user guides, is needed. On the retrieval level, one needs to limit contradictions in the use of auxiliary datasets (masks or continuous fields) of the same variables in separate processing chains. On the scientific level, consistency of multiple ECV CDRs means judging their relevant correlations, patterns, periodicity, trends, etc. (as appropriate for a given variable, process or cycle) in the light of underlying physical background knowledge (e.g., by jointly confronting them with a model). Through this link with background knowledge, "consistency" as defined in this study goes beyond "agreement" and relies rather on "compatibility". Finding inconsistencies in one or more ECV dataset(s) (i.e. patterns whose disagreements exceed underlying uncertainties, contradict physical principles or a well-founded model) often indicates errors in a dataset or model whose resolution can lead to new scientific understanding.

This study also provides an overview of the technical consistency of CCI CDRs (common format and metadata standards, common data portal, harmonized documentation, common uncertainty reporting). An open issue in this regard is

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

harmonization across programs and communities. Here, the CCI program has made an important step by adopting the netCDF format, with the CF and ACDD conventions (the de-facto standard in the modelling community) for its gridded satellite data records. The Climate Data Store (CDS) of the Copernicus Climate Change Service (C3S) is also based largely on CCI standards. Such common standards are a prerequisite for the use of automated data services for accessing multiple data sources with little manual interaction, hence facilitating use of the data in scientific studies across multiple ECVs.

The discussion of a concept for assessing consistency and related methods on the retrieval and scientific level shows how consistency with regard to different categorical and continuous auxiliary datasets can be tested and how the assessment of single-variable self-consistency and multiple quantity mutual consistency can be conducted. In all these methods a basic understanding of “the truth” needs to be employed. A relevant characteristic of an ECV and an appropriate metric (e.g. bias, correlation, contingency matrix, …) for its evaluation need to be chosen. A tabular summary of different methods to assess consistency is given in Table 2. For each of the different metrics, a threshold needs to be defined to judge on consistency of two datasets. This may well differ from commonly applied thresholds for validation purposes since also other processes than consistency may affect the datasets. We suggest as a minimum requirement that each consistency study states the applied thresholds, as is done for the examples in this paper. Whereas the methods used to assess consistency rely on well-established tools for calibration and validation, placing them into the systematic context with relevance to consistency as done here, can serve as a practical guideline to consistency assessment. A brief high-level analysis of the inter-dependencies of CCI ECVs at the retrieval and scientific levels (Table 1) is provided to understand where consistency is needed and thus needs to be checked. Finally a high-level

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

assessment of the current state of affairs regarding consistency assessment between the CDRs of the CCI program (Table 3) is compiled to outline possible further research needs.

When discussing consistency, datasets from sources other than satellite data (e.g. Earth system models) are often required to comprehensively study an Earth system cycle, and their uncertainties also need to be considered, together with uncertainties in simplified or estimated budget equations. It is well understood that establishing consistency between two or more variables requires targeted analysis. Within and outside CCI much effort has been spent on quantifying the sensitivities and dependencies of the retrieved quantities. However, a lot more remains to be done in this area.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

6. Acknowledgements

This study is based on ongoing work of altogether 30 projects of the ESA Climate Change Initiative (23 ECV projects, the Climate Model User Group project, cross-cutting outreach components on portal, toolbox, visualisation; CCI data standards and system engineering working group). We are grateful to ESA for creating the CCI program which has strengthened the consistency of the many research communities related to developing, processing, qualifying and using satellite CDRs. We are grateful to the several hundred scientists building the CCI community for making a consistent Earth observation based data repository real. The "operational" part of the CCI program has been transferred to the Copernicus Climate Change Service (C3S, (re-)processing to extend the CDRs, associated quality control, user support). We are also thankful for many other datasets from outside CCI and C3S which help cover all relevant ECVs: GOSAT Level 1 data from JAXA, GOSAT Level 2 data from NIES and NASA, OCO-2 Level 1 and Level 2 data from NASA, HOAPS data from EUMETSAT CM SAF, water vapour records from the G-VAP data archive, CAMS and ERA-5 data from ECMWF / Copernicus Atmosphere and Climate Change Services , SSM/I daily sea ice concentrations from IFREMER, and wind vectors from the Climate Forecast System Reanalysis.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

Appendix

Tables A1 and A2

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

7. References

Allan, R. P., 2012: The Role of Water Vapour in Earth's Energy Flows. Surv. Geophys., 33, 557 – 564, 33:557–564 DOI 10.1007/s10712-011-9157-8.

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

Bates, JJ, Privette, JL, Kearns, EJ; Glance, W, Zhao, XP, 2016: Sustained production of multidecadal climate records Lessons from the NOAA Climate Data Record Program. Bull. Amer. Meteor. Soc., 97, 1573 – 1581, doi/10.1175/BAMS-D-15-00015.1.

Bevan, S.L., North, P.R.J., Los, S.O., Grey, W.M.F, 2012: A global dataset of atmospheric aerosol optical depth and surface reflectance from AATSR. Rem. Sens. Environ., 116, 119–210.

Bojinski, S., M. Verstraete, T.C. Peterson, C. Richter, A. Simmons, and M. Zemp, 2014: The Concept of Essential Climate Variables in Support of Climate Research, Applications, and Policy. Bull. Amer. Meteor. Soc., 95, 1431–1443, https://doi.org/10.1175/BAMS-D-13-00047.1.

Boutin, J.; Vergely, J.-L.; Koehler, J.; Rouffi, F.; Reul, N. (2019): ESA Sea Surface Salinity Climate Change Initiative (Sea_Surface_Salinity_cci): Version 1.8 data collection. Centre for Environmental Data Analysis, 2019, doi:10.5285/9ef0ebf847564c2eabe62cac4899ec41. http://dx.doi.org/10.5285/9ef0ebf847564c2eabe62cac4899ec41

Buchwitz, M., M. Reuter, O. Schneising, H. Boesch, S. Guerlet, B. Dils, I. Aben, R. Armante, P. Bergamaschi, T. Blumenstock, H. Bovensmann, D. Brunner, B. Buchmann, J. P. Burrows, A. Butz, A. Chedin, F. Chevallier, C. D. Crevoisier, N. M. Deutscher, C. Frankenberg, F. Hase, O. P. Hasekamp, J. Heymann, T. Kaminski, A. Laeng, G. Lichtenberg, M. De Maziere, S. Noel, J.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

Notholt, J. Orphal, C. Popp, R. Parker, M. Scholze, R. Sussmann, G. P. Stiller, T. Warneke, C. Zehner, A. Bril, D. Crisp, D. W. T. Griffith, A. Kuze, C. ODell, S. Oshchepkov, V. Sherlock, H. Suto, P. Wennberg, D. Wunch, T. Yokota, Y. Yoshida, 2015: The Greenhouse Gas Climate Change Initiative (GHG-CCI): comparison and quality assessment of near-surface-sensitive satellite-derived CO2 and CH4 global data sets. Rem. Sens. Environ., 162, 344-362, doi:10.1016/j.rse.2013.04.024.

Chuvieco, E., Mouillot, F., van der Werf, G.R., San Miguel, J., Tanasse, M., Koutsias, N., García, M., Yebra, M., Padilla, M., Gitas, I., Heil, A., Hawbaker, T.J., & Giglio, L. (2019). Historical background and current developments for mapping burned area from satellite Earth observation. Remote Sensing of Environment, 225, 45-64.

Doherty, S.J., S. Bojinski, A. Henderson-Sellers, K. Noone, D. Goodrich, N.L. Bindoff, J.A. Church, K.A. Hibbard, T.R. Karl, L. Kajfez-Bogataj, A.H. Lynch, D.E. Parker, I.C. Prentice, V. Ramaswamy, R.W. Saunders, M.S. Smith, K. Steffen, T.F. Stocker, P.W. Thorne, K.E. Trenberth, M.M. Verstraete, and F.W. Zwiers, 2009: Lessons Learned from IPCC AR4: Scientific Developments Needed to Understand, Predict, and Respond to Climate Change. Bull. Amer. Meteor. Soc., 90, 497 – 514, https://doi.org/10.1175/2008BAMS2643.1.

Dorigo, W.A., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., Haas, E., Hamer, P. D., Hirschi, M., Ikonen, J., de Jeu, R., Kidd, R., Lahoz, W., Liu, Y. Y.,Miralles, D., Mistelbauer, T., Nicolai-Shaw, N., Parinussa, R., Pratola, C., Reimer, C., van der Schalie, R., Seneviratne, S. I. Smolander, T., Lecomte, P., 2017: ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. Rem. Sens. Environ., 203, 185 – 215, https://doi.org/10.1016/j.rse.2017.07.001.

ESA, 2019: CCI Data Standards version 2.0. ESA Climate Office, CCI-PRGM-EOPS-TN-13-0009, http://cci.esa.int/sites/default/files/filedepot/CCIDataStandards_v2-0_CCI-PRGM-EOPS-TN-13-0009.pdf

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

Ezraty, R., Girard-Ardhuin, F., Piolle, J. F., Kaleschke, L., and Heygster, G., 2007: Arctic and Antarctic sea ice concentration and Arctic sea ice drift estimated from Special Sensors Microwave data. User manual version 2.1. Ifremer/CERSAT, ftp.ifremer.fr/ifremer/cersat/products/gridded/psi-drift/documentation/ssmi.pdf.

Global Observing System for Climate, 2016: Implementation Needs. WMO, GCOS-200. http://library.wmo.int.

Gruber, A., Scanlon, T., van der Schalie, R., Wagner, W., and Dorigo, W., 2019: Evolution of the ESA CCI Soil Moisture Climate Data Records and their underlying merging methodology. Earth Syst. Sci. Data, 11, 717-739, https://doi.org/10.5194/essd-2019-21.

Hegglin, M. I., D. Plummer, J. Scinocca, T. G. Shepherd, J. Anderson, L. Froidevaux, B. Funke, D. Hurst, A. Rozanov, J. Urban, T. v. Clarmann, K. A. Walker, R. Wang, S. Tegtmeier, K. Weigel, 2014: Variation of stratospheric water vapour trends with altitude from merged satellite data. Nature Geoscience, 7, 768 – 776, doi: 10.1038/NGEO2236.

Hollmann, R., C.J. Merchant, R. Saunders, C. Downy, M. Buchwitz, A. Cazenave, E. Chuvieco, P. Defourny, G. de Leeuw, R. Forsberg, T. Holzer-Popp, F. Paul, S. Sandven, S. Sathyendranath, M. van Roozendael, and W. Wagner, 2013: The ESA Climate Change Initiative: Satellite Data Records for Essential Climate Variables. Bull. Amer. Meteor. Soc., 94, 1541–1552, https://doi.org/10.1175/BAMS-D-11-00254.1.

Holzer-Popp, Th., Bittner, M., Borg, E., Dech, St., Erbertseder, Th., Fichtelmann, B., Schroedter, M., 2002: Das automatische Atmosphärenkorrekturverfahren „DurchBlick“, in: Blaschke, T. (ed.), Fernerkundung und GIS: Neue Sensoren – innovative Methoden, H. Wichmann Verlag, Heidelberg, ISBN 3-87907-369-4, S. 78 – 87.

Immler, F. J., J. Dykema, T. Gardiner, D. N. Whiteman, P. W. Thorne, and H. Vomel, 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.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

IOCCG, 2010: Atmospheric Correction for Remotely-Sensed Ocean-Colour Products.Wang, M. (ed.), Reports of the International Ocean-Colour Coordinating Group, No.10, IOCCG, Dartmouth, Canada.

IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp, doi:10.1017/CBO9781107415324.

IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.

IPCC, 2018: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)]. World Meteorological Organization, Geneva, Switzerland.

IPCC, 2019a: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate [H.-O. Pörtner, D.C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, N.M. Weyer (eds.)]. In press.

IPCC, 2019b: Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

greenhouse gas fluxes in terrestrial ecosystems [P.R. Shukla, J. Skea, E. Calvo Buendia, V. Masson-Delmotte, H.-O. Pörtner, D. C. Roberts, P. Zhai, R. Slade, S. Connors, R. van Diemen, M. Ferrat, E. Haughey, S. Luz, S. Neogi, M. Pathak, J. Petzold, J. Portugal Pereira, P. Vyas, E. Huntley, K. Kissick, M. Belkacemi, J. Malley, (eds.)]. In press.

Kääb, A.; Winsvold, S.H.; Altena, B.; Nuth, C.; Nagler, T.; Wuite, J., 2016: Glacier Remote Sensing Using Sentinel-2. Part I: Radiometric and Geometric Performance, and Application to Ice Velocity. Remote Sens. 8, 598.

Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Hauck, J., Pongratz, J., Pickers, P. A., Korsbakken, J. I., Peters, G. P., Canadell, J. G., Arneth, A., Arora, V. K., Barbero, L., Bastos, A., Bopp, L., Chevallier, F., Chini, L. P., Ciais, P., Doney, S. C., Gkritzalis, T., Goll, D. S., Harris, I., Haverd, V., Hoffman, F. M., Hoppema, M., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A. K., Johannessen, T., Jones, C. D., Kato, E., Keeling, R. F., Goldewijk, K. K., Landschützer, P., Lefèvre, N., Lienert, S., Liu, Z., Lombardozzi, D., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Neill, C., Olsen, A., Ono, T., Patra, P., Peregon, A., Peters, W., Peylin, P., Pfeil, B., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rocher, M., Rödenbeck, C., Schuster, U., Schwinger, J., Séférian, R., Skjelvan, I., Steinhoff, T., Sutton, A., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F. N., van der Laan-Luijkx, I. T., van der Werf, G. R., Viovy, N., Walker, A. P., Wiltshire, A. J., Wright, R., Zaehle, S., Zheng, B., 2018: Global Carbon Budget 2018. Earth Syst. Sci. Data, 10, 2141-2194, https://doi.org/10.5194/essd-10-2141-2018.

Legeais, J.-F., Ablain, M., Zawadzki, L., Zuo, H., Johannessen, J. A., Scharffenberg, M. G., Fenoglio-Marc, L., Fernandes, M. J., Andersen, O. B., Rudenko, S., Cipollini, P., Quartly, G. D., Passaro, M., Cazenave, A., and Benveniste, J., 2018: An improved and homogeneous altimeter sea level record from the ESA Climate Change Initiative. Earth Syst. Sci. Data, 10, 281-301, https://doi.org/10.5194/essd-10-281-2018.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

Levizzani, V. and Cattani, E., 2019: Satellite Remote Sensing of Precipitation and the Terrestrial Water Cycle in a Changing Climate, Remote Sensing, 11(19), 2301.

Li, Z., Zhao, X., Kahn, R. A., Mishchenko, M., Remer, L., Lee, K.-H., Wang, M., Laszlo, I., Nakajima, T., and Maring, H. 2009: Uncertainties in satellite remote sensing of aerosols and impact on monitoring its long-term trend: a review and perspective, Ann. Geophys. 27, 2755–2770.

Lizundia-Loiola, J., Otón, G., Ramo, R., & Chuvieco, E. (2020). A spatio-temporal active-fire clustering approach for global burned area mapping at 250 m from MODIS data. Remote Sensing of Environment, 236, 111493.

Merchant, C. J., Paul, F., Popp, T., Ablain, M., Bontemps, S., Defourny, P., Hollmann, R., Lavergne, T., Laeng, A., de Leeuw, G., Mittaz, J., Poulsen, C., Povey, A. C., Reuter, M., Sathyendranath, S., Sandven, S., Sofeiva, V. F. and Wagner, W., 2017: Uncertainty information in climate data records from Earth observation. Earth Syst. Sci. Data, 9, 511 – 527, https://doi.org/10.5194/essd-9-511-2017.

Merchant, C.J. et al, 2019: Satellite-based time-series of sea-surface temperature since 1981 for climate applications, in preparation for Nature Science Data.

Meyssignac, B., A.B. Slangen, A. Melet, J.A. Church, X. Fettweis, B. Marzeion, C. Agosta, S.R. Ligtenberg, G. Spada, K. Richter, M.D. Palmer, C.D. Roberts, and N. Champollion, 2017: Evaluating Model Simulations of Twentieth-Century Sea-Level Rise. Part II: Regional Sea-Level Changes. J. Climate, 30, 8565–8593, https://doi.org/10.1175/JCLI-D-17-0112.1.

NASA, 2014: OMI/Aura and MODIS/Aqua Merged Cloud Product: https://cmr.earthdata.nasa.gov/search/concepts/C1265734652-GES_DISC.html.

North, P., Briggs, S., Plummer, S. & Settle, J., 1999: Retrieval of land surface bidirectional reflectance and aerosol opacity from ATSR-2 multiangle imagery. IEEE Trans. .Geosci.Rem. Sens., 37, 526 - 537.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

Paul, F. and 24 co-authors, 2015: The Glaciers Climate Change Initiative: Algorithms for creating glacier area, elevation change and velocity products. Rem. Sens. Environ., 162, 408 - 426.

Poulsen, C.A., Siddans, R., Thomas, G.E., Sayer, A.M., Grainger, R.G., Campmany, E., Dead, S.M., Arnold, C.; Watts, P.D., 2012: Cloud retrievals from satellite data using optimal estimation: Evaluation and application to ATSR, Atmos. Meas. Tech., 5, 1889–1910.

Popp, T., 1995: Correcting atmospheric masking to retrieve the spectral albedo of land surfaces from satellite measurements, International Journal of Remote Sensing, 16, 3483-3508.

Popp T., G. de Leeuw, C. Bingen, C. Brühl, V. Capelle, A. Chedin, L. Clarisse, O. Dubovik, R. Grainger, J. Griesfeller, A. Heckel, S. Kinne, L. Klüser, M. Kosmale, P. Kolmonen, L. Lelli, P. Litvinov, L. Mei, P. North, S. Pinnock, A. Povey, C. Robert, M. Schulz, L. Sogacheva, K. Stebel, D. Stein Zweers, G. Thomas, L. G. Tilstra, S. Vandenbussche, P. Veefkind, M. Vountas, Y. Xue, 2016: Development, Production and Evaluation of Aerosol Climate Data Records from European Satellite Observations (Aerosol_cci). Rem. Sens., 8, 421; doi:10.3390/rs8050421.

Pulliainen J., M. Aurela, T. Laurila, T. Aalto, M. Takala, M. Salminen, M. Kulmala, A. Barr, M. Heimann, A. Lindroth, A. Laaksonen, C. Derksen, A. Mäkelä, T. Markkanen, J. Lemmetyinen, J. Susiluoto, S. Dengel, I. Mammarella, J.-P. Tuovinen, T. Vesala, 2017: Early snowmelt significantly enhances boreal springtime carbon uptake. PNAS, 114, 11081-11086, https://doi.org/10.1073/pnas.1707889114.

Quartly, G. D., Legeais, J.-F., Ablain, M., Zawadzki, L., Fernandes, M. J., Rudenko, S., Carrère, L., García, P. N., Cipollini, P., Andersen, O. B., Poisson, J.-C., Mbajon Njiche, S., Cazenave, A., and Benveniste, J., 2017: A new phase in the production of quality-controlled sea level data. Earth Syst. Sci. Data, 9, 557-572, https://doi.org/10.5194/essd-9-557-2017.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

Queffeulou, P. and Croize-Fillon, D., 2015: Global altimeter SWH data set, Technical Report 11.2, IFREMER/CERSAT, ftp://ftp.ifremer.fr/ifremer/cersat/products/swath/altimeters/waves/documentation/altimeter_wave_merge__11.2.pdf.

Reuter, M., H. Boesch, H. Bovensmann, A. Bril, M. Buchwitz, A. Butz, J. P. Burrows, C. W. O'Dell, S. Guerlet, O. Hasekamp, J. Heymann, N. Kikuchi, S. Oshchepkov, R. Parker, S. Pfeifer, O. Schneising, T. Yokota, and Y. Yoshida, 2013: A joint effort to deliver satellite retrieved atmospheric CO_2 concentrations for surface flux inversions: the ensemble median algorithm EMMA. Atmos. Chem. Phys., 13, 1771 - 1780.

Reuter, M., M. Buchwitz, M. Hilker, J. Heymann, H. Bovensmann, J. Burrows, S. Houweling, Y. Liu, R. Nassar, F. Chevallier, P. Ciais, J. Marshall, and M. Reichstein, 2017: How much CO_2 is taken up by the European terrestrial biosphere? Bull. Amer. Meteor. Soc., 98, 665 - 671, doi:10.1175/BAMS-D-15-00310.1.

Reuter, M., Buchwitz, M., Schneising, O., Noel, S., Bovensmann, H., Burrows, J. P., Boesch, H., Di Noia, A., Anand, J., Parker, R. J., Somkuti, P., Wu, L., Hasekamp, O. P., Aben, I., Kuze, A., Suto, H., Shiomi, K., Yoshida, Y., Morino, I., Crisp, D., O'Dell, C., Notholt, J., Petri, C., Warneke, T., Velazco, V., Deutscher, N. M., Griffith, D. W. T., Kivi, R., Pollard, D., Hase, F., Sussmann, R., Te, Y. V., Strong, K., Roche, S., Sha, M. K., De Maziere, M., Feist, D. G., Iraci, L. T., Roehl, C., Retscher, C., and Schepers, D., 2020: Ensemble-based satellite-derived carbon dioxide and methane column-averaged dry-air mole fraction data sets (2003-2018) for carbon and climate applications, Atmos. Meas. Tech., 13, 789-819, https://doi.org/10.5194/amt-13-789-2020.

Rodell, M., H.K. Beaudoing, T.S. L'Ecuyer, W.S. Olson, J.S. Famiglietti, P.R. Houser, R. Adler, M.G. Bosilovich, C.A. Clayson, D. Chambers, E. Clark, E.J. Fetzer, X. Gao, G. Gu, K. Hilburn, G.J. Huffman, D.P. Lettenmaier, W.T. Liu, F.R. Robertson, C.A. Schlosser, J. Sheffield, and

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

E.F. Wood, 2015: The Observed State of the Water Cycle in the Early Twenty-First Century. J. Climate, 28, 8289–8318, https://doi.org/10.1175/JCLI-D-14-00555.1.

Sathyendranath, S, Brewin, RJW, Müeller, D, Brockmann, C, Deschamps, P-Y, Doerffer, R, Fomferra, N, Franz, BA, Grant, MG, Hu C, Krasemann, H, Lee, Z, Maritorena, S, Devred, E, Mélin, F, Peters, M, Smyth, T, Steinmetz, F, Swinton, J, Werdell, J, Regner, P, 2012: Ocean Colour Climate Change Initiative: Approach and Initial Results. IGARSS, 2012, 2024 – 2027, http://dx.doi.org/10.1109/IGARSS.2012.6350979.

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. Appl. Meteor. Clim., 55, 1633 - 1649, doi: /10.1175/JAMC-D-15-0304.1.

Schröder, M., Lockhoff, M., Fell, F., Forsythe, J., Trent, T., Bennartz, R., Borbas, E., Bosilovich, M. G., Castelli, E., Hersbach, H., Kachi, M., Kobayashi, S., Kursinski, E. R., Loyola, D., Mears, C., Preusker, R., Rossow, W. B., and Saha, S.: The GEWEX Water Vapor Assessment archive of water vapour products from satellite observations and reanalyses, Earth Syst. Sci. Data, 10, 1093–1117, https://doi.org/10.5194/essd-10-1093-2018, 2018.

Schröder, M., M. Lockhoff, L. Shi, T. August, R. Bennartz, H. Brogniez, X. Calbet, F. Fell, J. Forsythe, A. Gambacorta, S.-P. Ho, E. R. Kursinski, A. Reale, T. Trent, Q. Yang, 2019: The GEWEX water vapor assessment of global water vapour and temperature data records from satellites and reanalyses. Rem. Sens., 11, 251, https://doi.org/10.3390/rs11030251.

Schulz, J., Albert, P., Behr, H.-D., Caprion, D., Deneke, H., Dewitte, S., Dürr, B., Fuchs, P., Gratzki, A., Hechler, P., Hollmann, R., Johnston, S., Karlsson, K.-G., Manninen, T., Müller, R., Reuter, M., Riihelä, A., Roebeling, R., Selbach, N., Tetzlaff, A., Thomas, W., Werscheck, M., Wolters, E., and Zelenka, A., 2009: Operational climate monitoring from space: the EUMETSAT Satellite Application Facility on Climate Monitoring (CM-SAF). Atmos. Chem. Phys., 9, 1687 - 1709, https://doi.org/10.5194/acp-9-1687-2009.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

Siddans, R., 2016: S5P-NPP Cloud Processor ATBD, EUMETSAT Document number S5P-NPPC-RAL-ATBD-0001, available from https://sentinel.esa.int/documents/247904/2476257/Sentinel-5P-NPP-ATBD-NPP-Clouds.

Sogacheva, L., Kolmonen, P., Virtanen, T. H., Rodriguez, E., Saponaro, G., and de Leeuw, G., 2017: Post-processing to remove residual clouds from aerosol optical depth retrieved using the Advanced Along Track Scanning Radiometer, Atmos. Meas. Tech., 10, 491-505, doi:10.5194/amt-10-491-2017.

Sogacheva Larisa, Thomas Popp, Andrew M. Sayer, Oleg Dubovik, Michael J. Garay, Andreas Heckel, N. Christina Hsu, Hiren Jethva, Ralph A. Kahn, Pekka Kolmonen, Miriam Kosmale, Gerrit de Leeuw, Robert C. Levy, Pavel Litvinov, Alexei Lyapustin, Peter North, Omar Torres and Antti Arola, Merging regional and global aerosol optical depth records from major available satellite products, Atmospheric Chemistry and Physics, 20, 2031 – 2056, https://doi.org/10.5194/acp-20-2031-2020, 2020

Spreen, G., R. Kwok, and D. Menemenlis (2011), Trends in Arctic sea ice drift and role of wind forcing: 1992–2009, Geophys. Res. Lett., 38, L19501, doi:10.1029/2011GL048970.

Stengel M. et al., 2019: Cloud_cci AVHRR-PM dataset version 3: a 35yr spanning climatology of global cloud and radiation properties” in preparation, to be submitted to Earth Syst. Sci. Data

Stopa, J. E., F. Ardhuin, F. Girard-Ardhuin, 2016: Wave climate in the Arctic 1992-2014: seasonality and trends. The Cryosphere, 10, 1605 – 1629.

Tolman, H. L. and the WAVEWATCH III Development Group, 2014: User Manual and system documentation of WAVEWATCH III version 4.18. NOAA, Technical Note 316, NOAA/NWS/NCEP/MMAB.

Vinogradova, N., Lee, T., Boutin, J., Drushka, K., Fournier, S., Sabia, R., Stammer, D., Bayler, E., Reul, N., Gordon, A., Melnichenko, O., Li, L., Hackert, E., Martin, M., Kolodziejczyk, N.,

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

Hasson, A., Brown, S., Misra, S., & Lindstrom, E., 2019: Satellite Salinity Observing System: Recent Discoveries and the Way Forward. Frontiers in Marine Science, 6, 23p. https://doi.org/10.3389/fmars.2019.00243.

Waugh, D. W. and Eyring, V., 2008: Quantitative performance metrics for stratospheric-resolving chemistry-climate models. Atmos. Chem. Phys., 8, 5699 - 5713, https://doi.org/10.5194/acp-8-5699-2008.

WGClimate ECV Inventory Gap Analysis Report V1.1, 2018: available from http://ceos.org/document_management/Working_Groups/WGClimate/Documents/WGClimate_ECV-Inventory_Gap_Analysis_Report_v1.1.pdf.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

Table captions

Table 1: Links between ECVs on the retrieval (above the diagonal) and scientific (below the diagonal) level which need to be consistent if used together. Weak linkages are indicated in brackets. Cycles are indicated with the following acronyms: C=carbon cycle, W=water cycle, E=energy cycle. Processes are indicated with the following acronyms: r=radiation interaction, d=deposition, e=emission / evaporation, t=transport, c=chemical transformation, mtf=melting / thawing / freezing, i=ecosystem interaction, a=air sea fluxes of carbon and water, m=mask.

Table 2: Summary of assessment methods for consistency **on different levels and types**

Table 3: Consistency analysis status between pairs of CCI ECVs: intrinsically assured (*****), study needed (X), study done (c = comprehensive, e = exemplary, t = theoretical) - empty fields indicate that no study is needed, this link cannot be studied (e. g. due to resolution) or the link is considered weak. Numbered references for conducted studies are provided in the appendix (Table A2).

Table A1: Information on the datasets used for figure 10: versions, DOIs and references. The correlations between the SST Niño3.4 region (averaged 5°S to 5°N, 190°E to 240°E) time series and the other ECVS's Niño3.4 time series (and for SM, BA and AOD time series with Indonesia (averaged 10°S to 10°N, 100°E to 150°E) are given in the right column.

Table A2: Snapshot of publications or technical reports (available from ESA CCI program) until the submission of this manuscript behind entries on done consistency studies in Table 3.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

Tables

Table 1: Links between ECVs on the retrieval (above the diagonal) and scientific (below the diagonal) level which need to be consistent if used together. Weak linkages are indicated in brackets. Cycles are indicated with the following acronyms: C=carbon cycle, W=water cycle, E=energy cycle. Processes are indicated with the following acronyms: r=radiation interaction, d=deposition, e=emission / evaporation, t=transport, c=chemical transformation, mtf=melting / thawing / freezing, i=ecosystem interaction, a=air land/sea fluxes of carbon and water, m=mask.

ESA CCI ECVs	Aerosol	Clouds	GHGs	Ozone	Water vapour	Fire	Ice-Sheets	Land cover	Soil moisture	Glaciers	HR land cover	LST	Permafrost	Snow	Biomass	Lakes	Ocean Colour	Sea Ice	Sea Level	SST	Sea State	Sea surface salinity
Direct process dependencies	Retrieval dependencies																					
Aerosol		x	x	(x)	x	x	x	x				x		x		x	x			x		
Clouds	Wr		x	x	x	x	x	x		x	x	x		x		x	x	x		x		
GHGs	e				x									(x)						(x)		
Ozone		t	c		x									(x)		x	x			(x)		
Water vapour	EW	E	C	c		(x)	x					x		(x)		x	x		x	x		
Fire	CE		Ce	ce				x			x		(x)			x						
Ice-Sheets	d			r	W	d		x	x	x									x			
Land cover	de		C e			Cie t			x	x	x	x	x	x		(x)						
Soil moisture	e	E	e		We d	i		i		x	x	x		x	x	x	(x)	(x)	(x)	(x)	(x)	(x)
Glaciers	d					d	W	r			x		x	x		x		x				
HR land cover			Ce			Ct			i	m		x		x								
LST	Er	Er		r	EW r	ECe	Wr	r	Wr	m	r		x	x		x		x		x		
Permafrost		Er	Ce		We	Er	m	Er	Er	m	Er	EW r		x		(x)			(x)			
Snow	d	r		r	We	d	W	ri	mtf	Er m	ri	Wt mtf	Er m		(x)	x			(x)			
Biomass			C			Cc		ic	i			C		i								
Lakes	d e				W	d	Wt	ti	W	E mtf	t	EW r	WE e	W					x	x		
Ocean colour	d e		C	r		d							Cd	m		t		x		x	x	
Sea Ice				r		d						Wr	m				i		x	x	x	(x)
Sea Level					W		W		W	W			W	W		W		W		(x)	x	
SST	Er	Er	r	r	Er	E	mtf					EW t					Er	m	E		(x)	x
Sea State																	i		m			x
Sea surface salinity			C		ea		mtf			mtf			mtf	mtf			CW i	W mtf	WE	Wa	a	

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

Table 2: Summary of assessment methods for consistency **on different levels and types**

	Consistency type	**Required background knowledge**	**Assessment method**
Retrieval level			
	Categorical auxiliary data (“masks”)	Incompatible mask classes	Visual: combined images Contingency matrix Class combination maps
	Continuous auxiliary data	Target variable sensitivity to auxiliary variable	Visual: homogeneity Difference maps Statistical comparison
Scientific level			
	Self-consistency (single quantity)	**Behaviour of one quantity** Known record features Known map features Physical equation	Visual: features as expected Quantitative variability Trend analysis
	Mutual consistency (multiple quantities)	**Linkage between quantities** Physical model Understood Earth system phenomena	Difference maps Trend comparisons Correlations and other measures of co-variability

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

Table 3: Consistency analysis status between pairs of CCI ECVs: intrinsically assured (*), study needed (X), study done (c = comprehensive, e = exemplary, t = theoretical) - empty fields indicate that no study is needed, this link cannot be studied (e. g. due to resolution) or the link is considered weak. Numbered references for conducted studies are provided in the appendix (Table A2).

ESA CCI ECVs	Aerosol	Clouds	GHGs	Ozone	Water vapour	Fire	Ice-Sheets	Land cover	Soil moisture	Glaciers	HR land cover	LST	Permafrost	Snow	Biomass	Lakes	Ocean Colour	Sea Ice	Sea Level	SST	Sea State	Sea surface salinity
	Done / needed consistency analysis (retrieval or scientific level)																					
Aerosol		e [10]	*		X	X	X	X	e [26]			X		X		X	t [11]			X		
Clouds			*			X	X	X	X	X	X	X	X	X		e [24]	X	X		X		
GHGs					*	e [19]		X	X		X		t [1]		X		X			X		
Ozone						X	X					X		X		X	X			X		
Water vapour							X		X			X	t [2]	X		X	X		X	X		
Fire							e [14]	e [23]	e [23]	X	X	t [16,17]	t [3]	e [13]	e [27]	e [15]	X	e [28]		e [18]		
Ice-Sheets								X	X	c [21,22]		X	X	X	X				e [20]	X		
Land cover									X	*	X	X	t [4]	X	X	c [25]						
Soil moisture										X	X	X	t [5]	X	X	*	*	*		*	*	
Glaciers											X	X	e [6]	X		*		*	e [20]			
HR land cover												X	t [7,8]	X		c [25]						
LST													e [4]	X	X	X		X		X		
Permafrost														e [4]		t [9]	X	X	X			
Snow																X	X					
Biomass																						
Lakes																			e [20]	*		
Ocean colour																		X		*	X	X
Sea Ice																			X	*	X	X
Sea Level																					X	X
SST																						X
Sea State																						X
SSS																						

Technical level of consistency achieved

Common data format
Common data access portal
Common metadata standards
Common documentation standards
Common visualisation & analysis tools
(partly) also adopted by C3s Climate Data Store

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-19-0127.1.

Table A1: Information on the datasets used for figure 10: versions, DOIs and references. The correlations between the SST Niño3.4 region (averaged 5°S to 5°N, 190°E to 240°E) time series and the other ECVS's Niño3.4 time series (and for SM, BA and AOD time series with Indonesia (averaged 10°S to 10°N, 100°E to 150°E) are given in the right column.

ECV	**Dataset version, time period used, DOI, references:**	**Correlation of Niño3.4 SST with**
SST	Sea surface temperature ESA SST CCI ATSR and/or AVHRR product version v2.1, 1982-2016 DOI: n/a Merchant et al., 2019	Niño3.4 SST: 1.00
SL	Sea level height SL_cci data v2.0 1993-2015 DOI: 10.5270/esa-sea_level_cci-1993_2015-v_2.0-201612 Legeais et al., 2018 and Quartly et al., 2017	Niño3.4 SL: 0.87
SSS	Sea surface salinity SEASURFACESALINITY_CCI_DATA v1.8 2010-2018 DOI: 10.5285/9ef0ebf847564c2eabe62cac4899ec41 Boutin et al., 2019	Niño3.4 SSS: -0.63
Chlor_a	Chlorophyll-alpha CCI Chlor_a v3.1 (4km_GEO_PML)	Niño3.4Chlor_a: -0.68

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

	1998-2017 DOI: n/a Sathyendranath et al., 2012	
CFChigh	High level cloud fraction Cloud_cci AVHRR-PMv3 1982-2016 DOI: n/a Stengel et al., 2019	Niño3.4 CFChigh: 0.82
TCWV	Total column water vapour HOAPS 4 1988-2015 DOI:10.5676/EUM_SAF_CM/HOAPS/V002 Andersson et al., 2017, data from 2015 as beta version of HOAPS 4	Niño3.4 TCWV: 0.84
AOD550	Aerosol optical depth at 550 nm CCI ATSR-2/AATSR Swansea v4.1 1997-2011 https://esgf-node.llnl.gov/search/obs4mips/obs4mips.SU.ATSR2-AATSR.od550aer.mon.v20160922 |eridanus.eoc.dlr.de Bevan, S., et al., 2012; North, P., et al., 1999; Popp, et al., 2016	Indonesia AOD550: 0.52
Fire	Burned area	Indonesia Fire: 0.49

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

	FireCCI51 2001-2017 DOI: dx.doi.org/10.5285/3628cb2fdba443588155e15dee8e5352 Lizundia et al., 2020	
SM	Soil moisture ESA CCI SM merged v04.5 1991-2018 DOI: n/a Dorigo et al., 2017, Gruber et al., 2019	Indonesia SM: -0.57

1027

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-19-0127.1.

Table A2: Snapshot of publications or technical reports (available from ESA CCI program) until the submission of this manuscript behind entries on done consistency studies in Table 3.

1. Chadburn, S. E., Krinner, G., Porada, P., Bartsch, A., Beer, C., Belelli Marchesini, L., Boike, J., Ekici, A., Elberling, B., Friborg, T., Hugelius, G., Johansson, M., Kuhry, P., Kutzbach, L., Langer, M., Lund, M., Parmentier, F.-J. W., Peng, S., Van Huissteden, K., Wang, T., Westermann, S., Zhu, D., and Burke, E. J. 2017: Carbon stocks and fluxes in the high latitudes: using site-level data to evaluate Earth system models. *Biogeosciences*, **14**, 5143 - 5169, https://doi.org/10.5194/bg-14-5143-2017.
2. Reuter, M., M. Buchwitz, M. Hilker, J. Heymann, H. Bovensmann, J.P. Burrows, S. Houweling, Y.Y. Liu, R. Nassar, F. Chevallier, P. Ciais, J. Marshall, and M. Reichstein, 2017: How Much CO_2 Is Taken Up by the European Terrestrial Biosphere?. *Bull. Amer. Meteor. Soc.*, **98**, 665 – 671, https://doi.org/10.1175/BAMS-D-15-00310.1
3. Carolyn M. Gibson, Laura E. Chasmer, Dan K. Thompson, William L. uinton, Mike D. Flannigan & […]David Olefeldt, 2018: Wildfire as a major driver of recent permafrost thaw in boreal peatlands. *Nature Communications,* **9**, 3041, https://doi.org/10.1038/41467-018-05457-1.
4. Westermann, S., Peter, M., Langer, M., Schwamborn, G., Schirrmeister, L., Etzelmüller, B., Boike, J., 2017: Transient modeling of the ground thermal conditions using satellite data in the Lena River Delta, Siberia. *The Cryosphere*, **11**, 1441 - 1463, doi:10.5194/tc-11-1441-2017.
5. Dafflon, B., Oktem, R., Peterson, J., Ulrich, C., Tran, A. P., Romanovsky, V., & Hubbard, S. S., 2017: Coincident aboveground and belowground autonomous monitoring to quantify covariability in permafrost, soil, and vegetation properties in Arctic tundra.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

Journal of Geophysical Research: Biogeosciences, **122**, 1321 – 1342, https://doi.org/10.1002/2016JG003724.

6. Daiyrov M., C. Narama, T. Yamanokuchi, T. Tadono, A. Kääb, J. Ukita T., 2018: Regional geomorphological conditions related to recent changes of glacial lakes in the Issyk-Kul basin, northern Tien Shan. *Geosciences*, **8**, 99.
7. Cable, W.L.; Romanovsky, V.E.; Jorgenson, M.T., 2016: Scaling-up permafrost thermal measurements in western Alaska using an ecotype approach. *Cryosphere*, **10**, 2517 – 2532.
8. Zhang, N., Yasunari, T., & Ohta, T., 2011: Dynamics of the larch taiga–permafrost coupled system in siberia under climate change. *Environmental Research Letters*, **6**, 024–003, https://doi.org/10.1088/1748-9326/6/2/024003.
9. Nitze, I., G. Grosse, B. M. Jones, V. E. Romanovsky and J. Boike, 2018: Remote sensing quantifies widespread abundance of permafrost region disturbances across the Arctic and Subarctic. *Nature Communications*, **9**, 5423.
10. Klüser L., S. Stapelberg, Aerosol_cci Cloud_cci cloud mask consistency report v1.1. DLR / DWD / ESA (briefly summarized in the third example of section 4 in this paper).
11. Stebel, Kerstin, et al., 2017: Aerosol_cci2 Technical Note on consistency v1.0. DLR / ESA.
12. Adolf, C., Wunderle, S., Colombaroli, D., Weber, H., Gobet, E., Heiri, O., van Leeuwen, J. F. N., Bigler, C., Connor, S. E., Gałka, M., La Mantia, T., Makhortykh, S., Svitavská-Svobodová, H., Vannière, B., and Tinner, W., 2018: The sedimentary and remote-sensing reflection of biomass burning in Europe. *Global Ecology and Biogeography* **27**, 199–212. doi: 10.1111/geb.12682.
13. Cape, J., Coyle, M., and Dumitrean, P., 2012: The atmospheric lifetime of black carbon. *Atmos. Environ.*, **59**, 256 – 263, doi: 10.1016/j.atmosenv. 2012.05.030.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

14. Eichler, A., Tinner, W., Brütsch, S., Olivier, S., Papina, T., and Schwikowski, M., 2011: An ice-core based history of Siberian forest fires since AD 1250. Q*uaternary Science Reviews*, **30**, 1027 – 1034. doi: 10.1016/j.quascirev.2011.

15. Marlon, J. R., Kelly, R., Daniau, A.-L., Vannière, B., Power, M. J., Bartlein, P. J., Higuera, P. E., Blarquez, O., Brewer, S., Brücher, T., Feurdean, A., Romera, G. G., Iglesias, V., Maezumi, S. Y., Magi, B., Courtney Mustaphi, C. J., and Zhihai, T., 2016: Reconstructions of biomass burning from sediment-charcoal records to improve data–model comparisons. *Biogeosciences*, **13**, 3225 – 3244, doi: 10.5194/bg-13-3225-2016.

16. Pechony, O. and Shindell, D. T., 2010: Driving forces of global wildfires over the past millennium and the forthcoming century. Proc. of the National Academy of Sciences 107.45, 19167–19170, doi: 10.1073/pnas.1003669107.

17. Chuvieco, E., Mouillot, F., van der Werf, G.R., San Miguel, J., Tanasse, M., Koutsias, N., García, M., Yebra, M., Padilla, M., Gitas, I., Heil, A., Hawbaker, T.J., & Giglio, L., 2019: Historical background and current developments for mapping burned area from satellite Earth observation. *Rem. Sens. Environ.*, **225**, 45 - 64.

18. Chen, Y., Morton, D.C., Andela, N., Giglio, L., & Randerson, J.T., 2016: How much global burned area can be forecast on seasonal time scales using sea surface temperatures? *Environmental Research Letters,* **11**, 045001.

19. Heymann, J., M. Reuter, M. Buchwitz, O. Schneising, H. Bovensmann, J. P. Burrows, S. Massart, J. W. Kaiser, D. Crisp, 2017: CO2 emission of Indonesian fires in 2015 estimated from satellite-derived atmospheric CO2 concentrations, *Geophys. Res. Lett.*, **44**, 1537 – 1544, DOI: 10.1002/2016GL072042, pp. 18, 2017.

20. Cazenave, A., 2018: Global sea-level budget 1993-present. *Earth System Science Data*, **10**, 1551 - 1590, https://doi.org/10.5194/essd-10-1551-2018.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

21. Rastner, P., T. Bolch, N. Mölg, H. Machguth, R. Le Bris and F. Paul, 2012: The first complete inventory of the local glaciers and ice caps on Greenland. *The Cryosphere*, **6**, 1483 - 1495.
22. Huber, J., A. Cook, F. Paul, and M. Zemp, 2017: A complete glacier inventory of the Antarctic Peninsula based on Landsat7 images from 2000-2002 and other pre-existing datasets. *Earth Systems Science Data*, **9**, 115 - 131.
23. Forkel, M., Dorigo, W., Lasslop, G., Teubner, I., Chuvieco, E., and Thonicke, K. 2017: A data-driven approach to identify controls on global fire activity from satellite and climate observations (SOFIA V1). *Geosci. Model Dev.*, **10**, 4443 - 4476, https://doi.org/10.5194/gmd-10-4443-2017.
24. Brockmann, C., et al., 2013: Multi-Sensor Cloud Screening and Validation: IdePix and PixBox. *Proc. of the 2013 European Space Agency Living Planet Symposium. http://livingplanet2013. org/abstracts/850821.*
25. Pekel, J.-F., et al., 2016: High-resolution mapping of global surface water and its long-term changes. *Nature* **540**, 418.
26. Klingmüller, K., Pozzer, A., Metzger, S., Stenchikov, G. L., and Lelieveld, J., 2016: Aerosol optical depth trend over the Middle East, *Atmos. Chem. Phys.*, **16**, 5063 - 5073, https://doi.org/10.5194/acp-16-5063-2016.
27. Frolking, S., Palace, M.W., Clark, D., Chambers, J.Q., Shugart, H., & Hurtt, G.C., 2009: Forest disturbance and recovery: A general review in the context of spaceborne remote sensing of impacts on aboveground biomass and canopy structure. *J. Geophys. Res.* Biogeosciences, **114**.
28. Law, K.S., & Stohl, A., 2007: Arctic air pollution: Origins and impacts. *Science*, **315**, 1537 - 1540.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

Figure captions

Figure 1: Temporal coverage of CDRs for ECVs analysed by CCI. Filled bars indicate CDRs available in 2019, outlined bars CDRs that are planned within the ongoing phase of the CCI program.

Figure 2: The ECVs covered by ESA CCI CDRs, ordered according to the key Earth system cycle (energy, carbon, water) they help characterise. The cycles are inter-linked, and most water and carbon cycle ECVs are also relevant to the energy cycle, since energy is stored and transported in water and matter, at least on transient timescales.

Figure 3: Gaps in surface temperature fields (LST and SST from SLSTR on Sentinel-3A on 05/08/2018 at 10:38 UTC) due to masked clouds (grey), showing the absence of scatter at land-sea borders and sampling discontinuities across some land-sea boundaries due to different cloud-clearing approaches between LST and SST processing.

Figure 4: Consistency overview between Aerosol_cci (Swansea University) and Cloud_cci (FAME-C) AATSR cloud masks for observations of four selected days in September 2008. No cloud/no cloud and cloud/cloud situations are solely analysed as aerosol or clouds in Aerosol_cci and Cloud_cci, respectively. No cloud/cloud situations are wrongly analysed as aerosols and clouds, while cloud/no cloud situations are not analysed at all.

Figure 5: Mean AOD differences at 865 nm between ocean colour MERIS atmospheric correction by-product and aerosol ECV product from AATSR in May 2003 when both instruments retrieve AOD.

Figure 6: Top: Time series of monthly mean northern mid-latitude XCO_2 (red thick line) based on merging individual XCO_2 ensemble members (black lines) from GOSAT (since 2009) and OCO-2 (since 2014). The time series (2003-2018) begins with one XCO_2 product from SCIAMACHY/ENVISAT. **Bottom:** XCO_2 difference between ensemble members (black lines)

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

and the multi-sensor / multi-algorithm merged product (red line in top panel). Details see Reuter et al., 2020.

Figure 7: Trend estimates computed after (green) and before (black) homogenisation for all long-term TCWV data records available from the G-VAP data archive (Schröder et al., 2018). Trend estimates are sorted in ascending order without homogenisation. The grey horizontal line marks a trend of 0 kg/m^2/year (updated from Schröder et al., 2019).

Figure 8: The left panel shows the co-variation between a prototype version of the stratospheric water vapour CDR H_2O (produced within the Water_Vapour_cci) and ERA5 monthly zonal mean temperatures T at 100 hPa. The right panel shows the correlation between the two datasets.

Figure 9: Trends of monthly averaged significant wave height H_s data sets with the Mann–Kendall test (thatched areas) from satellite altimetry (left: ALT), and co-located model WW3 hindcast (right: CoLoc) both given in cm year^{-1}.

Figure 10: Zonal month-longitude cross sections (averaged 5°S and 5°N) for 150°E to 280°E normalized indices of a) sea surface temperature (SST), b) sea level height (SL), c) Sea Surface Salinity (SSS), d) chlorophyll-alpha (Chlor_a), e) high level cloud fraction (CFChigh), f) total column water vapour (TCWV). All ECVs are plotted for their respective full year availability. The black lines in the Hovmöller plots show the Niño3.4 box. g) Time series of Niño3.4 SST and Indonesia soil moisture (SM), burned area (Fire), and aerosol optical depth at 550 nm (AOD550). Information on the used datasets is provided in Table A1 in the Appendix.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

Figures

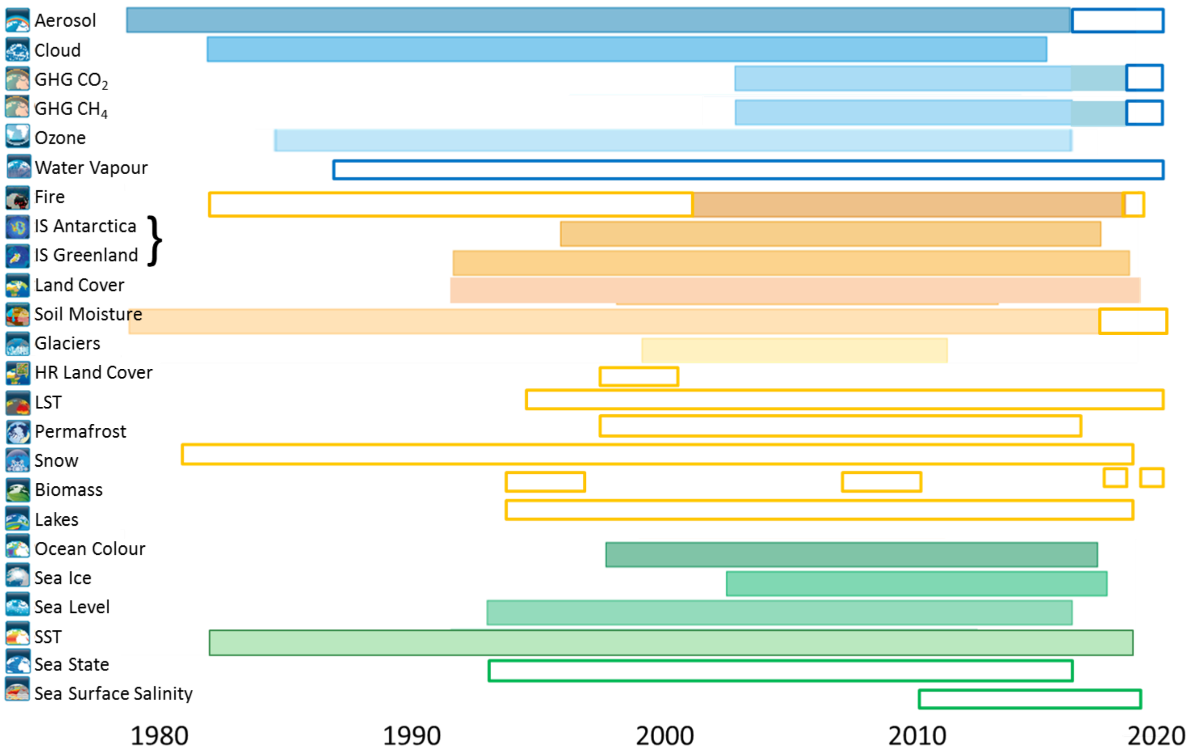


Figure 1: Temporal coverage of CDRs for ECVs analysed by CCI. Filled bars indicate CDRs available in 2019, outlined bars CDRs that are planned within the ongoing phase of the CCI program.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

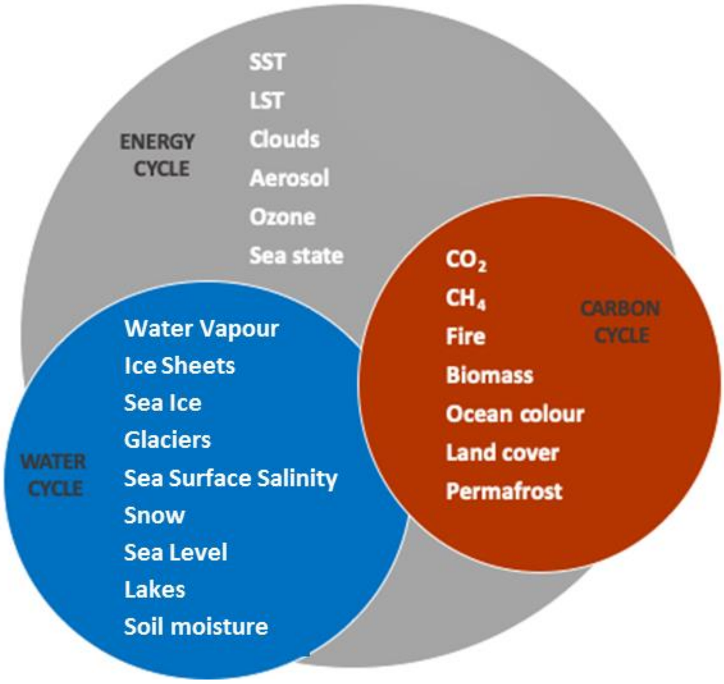


Figure 2: The ECVs covered by ESA CCI CDRs, ordered according to the key Earth system cycle (energy, carbon, water) they help characterise. The cycles are inter-linked, and most water and carbon cycle ECVs are also relevant to the energy cycle, since energy is stored and transported in water and matter, at least on transient timescales.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

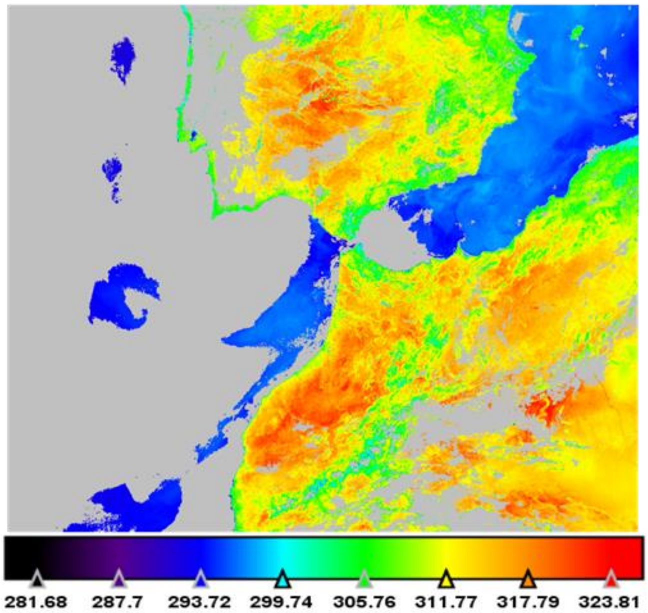


Figure 3: Gaps in surface temperature fields (LST and SST from SLSTR on Sentinel-3A on 05/08/2018 at 10:38 UTC) due to masked clouds (grey), showing the absence of scatter at land-sea borders and sampling discontinuities across some land-sea boundaries due to different cloud-clearing approaches between LST and SST processing.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

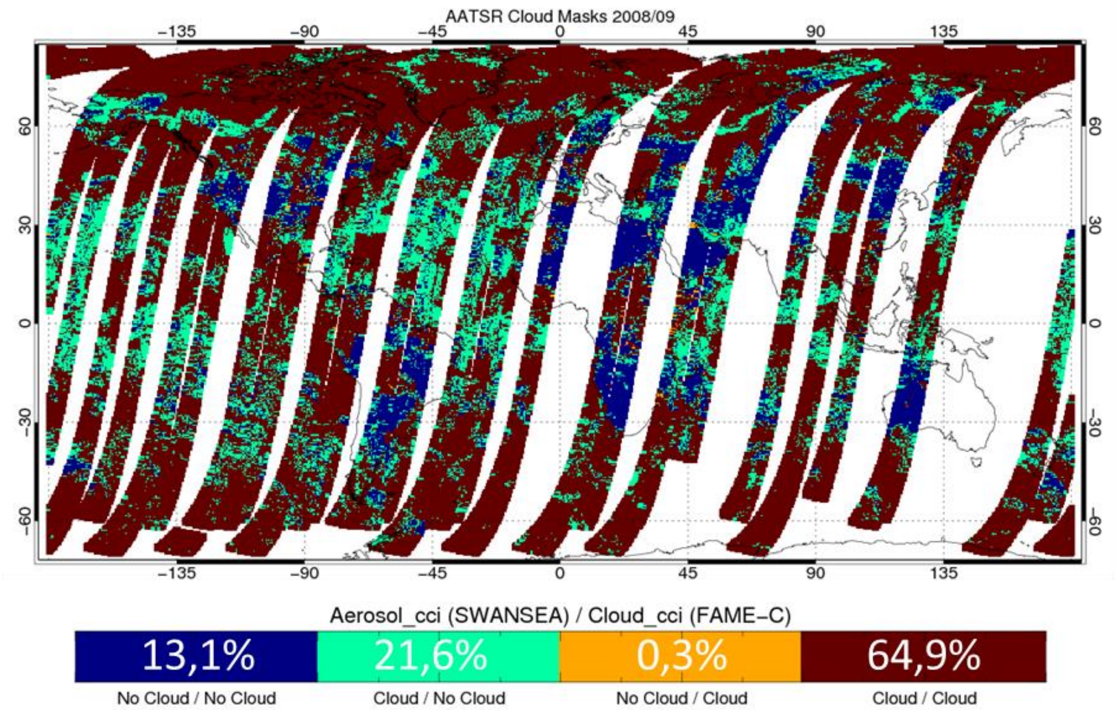


Figure 4: Consistency overview between Aerosol_cci (Swansea University) and Cloud_cci (FAME-C) AATSR cloud masks for observations of four selected days in September 2008. No cloud/no cloud and cloud/cloud situations are solely analysed as aerosol or clouds in Aerosol_cci and Cloud_cci, respectively. No cloud/cloud situations are wrongly analysed as aerosols and clouds, while cloud/no cloud situations are not analysed at all.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

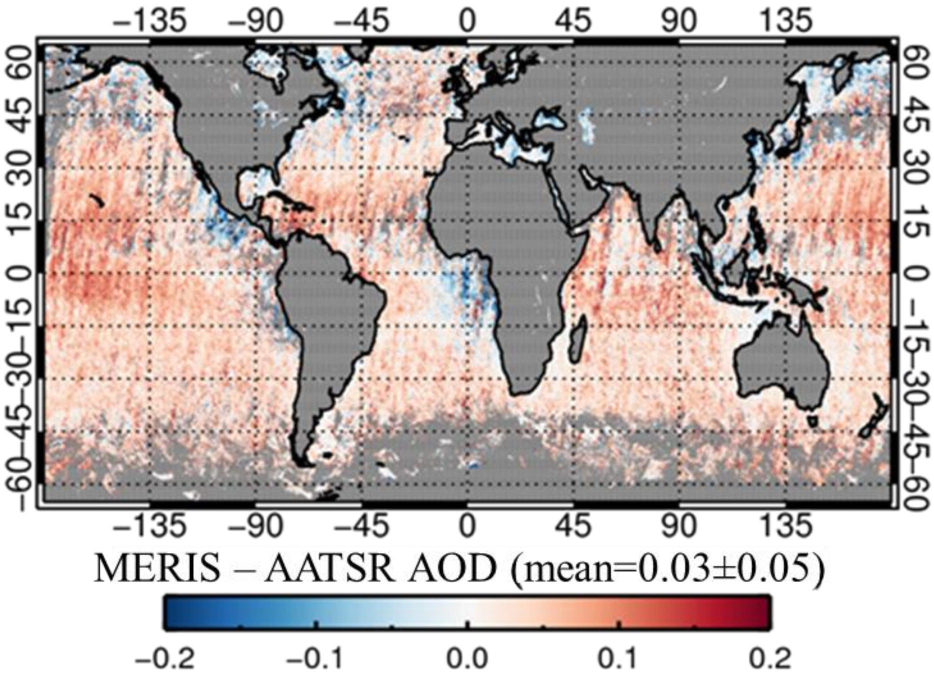


Figure 5: Mean AOD differences at 865 nm between ocean colour MERIS atmospheric correction by-product and aerosol ECV product from AATSR in May 2003 when both instruments retrieve AOD.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

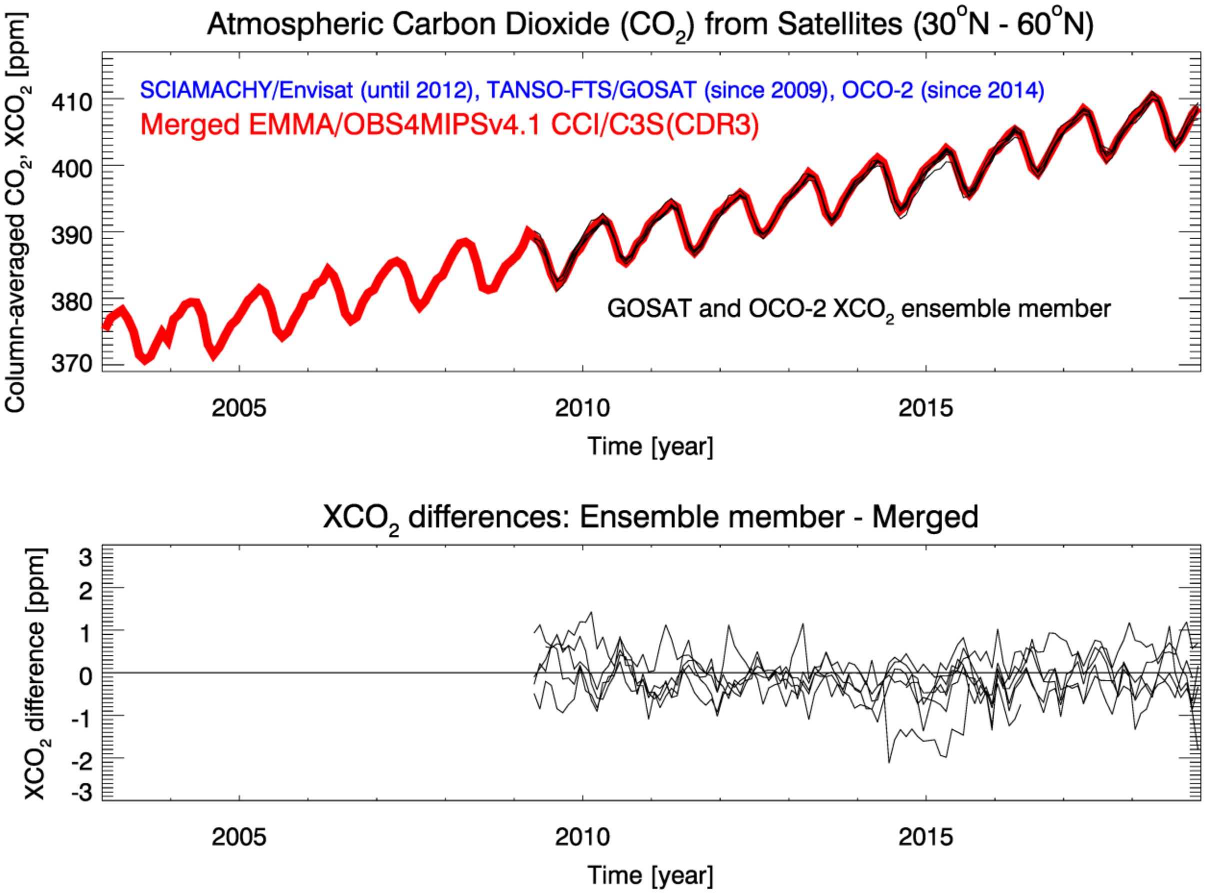


Figure 6: Top: Time series of monthly mean northern mid-latitude XCO_2 (red thick line) based on merging individual XCO_2 ensemble members (black lines) from GOSAT (since 2009) and OCO-2 (since 2014). The time series (2003-2018) begins with one XCO_2 product from SCIAMACHY/ENVISAT. **Bottom:** XCO_2 difference between ensemble members (black lines) and the multi-sensor / multi-algorithm merged product (red line in top panel). Details see Reuter et al., 2020.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

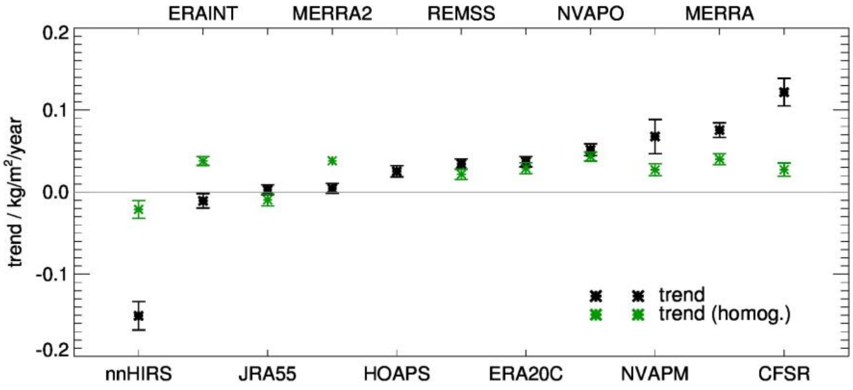


Figure 7: Trend estimates computed after (green) and before (black) homogenisation for all long-term TCWV data records available from the G-VAP data archive (Schröder et al., 2018). Trend estimates are sorted in ascending order without homogenisation. The grey horizontal line marks a trend of 0 kg/m^2/year (updated from Schröder et al., 2019).

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

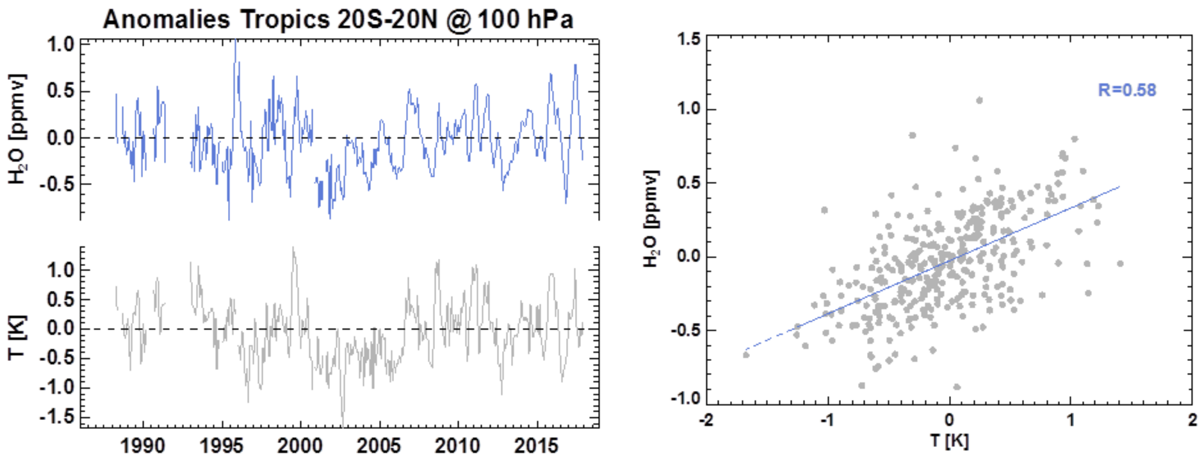
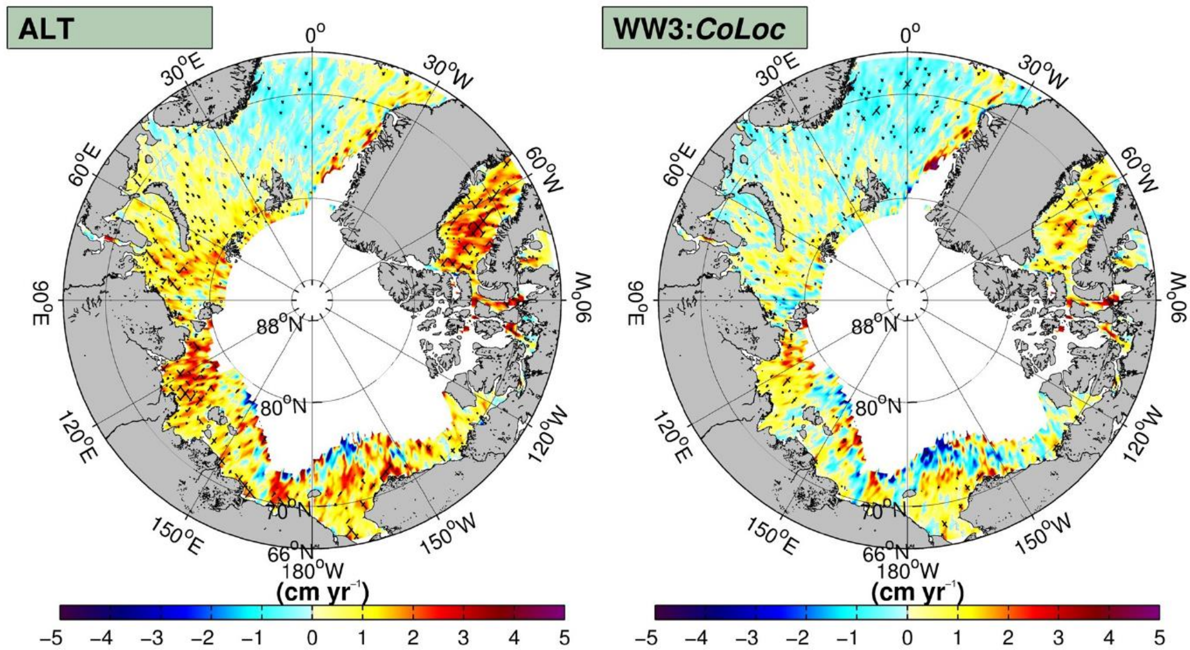


Figure 8: The left panel shows the co-variation between a prototype version of the stratospheric water vapour CDR H_2O (produced within the Water_Vapour_cci) and ERA5 monthly zonal mean temperatures T at 100 hPa. The right panel shows the correlation between the two datasets.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

1215



1216 **Figure 9:** Trends of monthly averaged significant wave height H_s data sets with the Mann–
1217 Kendall test (thatched areas) from satellite altimetry (left: ALT), and co-located model WW3
1218 hindcast (right: CoLoc) both given in cm year^{-1}.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.

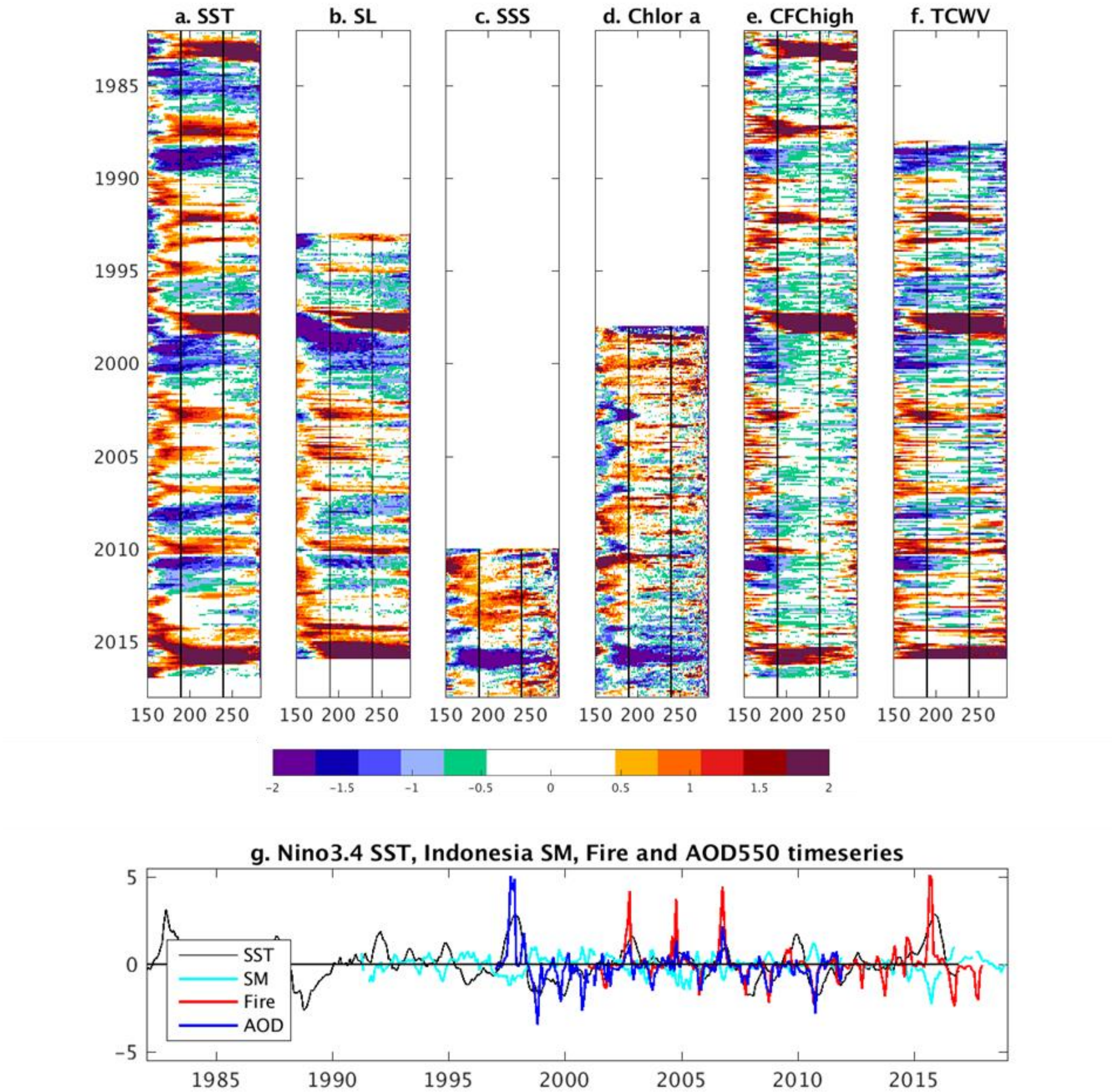


Figure 10: Zonal month-longitude cross sections (averaged 5°S and 5°N) for 150°E to 280°E normalized indices of a) sea surface temperature (SST), b) sea level height (SL), c) Sea Surface Salinity (SSS, d) chlorophyll-alpha (Chlor_a), e) high level cloud fraction (CFChigh), f) total column water vapour (TCWV). All ECVs are plotted for their respective full year availability. The black lines in the Hovmöller plots show the Niño3.4 box. g) Time series of Niño3.4 SST and Indonesia soil moisture (SM), burned area (Fire), and aerosol optical depth at 550 nm (AOD550). Information on the used datasets is provided in Table A1 in the Appendix.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-19-0127.1.