

Article

# Ten Priority Science Gaps in Assessing Climate Data Record Quality

Joanne Nightingale <sup>1,\*</sup>, Jonathan P.D. Mittaz <sup>2</sup>, Sarah Douglas <sup>1</sup>, Dick Dee <sup>3</sup>, James Ryder <sup>1</sup>, Michael Taylor <sup>2</sup> , Christopher Old <sup>2,4</sup>, Catherine Dieval <sup>2,5</sup>, Celine Fouron <sup>6</sup>, Guillaume Duveau <sup>6</sup> and Christopher Merchant <sup>7</sup> 

<sup>1</sup> Earth Observation, Climate and Optical Group, National Physical Laboratory, Hampton Road, Teddington TW11 0LW, UK; Sarah.Douglas@npl.co.uk (S.D.); James.Ryder@npl.co.uk (J.R.)

<sup>2</sup> Department of Meteorology, University of Reading, PO Box 243, Whiteknights, Reading RG6 6BB, UK; j.mittaz@reading.ac.uk (J.P.D.M.); michael.taylor@reading.ac.uk (M.T.); c.old@ed.ac.uk (C.O.); dieval@gfz-potsdam.de (C.D.)

<sup>3</sup> Copernicus Climate Change Service, ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, UK; Dick.Dee@ecmwf.int

<sup>4</sup> Institute for Energy Systems, School of Engineering, University of Edinburgh, Grant Institute Kings Buildings, W Mains Road, Edinburgh EH9 3JW, UK

<sup>5</sup> Deutsches GeoForschungsZentrum, Albert-Einstein-Strasse 42-46, 14473 Potsdam, Germany

<sup>6</sup> Telespazio France, 26 Avenue Jean François Champollion, 31100 Toulouse, France; celine.fouron@telespazio.com (C.F.); guillaume.duveau@gmail.com (G.D.)

<sup>7</sup> National Centre for Earth Observation, Reading RG6 6AL, UK; c.j.merchant@reading.ac.uk

\* Correspondence: Joanne.Nightingale@npl.co.uk; Tel.: +44-777-554-275

Received: 11 March 2019; Accepted: 18 April 2019; Published: 25 April 2019



**Abstract:** Decision makers need accessible robust evidence to introduce new policies to mitigate and adapt to climate change. There is an increasing amount of environmental information available to policy makers concerning observations and trends relating to the climate. However, this data is hosted across a multitude of websites often with inconsistent metadata and sparse information relating to the quality, accuracy and validity of the data. Subsequently, the task of comparing datasets to decide which is the most appropriate for a certain purpose is very complex and often infeasible. In support of the European Union’s Copernicus Climate Change Service (C3S) mission to provide authoritative information about the past, present and future climate in Europe and the rest of the world, each dataset to be provided through this service must undergo an evaluation of its climate relevance and scientific quality to help with data comparisons. This paper presents the framework for Evaluation and Quality Control (EQC) of climate data products derived from satellite and in situ observations to be catalogued within the C3S Climate Data Store (CDS). The EQC framework will be implemented by C3S as part of their operational quality assurance programme. It builds on past and present international investment in Quality Assurance for Earth Observation initiatives, extensive user requirements gathering exercises, as well as a broad evaluation of over 250 data products and a more in-depth evaluation of a selection of 24 individual data products derived from satellite and in situ observations across the land, ocean and atmosphere Essential Climate Variable (ECV) domains. A prototype Content Management System (CMS) to facilitate the process of collating, evaluating and presenting the quality aspects and status of each data product to data users is also described. The development of the EQC framework has highlighted cross-domain as well as ECV specific science knowledge gaps in relation to addressing the quality of climate data sets derived from satellite and in situ observations. We discuss 10 common priority science knowledge gaps that will require further research investment to ensure all quality aspects of climate data sets can be ascertained and provide users with the range of information necessary to confidently select relevant products for their specific application.

**Keywords:** Copernicus Climate Change Service (C3S); evaluation and quality control (EQC); climate data store (CDS); essential climate variable (ECV); climate data record (CDR); quality assurance (QA); satellite observations; in situ observations; traceability

## 1. Introduction

Long term observations of Earth system variables from Earth Observation (EO) satellites and in situ observation networks are essential for providing the foundation and scientific knowledge with which to understand the variability of natural and anthropogenic processes and to help mitigate and adapt to environmental and climate change. The Paris Agreement from 2015 [1], aiming at strengthening the global response to the threat of climate change, has requested systematic observation of the climate system for this purpose. Effective data content management systems are required to capitalise on the multitude of currently available and anticipated global climate data streams. These should facilitate data processing, integration and visualisation capabilities to support interpretation as well as the development of workflows for application-support services [2]. Further to the basic provision of these data streams there is a critical need for comprehensive metadata on the quality of the data to enable users to judge the fitness for use and ensure confidence in the information used to support decision making processes. A rigorous quantification of the accuracy and validity of climate information from EO satellite and in situ observations is fundamental to the scientific understanding of the Earth system and its response to change and progress in policymaking. Furthermore, attention must be paid to how this quality information is provided to data users [3–6].

The Copernicus Climate Change Service (C3S), implemented by the European Centre for Medium-Range Weather Forecasts (ECMWF) is one of six operational environmental information services established by the European Commission (EC) within the Copernicus Earth Observation Programme. C3S supports climate change adaptation and mitigation in Europe by ensuring reliable access to high-quality data on past, present and future climate, and by enabling users to make effective use of this data, e.g., for monitoring climate change and its impacts, for developing climate services in various industrial sectors, and for policy development and implementation. The backbone of C3S is a cloud-based Climate Data Store (CDS, [2,7]), which aims to make it easier for users with varying backgrounds to access complex climate datasets and turn them into useful information products. The CDS represents a single point of access to a catalogue of climate datasets including Essential Climate Variable (ECV, [8]) products derived from observations, model-based climate reanalyses, seasonal forecast data products, and climate model simulations including projections. In a progressive commitment to ensure that all datasets available through the C3S CDS are traceable, adequately documented and accompanied by quality information so that data users can make informed decisions for their application, C3S has made significant investments in the ongoing development of Evaluation and Quality Control (EQC) functionality. The purpose of the EQC is to collate and display quality assurance (QA) evidence on each of the individual data sets in a standardized manner allowing impartial evaluation by users, as well as a means for monitoring the status of and improving the climate data service information through time. The QA information collected will be prominently displayed in an accessible format on the CDS webpages associated with each data set in the catalogue.

As discussed in [3,9], regulatory frameworks requiring EO data and product producers to be held accountable for ensuring the quality, accuracy and validity of the information provided do not currently exist nor do the standards against which data quality should be monitored. This is true across all data providers from satellite derived data to in situ observation network information and model-based data products. Implementation of a common EQC framework suitable for data sets derived from satellite and in situ observations is therefore very timely. This paper outlines the scope, development and functionality of the EQC framework and its content management system to be implemented within the CDS. Additionally, we discuss key science gaps related to addressing and understanding the quality of

Climate Data Records (CDRs) derived from satellite and in situ observations. These science gaps will be discussed in relation to their scientific importance, as well as priorities for action and investment to ensure that, in the long term, all quality aspects of climate data sets can be ascertained and that the EQC will provide users with all information necessary to confidently select relevant products for their specific application.

## 2. User Requirements and QA Information

### 2.1. User Survey and Results

Extensive user consultation to gauge the current state of and need for quality assurance in climate data products derived from observations has been undertaken in several EU funded projects including, but not limited to; QA4ECV (Quality Assurance for Essential Climate Variables [10]); GAIA-CLIM (Gap Analysis for Integrated Atmospheric ECV CLImate Monitoring [11]); FIDUCEO (FIDelity and Uncertainty in Climate data records from Earth Observations [12]); CLIM-RUN (Climate Local Information in the Mediterranean region Responding to User Needs [13]); EUPORIAS (European Provision of Regional Impacts Assessments on Seasonal and Decadal Timescales [14]); CLIP-C (Climate Information Portal [15]); GlobTemp (GlobTemperature [16]); and CORE-CLIMAX (COordinating Earth observation data validation for RE-analysis for CLIMATE ServiceS) [17]. The findings from these projects were used to develop a targeted survey which included practical examples of the provision and use of Quality Indicators (QI's) such as basic metadata, documentation and traceability, validation and inter-comparison as well as algorithms and uncertainties, that should be provided in satellite and in situ observation derived CDRs. The survey was sent to a total of 582 potential users across Europe covering a range of sectors: agriculture and forestry, energy, health, infrastructure, transport, insurance, tourism, water management and coastal areas. A total of 80 complete responses to the survey were obtained and a further 20 users were interviewed in person.

Overall feedback was positive in the sense that users would take advantage of quality information if provided and detailed guidance documentation on how to handle the quality information appropriately was made available. There was a clear need for quality metrics of varying levels of detail and complexity to be provided for different levels of user. The key findings were like those collated in the complementary survey activities of other projects and demonstrated that:

1. There is a strong need for consolidated, short, simple guidance documents about the products, their quality metrics and how to interpret the quality metrics;
2. All documentation should be easily accessible and frequently updated to contain the most current information;
3. Traceability chains, developed as part of the QA4ECV project [3], were highly regarded by all users because they enable a quick and relatively complete understanding of the product algorithm;
4. Evidence that the product has been independently validated is key criteria for most data users;
5. Inter-comparison results are well used and considered very important in understanding the advantages and disadvantages of the data products relative to each other;
6. Access to maps and statistics at in situ measurement sites are considered very useful since access to this information can help users identify causes of discrepancies in data and understand typical seasonality in a location;
7. Most data users use pixel level quality flag information or would use them if provided;
8. Known issues or problems registers for data products were requested to allow users to understand the consistency of the product over time; and
9. Use cases and reasons for data products being produced are highly desired.

The valuable feedback from survey respondents helped shape the EQC Quality Assurance Templates (QATs) that are used to evaluate the scientific integrity and quality of ECV data acquired from both satellite and in situ observations.

## 2.2. ECV Product Inventory

For a demonstrator set of nine ECVs including: Precipitation; Surface Air Temperature; Leaf Area Index (LAI); fraction of absorbed photosynthetically active radiation (fAPAR); Soil Moisture; Sea Surface Temperature (SST); Ocean Colour; Ozone and Aerosols, an exhaustive inventory of observational data sets (both satellite and in situ) was compiled. Over 500 individual data products were found (Table 1). To be considered climate relevant to the C3S in this study, the datasets listed in the inventory had to be global, freely available, operationally produced over a long temporal period (shorter if recently funded for CDR development) and known to be used by the scientific community. Over 250 individual products across the nine ECVs were considered contenders for potential inclusion in the CDS catalogue (Table 1). An initial Quality Assurance Template (QAT) checklist was derived based on previous EU and internationally funded initiatives as well as the user requirements gathering process, to capture the status of several QI's for each of the climate relevant datasets. The initial set of QI's included information about documentation, product generation, quality flags, uncertainty characterisation, validation and inter-comparison [3]. A top-level evaluation of the QI check list for each of the 250 datasets revealed that most had some sort of documentation about the algorithm development and associated user guide, but many had little in the way of detailed quality assessment or uncertainty characterisation. Furthermore, the presentation of quality information between and within ECV product families was inconsistent, which ultimately hindered the ability to make a sound judgement on the overall quality of each data product.

**Table 1.** Total number of data products found for each of the nine demonstrator Essential Climate Variables (ECVs) and number of products considered to be climate relevant after filtering.

ECV	Total Number of Products Found	Number of Products after Climate Relevance Filtering
Precipitation (in situ) <sup>1</sup>	53	14
Surface Air Temperature (in situ) <sup>1</sup>	72	27
LAI	33	21
fAPAR	30	22
Soil Moisture	25	14
SST	72	27
Ocean Colour	40	28
Ozone	105	78
Aerosols	100	68

<sup>1</sup> Precipitation and Surface Air Temperature products were all derived from in situ observations, while the remaining ECV products were derived from satellite observations.

To develop a robust EQC process that would capture and enable standardisation of user relevant product quality information, a detailed scientific and gap analysis of a more manageable selection of key products for each of the eight demonstrator ECVs was necessary. Further filtering of the 250 products to approximately five key products per ECV was conducted by considering the available QI's for each of the individual products, along with additional criteria to ensure a mix of data product scenarios including: products that were produced from direct satellite or in situ observations (Level 2) as well as those that had been gridded (Level 3) or temporally and spatially interpolated (Level 4); products that merged both satellite and in situ observations; products from a variety of sensors; as well as products from a range of data providers (not just EU funded). The filtered data products that underwent detailed scientific analysis for quality information provision is listed in Table 2.

**Table 2.** The filtered data products that underwent detailed scientific analysis for quality information provision.

ECV	Product Name	Provider
Precipitation	Climatic Research Unit Timeseries (CRU TS) 4.0a	University of East Anglia
Precipitation	Global Historical Climatology Network (GHCN) Daily and Monthly Precipitation Product version 2	National Center for Atmospheric Research (NCAR)
Surface Air Temperature	GHCN daily and monthly V3 surface air temperature	NCAR
Surface Air Temperature	Goddard Institute for Space Studies Surface Temperature Analysis (GISSTEMP) monthly 2° anomaly data	National Aeronautics and Space Administration (NASA)
LAI and fAPAR	C3S GEOV1 LAI	VITO, Belgium
LAI and fAPAR	National Oceanic and Atmospheric Administration (NOAA) National Center for Environmental Information (NCEI) LAI v4 (AVH15)	NASA/NOAA NCEI
LAI and fAPAR	Meteosat Second Generation (MSG)/ Spinning Enhanced Visible and InfraRed Imager (SEVIRI)—Land Surface Analysis (LSA)-Satellite Application Facility (SAF)	European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT)
Soil moisture	C3S Soil Moisture—combined active and passive	ESA, Vienna University of Technology (TU Wien)
Soil moisture	Soil Moisture Ocean Salinity (SMOS)-L3-Global	Barcelona Expert Centre
SST	Advanced Very High Resolution Radiometer (AVHRR) SST Level 3 Pathfinder version 5.3	NOAA NCEI
SST	CCI/C3S Along Track Scanning Radiometer (ATSR) Level 3 Sea Surface Temperature version 1.1	European Space Agency (ESA)
Ocean Colour	Globcolour global merged—chlorophyll-a concentration	ACRI-ST, France
Ocean Colour	ESA Climate Change Initiative (CCI)/C3S V3.1—chlorophyll-a concentration	ESA, Plymouth Marine Laboratory
Ozone	Total Ozone (C3S Ozone CCI) Level-2, OMI/Aura	Royal Belgian Institute for Space Aeronomy (IASB-BIRA)
Ozone	Total Ozone (NASA Goddard Space Flight Center (GSFC)) Level-3, Solar Backscatter Ultraviolet Radiometer (SBUV) v8.6	NASA
Aerosol	NASA Moderate Resolution Imaging Spectroradiometer (MODIS) Terra/Aqua Deep Blue Level 3 version 6.1	NASA
Aerosol	CCI (A)ATSR/ Sea and Land Surface Temperature Radiometer (SLSTR) Dual/Single View Algorithm (ADV/ASV) Level 2 version 2.31	ESA
Albedo <sup>2</sup>	QA4ECV Spectral Albedo	European Union funded Quality Assurance for Essential Climate Variables (EU QA4ECV)
Albedo <sup>2</sup>	QA4ECV Broadband Albedo	EU QA4ECV
Formaldehyde (HCHO) <sup>2</sup>	QA4ECV HCHO ECV Precursor Product	EU QA4ECV
Nitrogen Dioxide (NO <sub>2</sub> ) <sup>2</sup>	QA4ECV NO <sub>2</sub> ECV Precursor Product	EU QA4ECV

<sup>2</sup> QA4ECV products were also evaluated in the EQC framework given the extensive QA undertaken on them as part of [3]. These products are no longer operationally developed but will likely be reproduced and refined as part of future CDR development funding opportunities.

### 3. EQC Framework Development

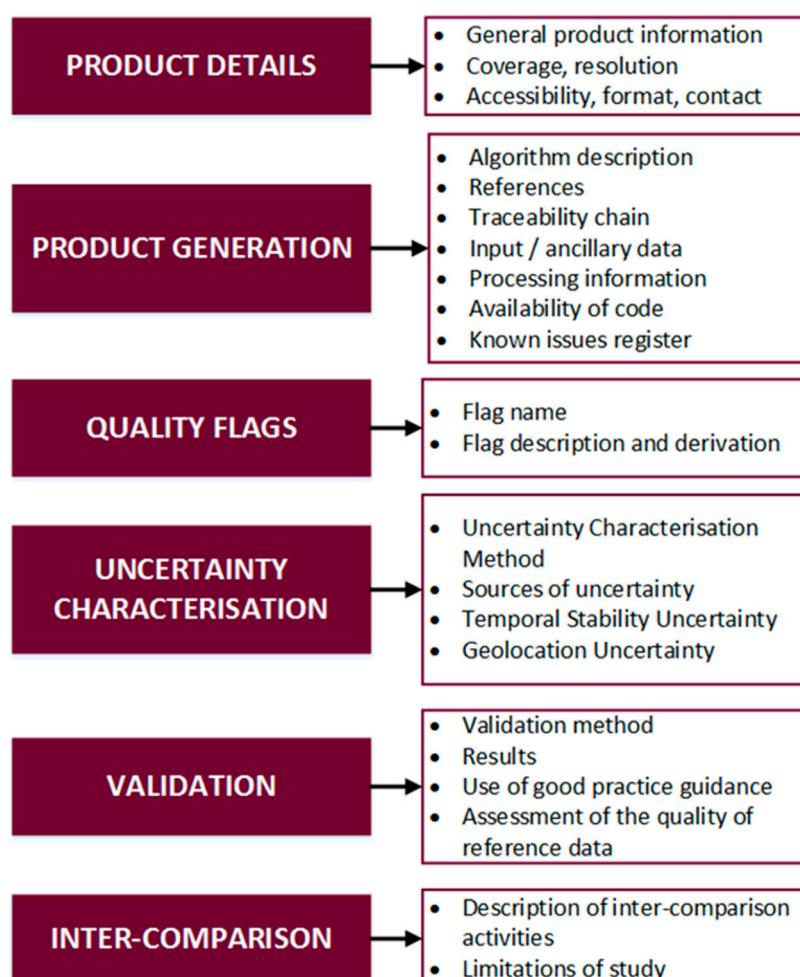
The following sections outline the development of the EQC functionality for the C3S CDS based on an in depth scientific quality analysis of over 20 individual data products representing nine ECVs. This involved the compilation of the Quality Assurance Template (QAT) and independent evaluation process to generate a published product Quality Assurance Report (QAR) as well as the parallel development of the EQC content management system to facilitate the processes.

### 3.1. QAT Development

Building on the concepts developed within the EU FP7 funded QA4ECV project [3], the Quality Assurance Template (QAT) consisted of six fundamental Quality Indicator sections including:

1. Product Details;
2. Product Generation;
3. Quality Flags;
4. Uncertainty Characterisation;
5. Validation; and
6. Inter-Comparison.

Figure 1 provides an overview of the QI sections and the nature of information gathered within each. Separate QAT's were developed specifically for both satellite and in situ observation derived data products. The templates are comprehensive with approximately 250 fields of information to be captured across the QI's. As the QAT is implemented in a webform (discussed in Section 3.3), each section is tailored to only present relevant questions for each ECV or data type. Further, drop down menus are provided to reduce free text fields and ensure the database operates efficiently. The QAT questions were designed with the aim of encouraging the data producer to not only relate the existing product quality information in the standardised manner, but to contemplate various aspects of product quality that may not have been previously considered.



**Figure 1.** Quality Assurance Template (QAT) Quality Indicators with a short overview of the type of information collated within each section.

On average it is expected to take a product producer and/or production team, having full scientific knowledge of their data product, approximately three hours to complete the entire QAT. Further, as part of the EQC content management system (CMS) outlined in Section 3.3, it is anticipated the product details could be imported from existing metadata structures and an autofill capability would enable existing product templates of similar products to be imported for editing to reduce time and effort.

### 3.2. QA Evaluation

A product evaluation method has been devised that will facilitate assessment of whether the product producer has provided sufficient information within each of the QI sections. It has two key purposes: (1) to allow a user to fully understand the status of the data product and make their own informed judgement as to its applicability for their application; and (2) for both users and funding organisations to determine if good practices are being followed in generating the product. A series of questions for each QI section were compiled for a Reviewer (independent product expert), to answer based on the extent of information presented in the QAT by the product producer. The questions are broad enough to encompass all ECV products and only require the Reviewer to check the most appropriate answer so as to minimise overall effort and reduce subjectivity between different evaluations where possible. Figure 2 shows an example of the evaluation questions within the Uncertainty Characterisation QI section that a reviewer would answer.

<b>What level of uncertainty characterisation has been carried out?</b>
<input type="radio"/> N/A <input type="radio"/> Full metrological analysis starting from L0 data <input type="radio"/> Partial metrological analysis <input type="radio"/> Statistical analysis <input type="radio"/> No analysis
<b>Has the methodology been described in sufficient detail for quality assurance?</b>
<input type="radio"/> N/A <input type="radio"/> Yes <input type="radio"/> Mostly <input type="radio"/> About half <input type="radio"/> A little <input type="radio"/> No
<b>Have all input data uncertainties been included (e.g. data, processes, algorithms, parameterisations)?</b>
<input type="radio"/> N/A <input type="radio"/> Yes <input type="radio"/> Mostly <input type="radio"/> About half <input type="radio"/> A little <input type="radio"/> No

**Figure 2.** Example of the evaluation questions within the Uncertainty Characterisation Quality Indicator (QI) section that an independent expert Reviewer would answer based on the information provided by the product producer in the QAT.

Similar to the evaluation process developed for QA4ECV [3], the EQC evaluation only assesses the fraction of information provided relative to all questions and reviewer assessment of information provided. Four levels of achievement ranging from Basic, Intermediate, Good and Excellent are defined. To achieve a rating of Excellent, almost all QI details per individual section must be provided with substantial credible detail, while a score of Basic would indicate that minimal explanation of a QI was provided and that good practices (if currently available) were not necessarily followed. Figure 3 shows an example of the Quality Evaluation Matrix (QEM) results summary achieved for two Ocean Colour data products. This evaluation indicates the QI sections where sparse information was provided across both datasets, potentially highlighting a scientific knowledge gap to be explored through further funding, as well as QI categories in which more information has only been provided for one product indicating a more in depth assessment of the product quality has been provided by the producer.

Details	Generation	Quality flags	Uncertainty Characterisation	Validation	Inter-comparison
Product Information	Input data and uncertainties	Quality Flags	Uncertainty Characterisation Method	Reference data representativeness	Scale of inter-comparison activities
Product Description	Sensor Calibration		Uncertainty sources included	Reference data uncertainty inclusion	Inter-comparison method
Coverage and Resolution	Algorithm method		Uncertainty values provided	Validation method	Product uncertainties inclusion
Data gaps	Algorithm tuning to reference data		Temporal stability	Validation results	Discrepancy between products identified and, if possible, resolved
Data set limitations and target applications	Sensitivity analysis		Geolocation uncertainty		
Documentation	Internal Processes				
	Traceability				

**Key**

- Basic
- Intermediate
- Good
- Excellent

(a) CCI/C3S V3.1 - chlorophyll-a concentration product

Details	Generation	Quality flags	Uncertainty Characterisation	Validation	Inter-comparison
Product Information	Input data and uncertainties	Quality Flags	Uncertainty Characterisation Method	Reference data representativeness	Scale of inter-comparison activities
Product Description	Sensor Calibration		Uncertainty sources included	Reference data uncertainty inclusion	Inter-comparison method
Coverage and Resolution	Algorithm method		Uncertainty values provided	Validation method	Product uncertainties inclusion
Data gaps	Algorithm tuning		Temporal stability	Validation results	Discrepancy between products identified and, if possible, resolved
Data set limitations and target applications	Sensitivity analysis		Geolocation uncertainty		
Documentation	Internal Processes				
	Traceability				

(b) Globcolour global merged - chlorophyll-a concentration product

**Figure 3.** Quality Evaluation Matrices (QEMs) for the (a) CCI/C3S V3.1—chlorophyll-a concentration product and (b) Globcolour global merged—chlorophyll-a concentration product. The CCI product has looked at the temporal stability in an effort to understand its fitness-for-purpose as a Climate Data Record, while Globcolour did no assessment of this. Neither product has been through a formal inter-comparison process highlighting a science gap.

By design, the QA assessment should not be used to determine if one ECV product is better or worse than other comparable data sets in an absolute sense but only in the amount of quality related information available. For example, a data set may have a high uncertainty associated with the values provided, but the producer may have done everything possible to ensure the best values given the data available and methods used, and may have provided all the information required in the QAT. This would give the product a high overall QA grade per QI, but the data set may not be

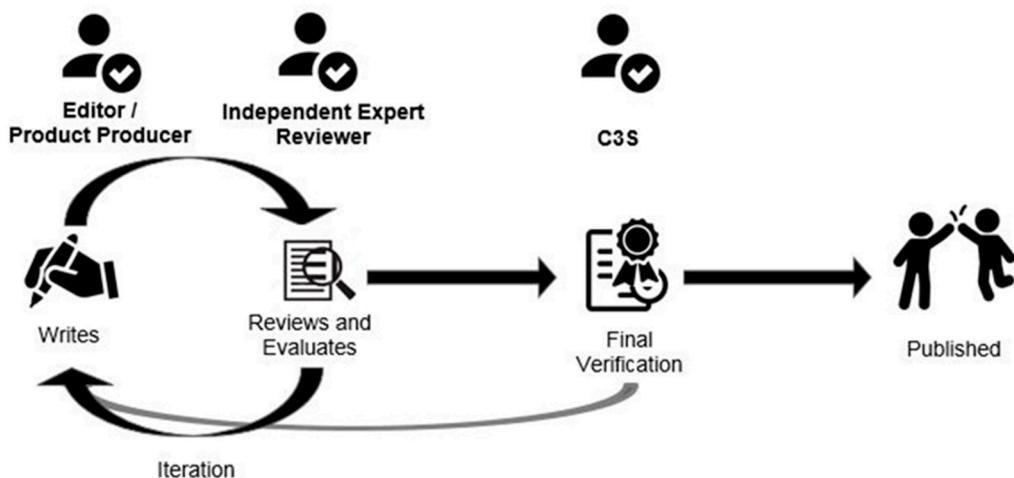
particularly useful beyond a very limited set of applications. Therefore, from a user application point of view the data may be considered of little utility, but from a QA point of view the assurance that best methods have been used to generate the data is high. It is anticipated as the EQC is implemented more broadly for a greater number of data products, and as further funding investments and international community efforts address product quality assessments through validation, inter-comparisons and development of uncertainty characterisation methods and guidance, there will be vast improvements in the understanding of the quality of data products, implementation of good practices and uptake of this by data users. Further, assessment of data products for specific applications is being undertaken as part of the C3S Sectoral Information Systems (SIS) and other EQC contracts. Through time, the EQC evaluation process will be refined and strengthened.

### 3.3. EQC Content Management System

In parallel to the manual progression of the QAT's and resulting individual product QARs, an EQC CMS was developed with the purpose of automating the process of collating, evaluating and presenting the quality status of each data product. The EQC CMS is coded in Drupal 8 and is directly compatible with the CDS infrastructure. It allows the creation of QARs with a workflow that consists of three key roles similar to those defined in [3]:

1. *Editors*, product producers who are responsible for filling out a QAT for their data product;
2. *Reviewers*, domain scientific experts who evaluate the QAR information completed by the Editors; and
3. *Approvers*, C3S representatives who provide a final check that the information is credible before the product QAR is issued publicly.

Figure 4 shows the workflow expected within the EQC function of the CDS, noting the iteration loop between the Editor and Reviewer to allow for refinement and enhancement of quality information provided. Based on the information provided by the Editor and the assessment conducted by the Reviewer, the EQC CMS generates a publishable QAR (online or printable in PDF format) as well the QEM described in Section 3.2.



**Figure 4.** Schematic overview of the C3S Evaluation and Quality Control (EQC) Content Management System (CMS) process. *Editors*, (product producers) are responsible for filling out a QAT for their data product; *Reviewers*, are domain scientific experts who evaluate the Quality Assurance Report (QAR) information completed by the Editors; and *Approvers* are C3S representatives who provide a final check that the information is credible before the product QAR is issued publicly. The iteration loop allows for refinement and enhancement of quality information provided.

To ensure the CDS catalogue is representative of the wide range of climate data products currently available (i.e., Table 1), the EQC will be applied incrementally. For example, currently EU funded data products will be expected to meet a higher level of quality information provision and evaluation than those data products that are no longer funded but still considered climate relevant, or are from international data providers. This does not necessarily mean they are of lesser quality or climate relevance, but rather more effort may be required to collate and perform thorough quality evaluations of these datasets to meet the C3S highest standard.

#### 4. Scientific Gap Analysis

Science knowledge gaps were identified in all products evaluated as part of the development of the EQC functionality. The gaps reflect information that was missing when filling out the QAT for 24 demonstration products (Table 2) as well as a more general reflection on what may be considered to be ‘good practice’ informed from other projects (such as those mentioned in Section 2.1) and international committees such as, but not limited to: Committee on Earth Observation Satellites (CEOS); Group on Earth Observations (GEO); Integrated Carbon Observing System (ICOS); and the Inter-governmental Panel on Climate Change (IPCC). The process highlighted cross-domain (land, ocean, atmosphere) as well as ECV-specific science knowledge gaps in relation to addressing the quality of CDRs derived from satellite and in situ observations. In Table 3 we outline the 10 most common and priority science knowledge gaps that will require further research investment to ensure all quality aspects of climate data sets can be ascertained and over time provide users the range of information necessary to confidently select relevant products for their specific application. Recommendations for addressing the science gaps are also provided and are chiefly targeted at data producers and agencies funding CDR product development. However, it is important to note that knowledge of the science gaps and research required to address these gaps is highly relevant to data users who should be aware of data quality issues prior to application of these datasets. Inevitably, it is the data users who will drive the requirement for better data and the provision of quality information with data products into the future.

**Table 3.** Common cross-ECV domain science knowledge gaps that require action to ensure quality aspects of climate data sets from satellite and in situ observations can be ascertained and ensure users have access to the range of information necessary to confidently select relevant products for their specific applications. In all cases, international coordination or endorsement of methods is desirable<sup>1</sup>. Timeframes are indicative of how long it would take to conduct the research to reach operational implementation.

	Science Knowledge Gap	Recommendation	Action	Importance	Research to Operations Timeframe
1.	Application of consistent and metrologically sound vocabulary to describe data and product quality.	The C3S EQC should encourage the consistent use of standardised vocabulary by providing a vocabulary list along with examples. This should be consistent with those being defined through international coordination bodies for satellite and in situ observations <sup>3</sup> .	C3S and EQC operators to establish and encourage use of terminology.	Very High Importance	Short term. (12 months)
2.	Sensor-to-sensor consistency in merged products from Level-1 data products onwards.	Development of procedures to apply metrologically-traceable methods of product stabilisation (e.g., harmonisation) that mutually also returns updated radiance calibration coefficients of sensor series.	Space agencies to ensure calibration coefficients are provided for development of downstream products.	Very High Importance	Harmonisation of a large number of sensors should be medium to long term goal (>5 years)
3.	Lack of long term in situ measurements and field-based campaigns globally that are specifically designed for satellite data/product validation and have documented evidence of metrological traceability.	Errors and uncertainties associated with the validation process need to be addressed and should include estimates of the reference data uncertainty and methodology (spatial/temporal/scaling) employed. The use of internationally endorsed good practice guidance should be encouraged if available. For example, these concepts are being developed through the ESA FRM (Fiducial Reference Measurement) projects.	Funding bodies to commission research and ensure adoption and coordination of good practices and guidance documentation.	High Importance	Adopting and developing further internationally endorsed good practice guidance is required in the medium term (2–5 years)
4.	Understanding uncertainties and error correlation associated with using Radiative Transfer Models (RTM) for ECV product generation.	Good practice guides should be developed to help ECV producers to use the optimum RTM/associated models for their application.	Data Provider and Research Community to develop good practice guidance.	High Importance	Guides should be produced in the medium term (2–5 years)
5.	Traceable assessment of Level-1 data in active and passive sensors.	Development of a framework for the metrological characterisation of satellite instruments that encompasses exploitation of ongoing pre-flight and post-launch calibration activities.	Funding bodies to commission research and ensure adoption.	High Importance	Research and operational understanding over the medium term (2–5 years)
6.	Implementation of end-to-end metrological traceability.	All Level-1 satellite-derived data records must have appropriate uncertainties included. These would be derived using a metrological approach to ensure that instrumental biases will also be reduced.	Funding bodies to commission research and ensure adoption.	High Importance	Research commissioning and operational understanding over the medium term. (>5 years)

**Table 3.** *Cont.*

<b>Science Knowledge Gap</b>	<b>Recommendation</b>	<b>Action</b>	<b>Importance</b>	<b>Research to Operations Timeframe</b>
7. Retrieval algorithm cross-comparisons (round-robbins) are required given the large number of data products and algorithms (Table 1).	Retrieval algorithm cross-comparisons activities should be funded in order to understand the relative performance and strengths of different methods.	Funding bodies to commission research and ensure adoption.	Medium Importance	Selected ECVs should begin in the short term. Ongoing in the medium term. (2–5 years)
8. Quantification and assessment of the quality of all ancillary data utilised in ECV product generation.	Data providers should justify the use of ancillary data and models in their products to ensure that results can be defended. If data is being used that is out of date but it is too complex to switch due to assumptions built into the processing scheme, then a sensitivity analysis of the consequences of including such data should then be performed.	C3S EQC to request justification in the EQC process.	Medium Importance	Requested justifications for existing products should be provided in the short term and operationalised into the future.
9. Development of consistent quality flags for each ECV product group.	Consistent and standardised quality flags will facilitate unbiased cross-comparison of the same ECV from different data providers. This will require coordination between product producers to agree on a set of consistent quality flags for their ECV product group.	C3S and funding bodies to coordinate a collaboration between ECV data providers.	Medium Importance	Collaboration set up in the short term. May be conducted within round-robin exercises. (R7). (<12 months)
10. Assessment of the implications for use of differing cloud masks, classification routines and gridding schemes used in all ECV products.	Cross-comparisons activities should be funded to address effects of cloud masking techniques, classification routines and gridding schemes as well as the evaluation of uncertainties in the process.	Funding bodies to commission research and ensure adoption.	Medium Importance	Recommended to be carried out in the medium term. (2–5 years)

<sup>3</sup> International coordination bodies for satellite and in situ observations include for example: CEOS; GEO; ICOS; IPCC.

#### 4.1. Recommendation 1—Standardised Metrological Vocabulary

Following the generic assessment of approximately 250 data products and detailed evaluation of 24 demonstrator data products, it is apparent that there is a pressing need for consistent use of vocabulary. In particular, the words ‘error’ and ‘uncertainty’ are widely misused [18]. Metrologists have standardised definitions of all terms related to measurement which can be found in the International Vocabulary of Metrology [19,20]. These terminologies and their use for Earth Observations (EO) data products are being evaluated as part of several European funded projects (i.e., QA4ECV, FIDUCEO) [3,18] and being adopted by CEOS, but an overarching consistent ECV QA vocabulary glossary is not yet available.

#### 4.2. Recommendation 2—Sensor-to-Sensor Consistency in Merged Products

The sort of time scale required for climate data is often longer than the lifetime of any individual sensor. This then means that all the sensors used to make a climate data record must be made consistent so that the changing of sensor does not introduce offsets into the data which may introduce spurious trends. Within the QAR, there is an explicit question relating to this topic so it is clear when this happens for any given ECV product. In truth, however, the methods used to enforce consistency can be rather ad-hoc. Ideally, making the sensors consistent should be based on a complete understanding of the characteristics of each sensor and its calibration. The sensors should also be corrected to an independent reference when available. Within the FIDUCEO project, this is achieved by harmonising the sensors which means recalibrating the sensors taking into account the known differences in, for example, the spectral response functions. The process also takes into account any error correlation structures between collocations etc., as part of a metrological approach. Looking through the QARs, almost all products have undertaken steps to make the sensors consistent. Within this, however, there is a range of different approaches from use of ground sites to sensor to sensor inter-comparison with simple bias corrections to scaling methods to more sophisticated methods. Given the importance of this step for many ECVs, all the different approaches need to be analysed and assessed for fitness of purpose and the impact on uncertainties on the final products. In particular, schemes that do simple bias corrections to correct for difference between the sensors need to be assessed to ensure that trends due to drifts in calibration error are properly accounted for. Again, ideally, this should be based on metrological techniques to ensure that all sources of error are accounted for but at the very least independent assessments of the methods used to make sensors consistent should be made.

#### 4.3. Recommendation 3—Validation Data and Methods

Validation is the process of assessing, by independent means, the quality of the data products derived from the system outputs [21]. Consistency of validation methodology across ECVs including the metrological assessment of the quality of reference data and documentation of product validation procedures for future usage is required. There are currently a range of different methods used to validate data products within and across ECVs which makes it difficult to directly compare validation studies. Several of the ECV communities have or are in the process of providing good practice guides which will improve this situation. Good practice guides for validation are being commissioned through the CEOS Working Group on Calibration and Validation (WGCV), see [22,23]. They are developed through in-kind contribution of a global network of experts for each ECV and can therefore take many years to be produced and made publicly available.

Evaluation of the validation methodologies used by different groups, reveals that certain sources of error are not being included in the analysis. Often the uncertainty in the reference (in situ or field measured) data is not included, though in part this may be because many reference sources still do not have accurate uncertainty estimates to be used. Representativeness, which can be related to the difference in spatial/temporal scales between the satellite data and the reference, is also not often taken into account. Further, because the ECVs tend to cover quite long periods of time, the quality and

sampling of the reference datasets also changes over time and this evolution of the reference networks should be taken into account when using the validation data to assess a given product, particularly when looking for trends in the data. The ESA Fiducial Reference Measurement (FRM) programme is supporting in situ measurement campaigns and the establishment of long term field sites specifically for the validation of satellite-derived data products. In support of CEOS activities, these ESA funded sites must: provide documented evidence of metrological traceability to SI (or appropriate international community standard) including a full uncertainty budget (instrumentation and usage); consider all spatial/temporal/scaling issues; be independent of any satellite geophysical retrieval process; provide long term sustainable mission validation information which may facilitate interoperability between sensors; and be carried out following (or developing as needed) community agreed good practice protocols. The FRM programme is currently supporting several ECVs within various projects including for example, Surface Temperature, Ocean Colour, Vegetation and Atmospheric Composition [24].

#### 4.4. Recommendation 4—Radiative Transfer Models

Many ECV products use Radiative Transfer Model (RTM) output as part of the retrieval process. Different products use different RTMs which will inevitably have different characteristics and error correlation structures. Some radiative models are ‘state of the art’ whereas other models are used due to heritage and may not be as up to date as possible. Even current RTMs will have remaining sources of error which may or may not be important for a given application and which need to be understood. The quality and uncertainties implicit in any input data to the RTM also need to be assessed since this will also contribute to the errors in the modelled values produced. Other RTM related issues include which emissivity models were used and how representative of the real world they are. In some wavelengths/surface types this can be very important and potentially be the source of significant error. The EC Joint Research Center (JRC) led Radiation Transfer Model Intercomparison (RAMI) initiative is a mechanism to benchmark models designed to simulate the transfer of radiation at or near the Earth’s terrestrial surface [25].

#### 4.5. Recommendation 5—Traceable Assessment of Level-1 Data

All satellite-derived ECV products start by using Level-1 data [26], and for passive sensors it generally consists of geolocated and calibrated radiances, while other measured quantities (apart from radiances) will be used for active sensors. All sensors need to have some form of calibration (on board and/or post-launch) to derive the required measurand at Level-1. It is often assumed that it is the responsibility of the Level-1 provider to ensure that the data is as well characterised as possible and that the data can be efficiently used without modification. However, there are cases where the assumption of a reliable Level-1 data set has been shown to be wrong within the lifetime of a given sensor series and where operational modifications and/or external recalibrations must be undertaken to reduce the calibration error. It is important to note that some of the sensors used to create CDRs were not designed with the stringent accuracies required by climate studies. For example, as there is no visible channel calibration system on-board the Advanced Very High Resolution Radiometer (AVHRR), visible channel calibration must be modelled after the observation based on ground target measurements to track the calibration degradation [27].

As instrumentation design improves with the addition of on-board calibration systems, both the prevalence and size of calibration errors have reduced, and for some applications the most modern sensors may not require any significant calibration correction. We do note, however, that even a well-designed sensor can itself have a poor calibration if there are inbuilt assumptions in the calibration process that are themselves not accurate, so it is not necessarily a given that modern sensors are bias free. In general, calibration errors usually present themselves in the form of biases in the Level-1 data when compared against trusted references (ideally traceable to the Système international d’unités, SI). For satellite data, another challenge is that pre-flight calibration may not be appropriate for in-orbit behaviour of the instrument [28–30], especially for the older sensors. In terms of general uncertainties

at Level-1, they are often simplistic such as a single quoted noise equivalent delta temperature ( $\text{NE}\Delta\text{T}$ ) or may not be present at all. It is therefore unsurprising that the use of Level-1 uncertainties by ECV producers is highly variable, from not using any Level-1 uncertainties at all to trying to use more complex uncertainty components. The FIDUCEO [12] project is one project that has been designed to demonstrate how such effects can be modelled and corrected for post-launch by adopting a measurement equation approach to recalibrate the data and propagate uncertainty information [20,31]. The presence of identified scientific gaps in metrological traceability from Level-0 to Level-1 for all satellite datasets means that it is still too early for derived products to claim the level of climate stability and/or accuracy over the required length of time to be considered useful for climate applications.

#### 4.6. Recommendation 6—Implementing End-to-End Metrological Traceability

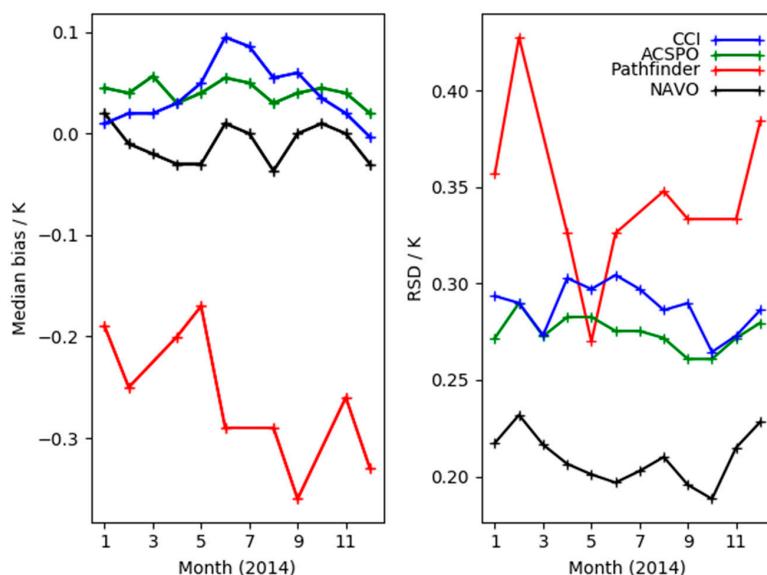
Assessment of uncertainties should be routine to the production of a CDR and should ideally take into account all sources of error present within the data and processing systems. Without justifiable uncertainties, accurate statements about trends and changes cannot be realistically made. It should be noted that while most products come with some sort of estimate of uncertainty, this does not mean that the uncertainties have been traced back to a reference or (in the best case scenario) to SI (Metrological traceability). For EO data this is defined as tracing all known sources of error from their original source through to the final derived product. To aid in implementing and demonstrating end-to-end metrological traceability, it is recommended that a traceability chain should be developed for each data product. A traceability chain is a diagrammatic and partly interactive representation of the processing steps taken to produce the final data product. It shows sub-processing chains and intermediate products/parameters, as well as provides a short description of each step and where to find more detail on the process implemented [3]. Developed as part of the QA4ECV project, traceability chains aid a user in understanding the data production and the assumptions that are made during implementation and are extremely popular among data users and producers alike. The traceability chain concept should be expanded further as a means of communicating metrological traceability within measurements and algorithms.

What is clear from projects such as FIDUCEO, where a detailed analysis of EO uncertainties has been undertaken, is that uncertainties are not simple but consist of different components which are related to how the underlying sources of error correlate (e.g., [32]). The error correlations that have been found relate to both spatially correlated and temporally correlated error sources as well as channel to channel correlations. All of these will be important when retrieving an ECV variable. To simplify this the FIDUCEO project has developed three different types of uncertainties which are called independent, structured and common and has also included channel to channel correlation matrices which may be important in ECV retrieval [32]. Under this scheme, independent uncertainties are where all components of uncertainty are considered random. It is this component which may already be available to some degree through estimates of the  $\text{NE}\Delta\text{T}$ . Structured uncertainties are those where some process has imposed a correlation structure on some spatial or temporal scale. One example is if the raw calibration data is averaged across scanlines which imposes an error correlation structure onto the uncertainties and so has to be dealt with separately if uncertainties at further levels of processing are to be correct. Finally, there are common uncertainties where the underlying errors are fully correlated over large spatial and temporal scales and so will not reduce if spatial or temporal averaging is subsequently used. Geolocation uncertainty will be important in determining uncertainties related to classification processes, which will feed into the final product uncertainty. It is also important for any validation studies to ensure that a proper understanding of representativeness between the reference data and the product itself. For example, in the case of the ESA CCI's Along Track Scanning Radiometer (ATSR) aerosol product, the validation was limited to locations where there was available Aerosol Robotic Network (AERONET) data. AERONET is a network of surface upward-looking sunphotometer sensors designed to produce high temporal resolution aerosol measurements at point locations. For this ATSR product, it is also not clear whether

a standard methodology for validation of space based aerosol data against AERONET has been used, like that developed by Ichuko [33], i.e., whether representativeness issues have been taken into account.

#### 4.7. Recommendation 7—Retrieval Algorithm Round-Robin Comparisons

It is vitally important that the retrieval methodologies applied are optimal given the data being used. It has become apparent, however, that there can still be a range of different algorithms used by different groups to derive climate data even when the input data is the same. For example there are at least four different SST products available from the Group for High Resolution Sea Surface Temperature project (GHRSST, [34]) which are based on identical Level-1 inputs from the time-series of AVHRR but which all have different validation statistics. Figure 5 shows the median bias and robust standard deviation (both the median and robust standard deviation are robust to outliers e.g., [35]) for 12 months of SST data from the four SST datasets observed in 2014. It can be seen that they are all different even though they are all measuring the same SST. Ideally it should be possible to develop an optimum algorithm which provides the best estimate in this case rather than having four different approaches.



**Figure 5.** Left hand panel shows the monthly median difference between four different Sea Surface Temperature (SST) products when compared to the drifting buoy network and the right panel shows the robust standard deviation, an outlier robust estimate of the underlying standard deviation. The four products are from ESA CCI, Advanced Clear Sky Processor for Ocean (ACSPO—the NOAA operational AVHRR product), Pathfinder (from the NOAA Pathfinder SST product) and the Naval Oceanographic Office (NAVO—the US Navy SST product). All products used the same input AVHRR Level-1 data so are measuring exactly the same SST but due to algorithmic differences the products are not the same.

The sort of problem highlighted above is, no doubt, present in most ECV products so more cross comparisons need to be undertaken to ensure that any given retrieval is as good as it can be. Just because a certain algorithm has a long heritage, it does not mean that it provides the optimal solution. Some producers do undertake round-robin exercises to try and ensure the optimal result but even when such exercises are performed a mixed picture can emerge. For example, in the case of the CCI Aerosol product, a round-robin was undertaken and it was finally decided to produce three different products which each seemed to work well in a particular domain (e.g., ocean or land) but could not by itself provide the best solution. What really needs to be done in cases like this is an investigation to work out why there are differences and use that information to develop a better set of algorithms overall. This does, however, then require significant extra work which many data producers will likely not wish to undertake. Studies need to be done to understand differences between different algorithms with the goal of developing the optimal retrieval based on what has been learnt.

#### 4.8. Recommendation 8—Quality of all Ancillary Input Data

Many products use ancillary data and or models as part of their retrieval scheme. These data range from climatological datasets, Numerical Weather Prediction (NWP) modelled data, models of surface properties, to models of aerosol. Different data producers will have made different choices regarding which models to use and sometimes the ancillary data used can be very old, likely due to code heritage reasons. There is therefore a need for data providers to justify the use of all ancillary data and/or model inputs relative to the latest knowledge about any given process.

A number of problems with some of the models have been captured during the product evaluation process. For example, three examples are highlighted.

1. For the ocean retrieval of aerosol a whitecap fraction model is used which is very old and probably should be updated [36]. There are much more recent models available and it has been shown that the Monahan model will lead to biases being introduced [37].
2. For the Soil Moisture CCI passive retrieval, an old model [38] is used. More modern models have been shown to outperform this model, so an update should be implemented; and
3. Many processes use climatological data as input to their retrieval. Care needs to be taken that the optimum data is used. For example, the CCI/C3S Aerosol product uses a Chlorophyll concentration climatology based on Coastal Zone Color Scanner Experiment (CZCS) data (a very old instrument) where there is almost certainly better data available.

#### 4.9. Recommendation 9—Consistent Quality Flags

Between products for the same ECV as well as across ECVs themselves, there is little consistency between the implementation of quality flags. Quality flags are very useful for the user and ideally should be easy to use and interpret allowing data filtering and enhancing knowledge of the production issues as well as pixel level uncertainties. Evaluation of the demonstrator ECV products revealed that data providers define quality flags differently, making comparisons between datasets difficult. Even in the case where the quality flags have been formally defined to be present across a range of products as is the case with SST, the actual use and meaning of different quality flags varies and can still vary from product to product. Recommendations on an initial set of data product quality flags that should be implemented widely have been provided in [3] and consist of the following: number of observations used in the calculation; snow/cloud cover; back-up algorithm implementation; fill-values utilised; pixel-based uncertainty estimates.

#### 4.10. Recommendation 10—Cloud Masks and Classification Routines

There are often times when some sort of classification is required to retrieve the correct parameter. Probably the most common of these are cloud masks but also includes classification of surface properties (i.e., land cover classes) and/or classification of parameter type such as aerosol type. Getting the classification wrong or using different interpolation methods to grid data can lead to significant biases in the final data. For example, the technique applied to transform a network of point in situ measurements to a set of gridded values may be greatly affected by the density of observations available. A comparison between two in situ derived gridded datasets—CRU and Global Precipitation Climatology Centre (GPCC) [39]—demonstrates that differences are slight for grid cells with many measurement stations in proximity. For cells where such stations are sparse however, anomalies at individual stations cause greater differences between the datasets.

Cloud masks are used in many ECV products, but often bespoke schemes are employed so it is very difficult to compare products. Most of the cloud masks used in the ECVs evaluated seem to be based on threshold based tests where a pixel is flagged as cloudy if it passes (or fails) a series of threshold based tests. This cloud masking technique has a long heritage and some cloud masking routines can have dozens of different threshold tests. One advantage of a threshold test is that the individual tests can focus on potentially problematic cloud types which may allow a more certain

detection of specific clouds can be hard to find. However, the key problem with using a threshold based system is that it is harder to take uncertainty into account since the thresholds are generally pass/fail and dataset noise is not considered. Alternatively, the Bayesian technique estimates the probability of a given pixel being clear or cloudy and generally uses a combination of clear sky modelling and cloud Probability Density Functions (PDF) to determine the likelihood of it being cloudy. As it is a probabilistic method, it can take into account uncertainties on the radiances/brightness temperatures.

Problems with cloud masking can have a demonstrable impact on the retrieved values. One such prominent example involved extensive scientific community debate concerning the interpretation of satellite derived estimates of Amazonian tropical rainforest response to changes in climate [40–42]. The presence of large cloud cover fraction and aerosol concentrations over the Amazon along with the various satellite data processing schemes employed by different product developers led to conflicting evidence over sensitivity of the rainforest to prolonged drought events [43]. Hilker et al. [43] showed the difference in Enhanced Vegetation Index (EVI) and Normalized Different Vegetation Index (NDVI) detectable change at 95% confidence with different atmospheric correction and cloud masking schemes. The study provided a direct statistical analysis of a measurable change in daily and composite surface reflectance obtained from the Moderate Resolution Imaging Spectroradiometer (MODIS) based on the noise level of data and the number of available observations post aerosol and cloud masking, which provided a greater number of observations to assess response in the tropical forest to climate fluctuations.

## 5. Summary and Future Recommendations

Here we have presented an initial framework for the Evaluation and Quality Control of climate data products derived from satellite and in situ observations to be catalogued within the C3S Climate Data Store. It builds on past and present international investment in Quality Assurance for Earth Observation initiatives, extensive user requirements gathering exercises, as well as a broad evaluation of over 250 data products and comprehensive evaluation of a selection of 24 individual satellite and in situ observation derived products across the land, ocean and atmosphere Essential Climate Variable (ECV) domains. An EQC CMS has been developed to facilitate the process of collating, evaluating and presenting the quality aspects and status of each data product to data users.

The development of the EQC framework highlighted cross-domain as well as ECV specific science knowledge gaps in relation to addressing the quality of climate data sets derived from satellite and in situ observations. The top 10 common priority science knowledge gaps that will require further research investment have been outlined in detail. These recommendations are chiefly targeted at data producers and agencies funding CDR product development. The science knowledge gaps vary in complexity and the level of effort required to address in a research context and implement operationally. Dependencies exist between the knowledge gaps and thus dedicated research in any one will help inform improved data transparency, traceability and climate applicability of all data products. The goal of the EQC functionality is to ensure users are provided with a range of product quality indicators, so they can confidently select relevant products for their specific application. Further, it is important to note that knowledge of the science gaps and research required to address these gaps is highly relevant to data users who should be aware of data quality issues prior to application of these datasets. Inevitably, it is the data users who will drive the requirement for the provision of better quality information with data products into the future.

### 5.1. Further Development of the EQC Functionality

The EQC framework will be implemented by C3S as part of their operational quality assurance programme. Further development and refinement of the EQC framework and CMS is ongoing. Below we provide several suggestions for this continued development in relation to three key areas including implementation, improvements and additional functionalities that were not implemented in the initial development phase.

### 5.1.1. Implementation

Each individual data product to be catalogued within the CDS will require:

1. A QAT to be completed;
2. An independent assessment of the QAT information; as well as
3. A CDS placeholder for the dataset.

The CMS will need to be expanded to handle these features for all the data types including observations, model-based climate reanalyses, seasonal forecast data products, and climate model simulations including projections. Data type specific QATs will need to be developed along with relevant evaluation questions, assessment process and publishable QARs. Enhancing the CMS functionality in relation to data import, auto-save and account synchronisation will ensure a seamless integration of these additional templates and processes into the CDS. When implementing the EQC functionality for the multitude of data products to be hosted through the CDS, it will be necessary to address several aspects such as the minimum requirements of QAR content before a data product can be listed in the CDS. It will also be necessary to find and recruit suitable product Reviewers (product evaluation experts) to ensure professional appraisals. To guarantee consistency in QAT evaluations within and between data sets it is recommended that a set of evaluation guidance for producers and evaluators be developed to facilitate this and that regular evaluation benchmarking activities are brought into the operational process.

### 5.1.2. Maintaining and Improving Quality Assurance

It is well known that data products are updated and improved through time in relation to funding cycles, as well as updates to sensor calibrations, improvements to algorithms through round-robin exercises and validation activities as well as simply through the extension of the data sets and new scientific advances. The EQC CMS will need to expand and evolve the QATs and evaluation fields and scoring to reflect these updates in scientific techniques. The CMS will also need to accommodate data preservation issues in relation to storing old versions of product QARs as new versions of data products become available and/or product contacts change. It is also recommended that in addition to coordination with and adoption of international good practices, the EQC dedicates resource to the development of guidance and training or workshops on the QA requirements for the CDS. Training on subjects such as the application of metrological in the context of ECV data should also be considered to help improve the amount of quality information (such as proper uncertainties) as well improve overall quality of the data.

### 5.1.3. Additional Functionalities

Additional useful functionalities of the EQC CMS may include: development of a QAR comparison tool to enable direct comparison of similar ECV data products; and the ability to track changes (time and date stamped) in the QAT throughout the review process to ensure both the product producer and expert reviewer are using the most current version of the template. Finally, it is recommended that the EQC invest in the development of a tool that is capable of making detailed and interactive product traceability chains to augment the product generation section of the QARs.

**Author Contributions:** J.N. is the main author and represents NPL who led the EQCO framework development for C3S; J.P.D.M. is a co-author and contributed to the design and development of the EQC; S.D., J.R., M.T., C.O., C.D. and C.M. contributed to the design and development of the EQC; D.D. is the Deputy Head of the C3S and provided input on the status and future of the C3S and EQC; C.F. and G.D. developed the EQC CMS.

**Funding:** The work reported here was carried out with EU funding under contract C3S\_51 Lot 2 with ECMWF. ECMWF implements the Copernicus Climate Change Service on behalf of the European Commission.

**Acknowledgments:** The authors wish to acknowledge colleagues who were former representatives of the University of Reading (Laura Carrea and Margaret Woodage) and Telespazio France (Caroline Quod and Christelle Barbey), who contributed to the development of the EQC for observations. We could also like to acknowledge the

individual product producers who worked with us to complete QARs for the 24 demonstrator products. As well as Alex Hale from NPL for project management support.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. UNFCCC. The Paris Agreement—2015. Available online: [http://unfccc.int/paris\\_agreement/items/9485.php](http://unfccc.int/paris_agreement/items/9485.php) (accessed on 21 April 2019).
2. Raoult, B.; Bergeron, C.; Alos, A.L.; Thepaut, J.-N.; Dee, D. Climate service develops user-friendly data store. In *Meteorology*; ECMWF: Reading, UK, 2017; pp. 24–27.
3. Nightingale, J.; Boersma, K.F.; Muller, J.-P.; Compernolle, S.; Lambert, J.-C.; Blessing, S.; Giering, R.; Gobron, N.; de Smedt, I.; Coheur, P.; et al. Quality assurance framework development based on six new ECV data products to enhance user confidence for climate applications. *Remote Sens.* **2018**, *10*, 21. [CrossRef]
4. ESA. *Earth Observation Science Strategy for ESA: A New Era for Scientific Advances and Societal Benefits*; ESA Communications: Noordwijk, The Netherlands, 2015.
5. Loew, A.; Bell, W.; Brocca, L.; Bulgin, C.; Burdanowitz, J.; Calbert, X.; Donner, R.; Ghent, D.; Gruber, A.; Kaminski, T.; et al. Validation practices for satellite-based Earth observation data across communities. *Rev. Geophys.* **2017**, *779*–817. [CrossRef]
6. Zeng, Y.; Su, Z.; Calvet, J.-C.; Manninen, T.; Swinnen, E.; Schulz, J.; Roebeling, R.; Poli, P.; Tan, D.; Riihela, A.; et al. Analysis of current validation practices in Europe for space-based climate data records of essential climate variables. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *42*, 150–161. [CrossRef]
7. CDS, C.S. C3S Climate Data Store. Available online: <https://cds.climate.copernicus.eu> (accessed on 21 April 2019).
8. Bojinski, S.; Verstraete, M.; Peterson, T.; Richter, C.; Simmons, A.; Zemp, M. The concept of essential climate variables in support of climate research, applications, and policy. *Am. Meteorol. Soc.* **2014**, *1431*–1443. [CrossRef]
9. Widlowski, J.-L. Conformity testing of satellite-derived quantitative surface variables. *Environ. Sci. Policy* **2015**, *51*, 149–169. [CrossRef]
10. QA4ECV. Quality Assurance for Essential Climate Variables. Available online: <http://www.qa4ecv.eu/> (accessed on 21 April 2019).
11. GAIA-CLIM. Gap Analysis for Integrated Atmospheric ECV CLImate Monitoring. Available online: <http://www.gaia-clim.eu/> (accessed on 21 April 2019).
12. FIDUCEO. FIDelity and Uncertainty in Climate data records from Earth Observations. Available online: <http://www.fiduceo.eu/> (accessed on 21 April 2019).
13. CLIM-RUN. Climate Local Information in the Mediterranean Region Responding to User Needs. Available online: <http://www.climrun.eu/> (accessed on 21 April 2019).
14. EUPORIAS. European Provision of Regional Impacts Assessments on Seasonal and Decadal Timescales. Available online: <http://www.euporias.eu/> (accessed on 21 April 2019).
15. CLIP-C. Climate Information Portal. Available online: <http://www.clipc.eu/> (accessed on 21 April 2019).
16. GLOBTEMP. GlobTemperature. Available online: <http://www.globtemperature.info/> (accessed on 21 April 2019).
17. CORE-CLIMAX. COordinating Earth Observation Data Validation for RE-Analysis for CLIMAtre ServiceS. Available online: <https://cordis.europa.eu/project/rcn/106564/reporting/en/> (accessed on 21 April 2019).
18. Merchant, C.J.; Paul, F.; Popp, T.; Ablain, M.; Bontemps, S.; Defourny, P.; Hollmann, R.; Lavergne, R.; Laeng, A.; de Leeuw, G.; et al. Uncertainty information in climate data records from Earth observation. *Earth Syst. Sci. Data* **2017**, *9*, 511–527. [CrossRef]
19. JCGM. *International Vocabulary of Metrology—Basic and General Concepts and Associated Terms (VIM)*; JCGM: Paris, France, 2012.
20. JCGM-100. *Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement*; JCGM: Pavillon de Breteuil, France, 2008.
21. Justice, C.; Belward, A.; Morisette, J.; Lewis, P.; Privette, J.; Baret, F. Developments in the validation of satellite sensor products for the study of the land surface. *Int. J. Remote Sens.* **2000**, *21*, 3383–3390. [CrossRef]
22. Fernandes, R.; Plummer, S.; Nightingale, J.; Baret, F.; Camacho, F.; Fang, H.; Garrigues, S.; Gobron, N.; Lang, M.; Lacaze, R.; et al. *Global Leaf Area Index Product Validation Good Practices*; Version 2.0; CEOS: Frascati, Italy, 2014.

23. Guillevic, P.; Götsche, F.; Nickeson, J.; Hulley, G.; Ghent, D.; Yu, Y.; Trigo, I.; Hook, S.; Sobrino, J.A.; Remedios, J.; et al. *Land Surface Temperature Product Validation Best Practice Protocol*; Version 1.1; CEOS: Frascati, Italy, 2018.
24. ESA. Fiducial Reference Measurements: FRM. 2019. Available online: <https://earth.esa.int/web/sppa/activities/frm> (accessed on 21 April 2019).
25. RAMI. RAdiation Transfer Model Intercomparison. Available online: <http://rami-benchmark.jrc.ec.europa.eu/HTML/> (accessed on 21 April 2019).
26. Parkinson, C.; Ward, A.; King, M. *Earth Science Reference Handbook: A Guide to NASA's Earth Science Program and Earth Observing Satellite Missions*; National Aeronautics and Space Administration: Washington, DC, USA, 2006.
27. Wu, X.; Sullivan, J.; Heidinger, A.K. Operational calibration of the Advanced Very High Resolution Radiometer (AVHRR) visible and near-infrared channels. *Can. J. Remote Sens.* **2010**, *36*. [CrossRef]
28. Mittaz, J.; Bali, M.; Harris, A. The calibration of broad band infrared sensors: Time variable biases and other issues. In Proceedings of the EUMETSAT Meteorological Satellite Conference, Vienna, Austria, 16–20 September 2013.
29. Mittaz, J.; Harris, A. A physical method for the calibration of the AVHRR/3 thermal IR channels. Part II: An in-orbit comparison of the AVHRR longwave thermal IR channels on board MetOp-A with IASI. *J. Atmos. Ocean. Technol.* **2011**, *28*, 16. [CrossRef]
30. Mittaz, J.; Harris, A.; Sullivan, J. A physical method for the calibration of the AVHRR/3 thermal IR channels 1: The prelaunch calibration data. *J. Atmos. Ocean. Technol.* **2009**, *26*, 996–1019. [CrossRef]
31. Woolliams, E.; Mittaz, J.; Merchant, C.J.; Hunt, S.; Harris, P. Applying metrological techniques to satellite fundamental climate data records. *J. Phys. Conf. Ser.* **2018**, *972*, 012003. [CrossRef]
32. Mittaz, J.; Merchant, C.J.; Woolliams, E.R. Applying principles of metrology to historical earth observations from satellites. *Metrologia* **2019**. [CrossRef]
33. Ichoku, C.; Chu, D.A.; Mattoo, S.; Kaufman, Y.; Remer, L.; Tanre, D.; Slutsker, I.; Holben, B. A spatio-temporal approach for global validation and analysis of MODIS aerosol products. *Geophys. Res. Lett.* **2002**, *29*, 1616–1619. [CrossRef]
34. GHRSST. Group for High Resolution Sea Surface Temperature. Available online: [www.ghrsst.org](http://www.ghrsst.org) (accessed on 21 April 2019).
35. Merchant, C.; Harris, R. Toward the elimination of bias in satellite retrievals of sea surface temperature: 2. comparison with in situ measurements. *JGR Oceans* **1999**, *104*, 23579–23590. [CrossRef]
36. Monahan, E.; Muircheartaigh, I. Optimal power-law description of oceanic whitecap coverage dependence on wind speed. *J. Phys. Oceanogr.* **1980**, *19*, 2094–2099. [CrossRef]
37. Albert, M.; Anguelova, M.; Manders, A.; Schaap, M.; de Leeuw, G. Parameterization of oceanic whitecap fraction based on satellite observations. *Atmos. Chem. Phys.* **2016**, *16*, 13725–13751. [CrossRef]
38. Wang, J.R.; Schmugge, T.J. An empirical model for the complex dielectric permittivity of soils as a function of water content. *IEEE Trans. Geosci. Remote Sens.* **1980**, *18*, 288–295. [CrossRef]
39. Harris, I.; Jones, P.; Osborn, T.; Lister, D. Updated high-resolution grids of monthly climatic observations—the CRU TS3.10 Dataset. *Int. J. Climatol.* **2014**, *34*, 623–642. [CrossRef]
40. Huete, A.; Didan, K.; Shimabukuro, Y.; Ratana, P.; Saleska, S.; Hutyra, L.; Yang, W.; Nemani, R.; Myneni, R. Amazon rainforests green-up with sunlight in dry season. *Geophys. Res. Lett.* **2006**, *33*, L06405. [CrossRef]
41. Saleska, S.; Didan, K.; Huete, A.; da Rocha, H. Amazon forests green-up during 2005 drought. *Science* **2007**, *318*. [CrossRef]
42. Samanta, A.; Ganguly, S.; Hashimoto, H.; Devadiga, S.; Vermote, E.; Knyazikhin, Y.; Nemani, R.; Myneni, R. Amazon forests did not green-up during the 2005 drought. *Geophys. Res. Lett.* **2010**, *37*. [CrossRef]
43. Hilker, T.; Lyapustin, A.; Hall, F.; Myneni, R.; Knyazikhin, Y.; Wang, Y.; Tucker, C.; Sellers, P. On the measurability of change in Amazon vegetation from MODIS. *Remote Sens. Environ.* **2015**, *166*, 233–242. [CrossRef]

