CMUG Phase 2 Deliverable

Reference: D2.3: Technical report on product assessment

Due date: December 2016

Submission date: February 2017

Version: 3.1

Climate Modelling User Group

Deliverable 2.3

Technical report on product assessment

Centres providing input: Met Office, MPI-M, ECMWF, MétéoFrance, IPSL, SMHI, DLR

Version nr.	Date	Status
0.6	22 May 2015	First CMUG assessment of CCI ECV product documentation
2.0	7 Dec 2015	Second CMUG assessment of CCI ECV product documentation
3.0	11 Jan 2017	Third CMUG assessment of CCI ECV product documentation
3.1	6 Feb 2017	Aerosol ECV updated **UPDATES SHOWN IN RED TEXT IN THE REPORT**

CMUG Phase 2 Deliverable 2.3

Technical report on product assessment

Table of Contents

Technical report on product assessment

1. Purpose and scope of the Technical Report

The purpose of this document is to review the product assessments of each CCI team and give feedback to ESA and the CCI teams. It provides comments and technical advice on some of the CCI Project deliverables, including the "Product Validation and Inter-comparison Report", "Climate Assessment Report" and "Uncertainty Characterisation Report". Other CCI project reports were also assessed when found to be relevant. This document is intended as user feedback to CCI Projects (but is not part of a formal review). This report also compares the CCI ECV products with the specifications described in the GCOS products guidelines[1] to assess their compliancy.

Some of the issues to be addressed in this document are:

- Is the validation of the products adequate?
- Are the error characteristics provided by CCI Projects adequate?
- What are the different components of the uncertainty?
- What is the anticipated impact of the CCI data on the climate research community?
- Is the user assessment complete and representative of the wider community?
- Does the product meet (where feasible/applicable) GCOS requirements for satellite ECVs?

The structure of the report is that each ECV is assessed in its own section and within this there are sub-sections for *Quality* (covers PVIR and PVP), *Uncertainty* (covers UCR), and *Maturity of data* (covers CAR and GCOS requirements).

2. Comments on CCI validation and user assessment reports

The comments in this report refer to the relevant documents available to CMUG near the end of Phase 2 (December 2016). The documents reviewed are listed in Table 1.

A general comment applicable to all ECVs is although there are good ATBDs (Algorithm Theoretical Baseline Document) for all ECVs *before* the round robin comparisons there isn't always an ATBD of the finally selected product.

It is also noted that in accessing documents from CCI project websites the directory structures and naming conventions for the files is not always consistent. Where this was so it had the potential for delaying access or accessing an incorrect file.

[1] 2011 update available at http://www.wmo.int/pages/prog/gcos/Publications/gcos-154.pdf

ECV	Product Validation	Climate Assessment	Uncertainty Characterisation	Other Docs (e.g. PVP, ATBD)	CMUG lead / update
SST	PVIR v1.0 (22.01.14)	CAR v1.0 reviewed in Phase 1	UCR v3 04.12.13	PVP v2.0 (04.02.14) Climate Data Research Package v1.0 (30.09.13)	Met Office 14/12/2016
Ocean Colour	**PVIR v2.03 (09.10.15)**	**CAR v2.04 (04.02.16)**	**Comprehensive Error Characterisation Report (CECR) v2.0 (21.09.15)**	**PVASR v3.0 (Pt1 23.12.15; Pt 2 15.01.16). PUG v3.0.1 (02.11.16) (The PVP was reviewed in Phase 1 and does not appear to have been updated)**	Met Office 19/12/2016
Sea Level	**PVIR v1.1 (01.12.16)**	**SLCCI_CAR v1.3 (14.09.16)**	**Error Characterization Report v2.2 (29.07.16)**	(PVP was reviewed in Phase 1) **SLCCI-ATBDv1-016 V3.3 (23.08.16)**	MétéoFrance 27/12/2016
Sea Ice	PVIR v1.0 25.11.2014	CAR v1.0 26.11.2014	CECR v1.1 (20.08.12)	PUG (The PVP was reviewed in Phase 1)	MPI-M 06/12/2016
Clouds	PVIR v2.1 (09.12.13)	CAR v1.0 (08.11.13)	CECR V3.0 (21.05.15)	PVP V2.0 (12.06.13)	SMHI 19/12/2016
Aerosol	**PVIR v2.4 (06.04.16)**	**CAR v1.6 (27.07.16)**	**CECR v1.6 (27.07.16)**	(The PVP v1.5 was reviewed in Phase 1)	DLR 11/01/2016
Greenhouse Gases	**PVIR-GHG-CCI-v4 (24.02.16)**	**GHGCCI_CAR_v3 (03.05.16)**	Not available	(ATBD v1.0 (15.03.12) was reviewed in Phase 1)	ECMWF 14/12/2016
Ozone	**PVIR v2.0 (30.06.16)**	CAR v1.0 reviewed in Phase 1	**Ozone_cci_KIT_CECR_02_01_01_v2 (20.05.16)**	PVP v1.0 reviewed in Phase 1	ECMWF 14/12/2016
Land cover	IPVR v1.2 (17.12.12.)	CAR v1.1 (04.07.14)	Not available	(The PVP v1.3 was reviewed in Phase 1)	IPSL 14/12/2016
Fire	PVR-II v1.3 (24.10.14)	CAR v2.1 (31.10.14)	CECR v2.1 (11.11.14)	(The PVP was reviewed in Phase 1)	MPI-M 19/12/2016
Soil moisture	PVIR v1.0 (27.11.2014)	Not available	**CECR v01.0 (23.06.16)**	PVASR v1.0 (27.06.13) (The PVP v1.1 was reviewed in Phase 1)	MPI-M 11-01-2017
Ice Sheets (Greenland)	PVIR v1.4 (28.09.15) Phase 1	CAR v2.1 (28.09.15) Phase 1	CECR v1.2 (13.06.13) Phase 1	PVP v1.2.1 (18.10.12) Phase 1	Met Office 19/12/2016
Glaciers	Not available	CAR v0.5 (28.05.14) Phase 1	UCR v2 (06.06.16)	None	Met Office 19/12/2016

Table 1. Version of documents available for review. Items in red are updated in this version of the report.

2.1 Sea Surface Temperature

Quality

Most of the relevant material is in chapters 7 to 9 in the Product Validation Plan (the PVP assessed here is a newer version than that assessed in CMUG Phase 1). There is also relevant material in the PVIR (Product Validation and Intercomparison Report, CAR (Climate Assessment Report) and UCR (Uncertainty Characterisation Report).

For the Product Validation Plan the methods to be adopted are those approved by the GHRSST science team and conform to the guidelines under the QA4EO framework under the CEOS-WGCV. This international oversight of the validation plans is to be encouraged by all CCI teams.

The validation criteria for in-situ measurements of being within 2 hours of the satellite overpass is rather relaxed especially during the day when diurnal thermoclines are present. However it is recognized that with current matchup datasets this has to be a compromise.

The concept of validation confirmation levels is proposed which is an attempt to validate the uncertainty in the *in-situ* data as well as the product. This is an interesting proposal. CMUG have requested access to these maps but to date they are not available.

It is not clear why other satellite SST datasets (IR and MW) were not used for the validation. In particular the ARC dataset is an excellent precursor dataset for the CCI SST level 2 and 3 products. The CMUG analysis has concentrated on the comparison with ARC and in-situ data. The microwave SST datasets (TMI, AMSR-E) in particular covered a period (from 1998) when the buoy coverage was still sub-optimal and give good coverage well away from landmasses. However the analyses listed in Table 8-1 of the PVP are probably a better way to do this.

In section 9.2 of the PVP the validation of the level 4 product in a climate modeling framework is proposed. Both the time means (monthly and annual 20 years) and the variability is assessed using HADGEM3. Also engagement with the CMUG is proposed which has been possible through the CMUG assessment of the CCI level 2 ATSR SSTs.

The PVIR describes the validation of the SST product at the end of phase 1 using both in-situ data from buoys and also ship borne radiometers. A lot of comparisons are shown to demonstrate the accuracy of the CCI SST product. The CMUG did a rather more limited assessment of the CCI SST products with similar results in the CMUG phase 1 D3.1v2 report. An important conclusion of the PVIR is that the assessment of the CCI SST product improves from 1991 to 2010 as the buoy coverage improves during this time. An assessment of the CCI data in the early 90’s is more challenging when the sampling of the buoy network was sparser. The PVIR confirms the CMUG results of the slightly higher bias of the CCI SST relative to the buoys compared with ARC except for ATSR-1. The PVIR also includes validation results for the AVHRR CCI SST CDR, which CMUG did not assess, and a dedicated section examining data quality for polar latitudes and showing a better performance of all the products in the Southern Ocean, compared to the Arctic Ocean. The PVIR also compares various SST analyses (level 4) including the OSTIA_CCI with ARGO measurements and shows it compares favourably with other SST analyses. One metric is the ‘sharpness’ of the gradients in the different SST analyses and the OSTIA_CCI does well in this respect.

Uncertainty

For the uncertainty characterization report this is a useful reference document for all ECVs and how to treat the uncertainties. It describes all the different contributions to consider for a complete error budget for satellite datasets. It would be useful to publish this report more widely for the community in general. However this document does not address the validation of the SST uncertainties provided in the dataset even though the underlying principles on how it is derived are sound but this is provided in the PVIR. Useful maps of the confidence in the SST uncertainty validation are shown. The CMUG have also provided an assessment of the uncertainties (see D3.1v2).

Maturity of data

The Climate Assessment Report was made available in January 2014 and is a comprehensive assessment of the SST CCI products. **It is a good model for all other ECV teams.** There is a detailed investigation of the stability of the CCI SST products by comparison with the tropical buoy array and also the GHRSST ensemble of products. The GCOS stability requirement is met in the tropical Pacific and comparable to that of ARC data. The only surprise is the comparison with ARC. A section on using the CCI data to assess the Met Office HadGEM3 climate model is also presented showing it is a suitable dataset for the evaluation of coupled model mean states. The AVHRR data is shown to suffer from intermittent biases possibly due to desert dust contaminating the retrieval. There were also reports by 9 "trail blazers" who used the CCI data. All gave promising assessments of the CCI datasets. Finally a case study compared the consistency of the CCI SST product with ocean colour in one case study showing a comparison of fronts in the ocean off the coast of Mexico. They concluded the data were consistent on daily and 4km scales in the regions of the fronts.

A section on feedback from users was given with an issue of data download speeds being highlighted several users. Also several minor issues with reading the data and treatment of associated flags. CMUG in D3.1v2 have highlighted problems with the time associated with the data which is different according to which depth of the data you are interested in.

The SST CCI team is now extending the SST CDR by processing the AVHRR radiances to derive SSTs prior to the launch of the ATSR instruments in 1991. These data are not available for assessment yet but will not have the accuracy of the ATSR data record. Also work is continuing on improving the ATSR SST CDR developed in phase 1 and a new dataset will be available before the end of CDOP-2.

Validation methods approved by the GHRSST science team and conforming to the guidelines under the QA4EO framework under the CEOS-WGCV will be adopted. This international oversight of the validation plans is to be encouraged by all CCI teams. The GCOS stability requirement is met in the tropical Pacific and comparable to that of the pre-cursor ARC data. However in general for regions of 100km scale an accuracy of 0.1K with the CCI data is not quite achieved being closer to 0.15K. Areas with persistent cloud cover are particularly challenging in terms of achieving accuracy requirements.

There is a document on the Climate Data Research Package which provides an overview of the data produced by the ESA SST CCI project. This gives users a description of the data archive which has a 'Long Term Product' and 'Demonstration Products'. It also gives climate users the links to the data archive[2] from which the data can be downloaded. This is a clear web page addressing formats, citation, documents for users and contact details. There are also a range of related datasets available which can be used for validation of the CCI and other products. It would be good from an integrated CCI perspective that the web site provides direct links to the related CCI datasets such as ocean colour and sea-ice. This would encourage users interested in SST to also look at these.

[2] http://catalogue.ceda.ac.uk/uuid/1dc189bbf94209b48ed446c0e9a078af

2.2 Ocean Colour

Quality

The three Product Validation and Algorithm Selection Reports (PVASR) address the assessment of the results from the round-robin comparisons in two areas. Part 1 is for the atmospheric correction and part 2 is for the in-water retrievals and part 3 covers the identification of cloud/ice fee pixels for which a valid retrieval can be made. No overall assessment of the final product could be found in the validation folder (see Fig. 1) which is a major shortcoming. It seems it is left to CMUG to assess the CCI ocean colour CCI product as reported in the CMUG report D3.1v2. Only Part 1 and Part 2 have been updated for v2 and v3. The v3 version of Part 2 available on the website appears to be a late, but not completely finalised, draft, as it contains a few comments from the authors to themselves and lacks any cover sheet.

/ public / Product Validation and Algorithm Selection Report

Name	Modified	Size
OC-CCI-PVASR-PART1-AC_26-11-12.pdf	13/12/13 15:20	49 MB
OC-CCI-PVASR-PART3-PixelIdentification-26-11-12.pdf	30/01/13 11:30	768 KB
PVASR-PART2-In-water-13-11-2012.zip	15/11/12 10:34	9 MB

Figure 1 Contents of Ocean Colour Product Validation Docs Folder on 15 Jan 2014.

In the outlook of the Product Validation and Algorithm Selection Report it is stated "*The auxiliary meteorology data should be harmonised. For SeaDAS processing SeaWiFS, MODIS and MERIS data NCEP is applied and but MEGS, POLYMER, Forward NN uses the ECMWF data in the MERIS product*". What should have been made clear, is that the ERA-Interim fields MUST be used for the water vapour correction in phase 2 of the CCI for all ocean colour products. This will avoid sudden discontinuities seen in the water vapour field of the operational ECMWF fields. CMUG made this point strongly at the beginning of the CCI project. According to the v2 report it appears that the choice was made to use NCEP instead, which at least is consistent, and the v3 report does not appear to state what was used.

- While validation of the products is a continuous process, there are still concerns with regards to the under-sampling of the in-situ datasets particularly in the low and high productive (chlorophyll-a concentration) regions.

- Originally, the units of the chlor_a_rms_uncertainty and chlor_a_bias_uncertainty variables in the OC-CCI NetCDF files were not clear. These have since been renamed and the documentation improved, so this point has been addressed.

The OC-CCI team's proposition for periodic comparisons of algorithms when there is a significant changes to either in-situ observations or retrieval methods, followed by mission re-processing, is commendable (Ref: UCD). However, there should also be a system (perhaps to be considered by ESA) to archive the previous version(s) with corresponding training and validation dataset to maintain backward compatibility and traceability.

The Product Validation and Inter-comparison Report (PVIR) was first made available in May 2014 focussing on OC-CCI v1, and has been updated in October 2015 to focus on OC-CCI v2. A report focussing on the latest v3 release does not yet appear to be available, so the v2 report is reviewed here. The v1 report comprised a thorough assessment of the data sets used for validation of the newly generated OC-CCI datasets, which included *in situ* observations as well as merged, satellite-derived data sets like GlobColour and MEASURES. The v2 report gives a briefer description of the validation datasets, but references the v1 report so the reader is aware of the information. Included is a description of a new non-publicly available database from IMR, which has been used. Overall, a large database of in situ observations has been compiled and processed, and usefully separated out HPLC and fluorescence chlorophyll.

Most of the assessment focused on chlorophyll-a concentration, because of the lack of observations available for the assessment of other CCI products. Assessment methods are similar to those used in the v1 report, and presented as a continuation. More comprehensive assessment is presented in the CECR, with a figure from that reproduced in the PVIR to demonstrate that GCOS accuracy requirements are being met. Included here is a trend analysis, primarily a qualitative evaluation of the CCI data across the global ocean as well as in biogeochemical Longhurst provinces. The latter is a relatively novel and useful approach. This highlighted increased chlorophyll variability in periods with only one available sensor.

The OC-CCI team implemented a bias correction scheme, that used SeaWiFS as the reference data set, in order to remove biases between sensor-specific data, for each grid point of the global domain and for each band of remote sensing reflectance. The CCI team found there were no obvious biases in the merged, monthly products, due to spurious signals associated with the merging of different missions. These results were presented in the v1 report, and are merely referenced in the v2 report, rather than being repeated with the new data set.

The OC-CCI v2 dataset appears to be consistent with its precursors, including the v1 dataset. Like the v1 data set, the OC-CCI v2 data set was found to be most similar to SeaWiFS derived data, although generally, data records derived from single missions were closer to OC-CCI than other merged mission initiatives, and less biased by the inclusion of data from new OC sensors.

The report highlighted trend signals in the OC-CCI that were (in some cases, but not always) identified in previous scientific studies and called for modelling studies to support the investigation of their spatial and temporal evolution. Particularly large scale, positive trends are seen stretching from the Equatorial Western Pacific to South America and then moving west through the South Pacific and also from the Eastern tropical Pacific to the western Equatorial Pacific. Large scale, negative trends are evident in the North Pacific and in the subtropical Eastern North Atlantic.

Uncertainty

The OC uncertainty characterisation report was assessed in CMUG Phase 1. Since then, two Comprehensive Error Characterisation Reports (CECR) have been made available, v1 and v2. The v2 report is reviewed here, which focuses on the v2 product release. A report for v3 does not yet appear to be available. The report details validation of the products (see previous section) as well as the uncertainties.

The report gives a concise overview of the uncertainty generation procedure, and discusses the reasoning behind the choices made, including the evolution from v1 to v2. The uncertainties are based on the assignment of observations to different water classes, rather than the working through of sources of error which is applied for some ECVs. This decision is based on user requirements. In v1, water classes are based on in situ data, whereas in v2 they are based on remote sensing data. This gives a considerably larger training set, but means that data must be very careful selected, an issue discussed in the report. The CECR demonstrates the v1 and v2 products to be generally similar, but with some improvements due to the change in uncertainty procedure.

In early versions of the Product User Guide (PUG), very limited description was given of the uncertainties, making their use very unclear. In the PUGs for the v2 and v3 products this has been greatly improved, with a clear and logical description given.

Maturity of data

Monthly OC products have been added since the release of OC-CCI v1.0, in December 2013, in Phase 1. In phase 2, two further product versions, v2 and v3, have been released. The most recent version of the CAR is for v2.

V1 of the Climate Assessment Report (CAR) was made available in March 2014. It covered aspects of the algorithm development process, the validation of OC-CCI products (largely covered by the PVIR) and a number of studies on ocean colour decadal variability and applications, such as data assimilation and model skill assessment. The latter studies were based on a combination of precursor ocean colour data (individual SeaWiFS or MODIS products) and no attempt was made to utilise the GlobColour merged data. This is a significant shortcoming as such a comparison would have been a true benchmark of the OC-CCI products (both in terms of reproducing climate variability and exploitation of the data via data assimilation techniques) and highly relevant to the CCI project. The studies presented in this report that used OC-CCI data were based, like mentioned before for the PVIR report, on interim versions (either OC-CCI v0.95 or v0.95.1) not publicly released.

In v2 of the CAR, the above issues have been addressed. Assessment is provided of both the publicly available v2 and v1 products, building on that presented in the v1 CAR. A useful and interesting range of applications is presented, including validation against the GCOS requirements, global and regional trend analysis, assessment of stability and uncertainties, application to phenology and multivariate analysis, and data assimilation. The trend analysis includes comparison with both single sensor and merged precursor data, and demonstrates the benefit of a bias correction approach for climate studies.

Despite the OC-CCI products being specifically developed for Case 1 waters, the data assimilation sections mostly focus on Case 2 water studies. This is an important gap in the analysis. On the other hand, it is encouraging to see the utilisation of the error characteristics of OC-CCI data in data assimilation studies. Furthermore, even in Case 2 waters their use appears to be of benefit for reanalyses. For the v3 dataset, which promises greater accuracy for Case 2 waters, it will be interesting to see how their accuracy compares with v2.

The OC-CCI v1.0 dataset includes two GCOS variables: chlorophyll-a concentration and water-leaving radiances. The GCOS requirements for these two variables, as stated in the update to the satellite supplement to the GCOS Implementation Plan (GCOS, 2011), is for an accuracy of 30% for chlorophyll-a concentration and of 5% for water-leaving radiances. In terms stability, GCOS set their requirements as 3% for chlorophyll-a concentration and 0.5% for water-leaving radiances. The validation of these two OC-CCI v1.0 data products, against *in situ* observations, concluded that the GCOS requirement is met for most of the range in chlorophyll concentrations (except for concentrations lower than 0.1 mg $Chl.m^{-3}$) and for most water leaving radiances (with best results for the shortest wavelength of 412 nm), but slightly missing the GCOS target at longer wavelengths, as the frequency of higher relative errors increases with increasing wavelengths. These conclusions remain true for the v2 products. It appears that GCOS requirements are now being met for the full range of chlorophyll concentrations, although this does not seem to be explicitly stated. The subtropical gyres, where chlorophyll concentration is typically very low, and the highly productive coastal waters, where chlorophyll tends to be very high, are the geographical regions that exhibit the largest relative errors.

Whilst not GCOS variables, there is a growing user requirement for products such as phytoplankton functional types (PFTs). It is encouraging to see the OC-CCI team considering this issue, and presenting an initial demonstration of such an application.

In the first sentence of section 3.1 of the CAR, it states that SeaWiFS and MODIS are from ESA and that MERIS is from NASA: these affiliations should be reversed.

2.3 Sea Surface Height

Quality

Four documents are addressing the quality and uncertainties of the SL CCI products.

The v2.1 version of the ATBD allows knowing what are the algorithms that are used to produce the v1 version of the SL-CCI products (gridded and mean Sea level) and the v3.3 version of the ATBD describes those applied for the development of the v2.0 of the SL-CCI products. They allow addressing the improvements that have been brought in the different steps of the derivation of the geophysical parameters from the original satellite observation (new orbits, new wet troposphere correction, new tide model for the calculation of the tide contribution to sea level change …). However the overall quality and uncertainty are discussed in other documents.

In the PVIR published at the beginning of phase II (September 2014), the authors certify the end-to-end quality of ECVs and analyse the total contribution of improvements done in the final products (v1.0 version of the SL-CCI) by comparison with the products existing before the beginning of the project (from AVISO). The comparison to in-situ observations (tide gauges and ARGO) reveals the differences of variability between satellite and in-situ observation at different time scales, but also some improvements achieved with the new products (in particular for interannual variability). An east-west difference of trends between the Sea Level Anomalies derived from the two satellite datasets over the 2003-2010 period, attributed to a change in the used orbit model, remains however intriguing for a non-specialist. However the impact of the orbit is well addressed in the error report document (see below the “Uncertainty” sub-section). One interesting addition to the PVIR compared to the one published at the end of phase 1 (v1.0 dated from the September 2013), is a list of recommendations aiming at improving the overall quality of the SL CCI.

Additional improvements are described in the PVIR published in December 2016 that is focused on the comparison of the gridded SL-CCI (v2.0) with the previous version of the product (v1.1). The main part of the above-mentioned recommendations, have been implemented and some of them have a significant impact. This is the case for the inclusion of the treatment of new satellite altimeters (Cryosat-2 and SARAL/Altika) that allows to better address sea level change at the highest latitudes and, combined in particular with the use of a new tide model, has an effect on the calculated sea level variance. Another significant impact comes from the use of the same wet troposphere correction algorithm for all the missions that affects the GMSL (Global Mean Sea Level) decadal variability. A slight improvement of the reproduced seasonal cycle of the sea level when compared to in-situ observations is also noted.

The final version of the SL PVP was already examined in the CMUG Phase 1 D2.2 report. One comment suggested a more regional (basins and latitudinal bands) evaluation and ocean-process based evaluation to complement the aggregated diagnostics proposed in the PVP document.

This is achieved in the June 2014 version of the CAR (v1.2) which presents a set of scientific evaluation of the products, several of them being focussed on regional assessments (regional mean sea level trends …) and process-oriented (steric mean sea level …). In this version of the document some more emphasis is given on the ability to reproduce the trends in the Arctic ocean with the CCI sea level. This analysis is developed in the last version of the CAR (v1.3) thanks to the addition of a comparison to models assimilating in-situ observations and a focus on key sub-regions of the Arctic Ocean (Lofoten Basin and Sub Polar Gyre). One important conclusion of the new regional assessment included in this document is that the results of the assimilation of the SL-CCI in the GECCO2 ocean synthesis framework don't allow detecting an improvement or degradation of the regional trend pattern between AVISO SL product and the v1.1 version of the SL-CCI. However, the assimilation experiments demonstrate the improvement of the SL products from AVISO to SL-CCI v1.0 and to CL-CCI v1.1 in the Tropical regions, when analysing the consistency between the assimilating model and the data. The quality of the SL-CCI is also checked in this document through an intercomparison of the corresponding GMSL with other existing GMSL products over a common period extending from 2005 to 2014. The new CCI data (v1.1 in this document) lead to the best closure when they are compared to the addition of the contributions to the GMSL (steric and mass change) estimated with independent observations (ARGO and GRACE).

Uncertainty

The Error Characterization Report document allows having a clear view of the impact of the orbit model on the trends and on its errors. The improvement of the hemispheric trends is well established and the contribution of the orbit model to the uncertainty is evaluated through a clear reasoning (comparison of different orbit solutions). This report also deal in the same way with the other sources of uncertainties giving the main basis of the estimates for each contribution: comparison between microwave radiometers on-board altimetric satellites and analyses from atmospheric models for the contribution of wet troposphere corrections; cross-comparison between global mean wind speed derived from altimetry with atmospheric reanalysis and intercomparison of tide models for the altimeter instrumental parameters contribution; the analysis of overlapping periods of measurements when available to estimate the contribution of the biases linked to the altimeter missions. The mention of published references allows going further in the understanding of the uncertainty evaluation process. The assumptions that are done to combine the different sources of uncertainties are also well stated.

Other ways to check the errors is also presented for the GMSL. The first one consists in the comparison between the GMSL derived from the altimetry and the one derived from ARGO profiles and tide gauges. This gives results consistent with the total error estimates for the long-term trend of this specific parameter. Another approach, applied to the Jason periods, allows the determination of a confidence envelop for the GMSL as a function of time and to separate the uncertainties due to different varying parameters. It consists in analysing an ensemble of GMSL time-series of a priori equivalent qualities generated by tuning four identified parameters of the generation algorithm. This analysis shows the predominance of the selection of standards (tide model, reference mean surface …). An important new result presented in the last version of the report is a map of regional trend uncertainties which calculation is not detailed and not yet published. This map has already been confronted with the map of the spread of sea level trends obtained from an ensemble of model reanalyses (ORAS5), showing consistent estimate and geographical variability (according to the CAR).

The authors highlight the fact that the resulting total errors, at the global and at the regional scales, are often higher than the user requirements identified at the beginning of the project. This concerns more precisely the long-term trend total error (stability) that is estimated to be lower than 0.5mm/yr for the global mean sea level and lower than 3mm/yr for the regional mean seal level. The corresponding GCOS (and CMUG) requirements are respectively lower than 0.3mm/yr and lower than 1mm/yr. However, it appears that this was not a limitation in the analysis done by CMUG over the Mediterranean basin since the new estimate of the uncertainty of the trend over this region appears to be lower than 3mm/yr when looking at the above-mentioned uncertainty map (to check further).

Some ways of reducing the uncertainties are mentioned at the end of the document including a TOPEX data reprocessing, new orbit solutions and improved wet troposphere corrections.

Maturity of data

The user assessment presented in the successive version of the CAR is a representative illustration of the use of the data by the scientific community. It covers its common use for ocean model assimilation and for trend analysis. In particular, the agreement between the GMSL temporal variations simulated by the ECMWF ocean reanalyses and observed though the SL-CCI GMSL product, allows having some confidence in the attribution of the sea level change to steric or mass changes that are inferred from the model diagnostics. The analysis of sea level in Sub-Polar Gyre demonstrates the ability of the NorCPM model assimilating SST to reproduce the observed interannual to decadal variability over the last decades. All these application illustrate the potential use of the data by the climate modelling community over a very wide range of scales, from sub-regional to global.

The comparison between the SL-CCI dataset (v1.0) and the SL AVISO dataset is done all along the version of June 2014 of the CAR. The results show that the largest differences in interannual variability are located along the sea-ice edge, in the ITCZ region and in regions of large eddy activity like in the western boundary currents, but with very good agreement elsewhere. Trends over the last decades are different at the hemispheric scale and in some specific regions like in the equatorial Pacific. The analyses conducted by the CMUG group over the Mediterranean area are consistent with these findings but with closer agreement between the two products over this specific region. As detailed in the sub-section on "Quality", the CAR published in August 2016 and the PVIR published in December 2016 complete this view of the improvement from AVISO products to the v1.1 and then v2.0 versions of the SL-CCI products.

These improvements made along the two phases of the project clearly reinforce the maturity of the dataset. The ways in which GCOS requirements are met are described in the last part of the "Uncertainty" sub-section.

2.4 Sea Ice

Quality

The Product Validation & Intercomparison Report (PVIR), the Product User Guide (PUG) and the Climate Assessment Report (CAR) of this CCI's activity all contain an assessment of the quality of this data record.

In general, however, this CCI provides a far more detailed discussion and quantification of error sources than comparable projects did before. This is obviously highly relevant for the climate research community, where quantitative estimates of observational uncertainties are necessary for any meaningful use of observational data.

As part of the sea ice CCI, climate data record of both sea-ice thickness (SIT) and sea-ice concentration (SIC) have been developed. The former record is only available for the Northern hemisphere during winter, while the latter is available for both hemispheres all year round. Both records have a relatively large number of sources for observational uncertainty that are very coherently discussed in the PUG.

For sea-ice thickness, there are four main error sources as summarized in the PUG: First, for the transformation from freeboard (i.e. the amount of snow and ice above the water level) to actual ice thickness, the snow thickness must be known. Since this information is currently not available from measurements, instead a climatology is used that no longer seems to be valid in many places. Second, it is unclear whether the radar signal actually stems always from the snow-ice interface, as assumed for the calculation of ice thickness, or whether it sometimes primarily stems from within the snow pack. Third, there is speckle whose influence can be reduced by averaging over a larger number of measurements, which then of course reduces both spatial and temporal resolution of the record. And fourth, the large footprint of the radar which might cause a mis-representation of actual ice-thickness distribution.

The PVIR contains an extensive quantification of possible biases of the CCI SIT product relative to independent products. Compared to ground-based in-situ measurements and EM measurements, the CCI SIT product over-estimates ice thickness by typical 0.5 m to 1 m, and even more in extreme cases. The root-mean square error relative to airplane measurements during the Operation Ice Bridge is typically close to 1 m, too. Compared to this data record, the biases are however both positive and negative, in contrast to the EM and in-situ measurements, where CCI SIT data usually shows too large ice thickness. Compared to data from upward looking sonar (ULS)Th, the SIT record overestimates ice thickness by typically about 0.75 m and does not show the significant thinning of the ice that is displayed by the ULS data.

There is no systematic description of the individual contributions of the possible error sources to these discrepancies, which would be very helpful for any user of the data. The very large biases shown by this SIT record make the record currently not suitable for standardized use for either model evaluation or model initialization. Hence, the usefulness of this data for the climate-research community is currently somewhat limited. A respective warning should possibly be included more clearly in the user guide. A better characterization of the most likely underlying error sources is also desirable. Nevertheless, the fact that these uncertainties are quantified as part of the sea-ice CCI is very advantageous for any user of the data.

For sea-ice concentration (SIC), there's also a very nice description of possible error sources in the PUG. During summer, the main limitation of the accuracy of this record is related to melt ponds that form on the sea-ice surface. These ponds are always seen as open water by any passive microwave (PM) instrument, which causes a systematic underestimation of sea-ice concentration during summer. The accuracy of PM retrievals is also limited for thin ice of less than 30 cm thickness, because then a significant amount of the retrieved PM signature stems from the underlying water, which then causes an estimate of SIC that is lower than the actual value. The influence of land pixels also deteriorates the PM signature of sea ice, which is why near-land pixels are excluded from the entire record, in contrast to other existing records. The PM signature is also influenced by the atmosphere, which hence is an additional error source. Other systematic errors are caused by smearing out of the retrieved signature on scales smaller than the satellite footprint.

Uncertainty

During phase I of SICCI, a Comprehensive Error Characterisation Report (CECR) was compiled that summarizes the main sources of uncertainty in remotely-sensed sea-ice concentration and sea-ice thickness.

For sea-ice concentration, the report gives a comprehensive overview of errors that are caused by physical limitations of the retrieval method and of errors caused by the numerical analysis of retrieved physical parameters. It does primarily summarize previously known results for a variety of different sensors and does not put these into the concrete context of SICCI. As such, the report is of somewhat limited use in assessing the overall uncertainty of the sea-ice concentration that is derived from this particular project, and is primarily useful as an overview of pros and cons of various sensors that can be used for estimating sea-ice concentration.

For sea-ice thickness, in contrast, the report provides some new estimates of uncertainties for the various methods that can be used to assess sea-ice thickness from space. As such, the report goes beyond published work, but again there is no direct reference to the sea-ice thickness work that has been carried out within SICCI, which again limits the usefulness of this report for the end-user of SICCI data. One would have wished for a concrete discussion of how the sea-ice thickness product that has been compiled within SICCI is limited by the various factors that are discussed here.

In reading the entire CECR, a clear imbalance between the treatment of SIC and SIT is apparent, with the former not even being mentioned in the conclusion section. It is obvious that both main authors of this report were primarily concerned with SIT during phase one of SICCI, which is why almost three times as many pages of this report are dedicated to SIT compared to SIC.

The uncertainty estimate that is contained in the actual SIC product itself provided by the SIC record only contains an estimate of those uncertainties that can be quantified for example by estimating differences between several over passes of the satellite. They do, however, not contain estimates of the uncertainty of sea-ice coverage caused by the existence of melt ponds, which is the main error source during summer. It would be very helpful for users of this record if at least a rough estimate of this uncertainty was provided, for example based on a climatological record of melt-pond occurrence from MODIS data.

For the uncertainties that can be quantified, the quality of the SICCI record is very high. In particular, the detailed description of the various error sources through the year and their inclusion into the actual record is a clear advantage of this record over existing ones. The quantifiable uncertainty of retrieved sea-ice concentration for 100 % ice coverage and for open water are usually below 3 %, which is clearly an acceptable value for climate research, in particular given the much larger influence of the unknown error sources such as melt ponds and the thin ice fraction. Roughly the same value is obtained for thick ice of intermediate concentration by comparison with high-resolution Landsat images. Including thin ice gives twice as large an uncertainty of around 6 %, which can be too large for some applications, in particular those related to very low sea-ice concentration and initial ice formation.

In the Southern hemisphere, the SICCI record provides good agreement with the ASPeCt data set, though the Comiso Bootstrap algorithm performs slightly better.

The description of these uncertainties in the PVIR is sufficient and very helpful. No other existing satellite product has an even remotely similarly extensive assessment of quality. It might be helpful to summarize these findings briefly in the PUG such that users get a short quantitative overview of possible error sources.

In addition, the sole inclusion of the quantifiable uncertainties might pretend a smaller uncertainty than there actually is. This should clearly be mentioned in the PUG. Otherwise there's a chance that users underestimate the possible error range of sea-ice concentration based on the uncertainty estimate that's included in the record, which does not include the sometimes large impact from melt ponds and thin sea ice on the total error.

Maturity of data

For sea-ice CCI, the available data of sea-ice concentration and of sea-ice thickness have been compared to existing products to assess the maturity of the data. These exercises are described comprehensively in the projects Climate Assessment Report (CAR).

For sea-ice thickness, the estimates of SICCI have a clear bias compared to existing remotely-sensed data sets and also compared to existing in-situ data. For the modal and for the median ice thickness, SICCI estimates of ice thickness are on average around one meter larger compared to existing data. Since a similar bias exists similar to independent in-situ data, it is very likely that this bias primarily stems from the SICCI product itself. Hence, this data is not mature for usage in climate-related applications.

For sea-ice concentration, the anomalies of both main integrated quantities sea-ice extent and sea-ice area agree extremely well with existing estimates. However, regionally substantial differences of more than 20 % of ice concentration are not unusual, which are, primarily, a reflection of the different philosophies that are taken to compensate for physical limits on the retrievals. For example, while SICCI sea-ice concentration only describes the concentration of melt-pond free sea ice, other algorithms attempt to use an ad-hoc bias correction to compensate for the melt-pond covered sea-ice in their estimate of sea-ice concentration. Since these corrections cannot be physically based, their usefulness for climate-related applications is limited and the approach taken by SICCI is much preferred. The accuracy of SICCI retrievals against the independent data of the Round Robin data package is extremely high, and matches the independent data to better than 1 % in open water and to usually better than 3 % for 100 % sea-ice coverage.

Regarding GCOS requirements, both SICCI SIC and SIT only meet the requirements partly. Regarding resolution, neither SIC nor SIT meet GCOS-requirements. These are for SIC 10-15 km, which, however, is simply not possible with the passive-microwave frequencies used within SICCI that limit spatial resolution to 25 km. For SIT, GCOS requires 25 km resolution, which again is a higher resolution than the 100 km resolution delivered by SICCI.

The GCOS-required accuracy of 5 % for SIC is met by SICCI, at least if one accepts that melt-pond covered sea-ice simply cannot be detected as sea ice from passive microwave. For SIT, a bias of around 1 m is clearly incompatible with the GCOS requirement of 0.1 m accuracy.

The temporal resolution as required by GCOS is met by SIT with its monthly resolution, and exceeded by SIC, where GCOS requires weekly data while SICCI delivers daily data.

There is no analysis of long-term stability of either SIC or SIT within SICCI, which should be addressed in phase II of that project.

2.5 Clouds

The Comprehensive Error Characterisation Document and the Product Validation Plan (PVP) have been assessed in Phase 1. Here, we assess the updated version of the PVP from June 2013, the Product Validation and Inter-comparison Report (PVIR) from December 2013 and the Climate Assessment Report (CAR) from November 2013.

Quality

The PVP contain detailed summary of the different validation datasets and a general outline of the comparisons. The updated report includes additional validation datasets, the SSM/I LWP dataset and ground based station data, the Alpine Surface Radiation Budget (ASRB) and Baseline Surface Radiation Network (BSRN) for polar stations. This can improve the validation over high altitudes and snow covered regions.

In PVIR extensive validations of the two Cloud-cci retrieval algorithms, the Community optimal estimation Cloud retrieval For Climate (CC4CL) for Modis, AVHRR and AATSR and the FUB AATSR MERIS Cloud retrieval (FAME-C) have been made. The data was compared to measurements from geostationary satellites (MSG-SEVIRI) and active sensors (CloudSat-CPR, CALIPSO-CALIOP) and to existing cloud climatologies (CM-SAF CLARA A1, MODIS, UWisc MWR).

For the level-2 comparisons, the active A-Train sensors were used to validate the CC4CL cloud mask. The algorithm performs similarly well as comparable existing algorithms. However, CC4CL overestimate clouds in twilight conditions, underestimate cloud amount in tropical regions and underestimate Cloud Top Height (CTH). These issues need to be addressed, does e.g. the twilight problem affect the observed diurnal cycle?

For FAME-C collocation of ENVISAT and A-train data is only possible at high latitudes, case studies showed that FAME-C also tend to underestimate CTH. A summary table for the level 2 data with the validation results organized according to “external” parameters such as surface type, solar zenith angle as described in the PVP would be useful. This would complement the risk flags for the retrievals that include information about underlying surfaces, sun glint, cirrus and multilayer cloud occurrences as described at the end of the PVIR.

The level-3 results are well summarized in Table1.1 with colour codes indicating if the bias and std fulfil the GCOS goals. The data was compared to SYNOP and existing cloud climatologies (CM-SAF CLARA A1, MODIS, UWisc MWR LWP).For cloud cover and CTH both algorithms, CC4CL and FAME-C perform well on a global average scale. For cloud water paths the values are outside the GCOS thresholds. As pointed out in the text, the Cloud CCI three year of data can overestimate the variability, which in addition to some known issues and bugs that were found in phase 1 affects the results. In phase 2 some of these issues should have disappeared. In addition to Table1.1, summary geographical maps showing the regional variation in bias, std and uncertainties would give useful information for developers and users.

Uncertainty

The Comprehensive Error Characterisation Document was assessed in Phase 1, it provides a description of how the errors are determined (e.g. it outlines the relevant equation from the ORAC algorithm). Alternative ways to derive the uncertainty for level 3 products from L2 error are discussed in the CECD/CECR(?). An update on which method that was chosen would be of interests.

Quantitative information on the errors of the CCI cloud products are given in PVIR. For phase 1 problems were detected for the cloud water path, it was found that the error/uncertainty on a global scale is beyond the GCOS requirements for all datasets, with positive biases between 0 and 80 g/m2 and bias-corrected rmse values between 50 to 100 g/m2 (CC4CL datasets) and up to 140 g/m2 (FAME-C). Problem regions, such as the Arctic was also noted in the uncertainties with increasing values polewards of +/-65°Lat.

According to the PVP, the CCI Cloud uncertainties were planned to be used in the validation activities. This should be pursued in the phase 2 validation exercises investigating if the differences compared to other datasets are reflected in the uncertainties. The uncertainties as described in the CECD/CECR(?) are for the optimal estimation retrieval, but uncertainties associated with the cloud mask are not discussed. Are there any plans or ideas how they can be considered too?

Maturity of data

The Cloud-CCI L3 datasets fulfillment of the GCOS requirements are reported in PVIR (Table 1-1) as summarizes of the results of the L3 validation of the datasets compared to the GCOS accuracy requirements. Cloud cover and cloud top height meet the requirements for bias for three out of four instrument groups and partly fulfill them for the standard deviations. For cloud water path, most values are outside the GCOS thresholds.

As stated in the PVIR a number of improvements of the retrievals have to be investigated in Cloud_cci phase 2. “In particular the correct identification of aerosol and cloud contaminated pixels and improved assignment of the cloud phase would greatly improve the results. An IR only retrieval to improve cloud top height could also be investigated.” Additionally, it is noted in the PVIR that the GCOS requirements are goals for long-time climatologies and in Cloud CCI only three demonstrator data sets have been produced. Once phase 2 decadal data is available, a new assessment can determine the maturity of the data.

An important point made in the PVIR is that validation of cloud (or any other) properties should not be based on a single dataset of in situ observations, but on a collection of data sources (i.e. SYNOP, APCADA, and active remote sensing), which would allow to provide the estimation of the uncertainty of the validation data, similar ideas as put forward by CCI SST for in-situ data.

In the Cloud CAR, the data products for 2007-2009 have been compared to existing global satellite datasets, reanalysis data and regional and global climate model simulations.

The GEWEX Cloud Assessment cloud products global datasets (Stubenrauch et al. 2013) and ERA-Interim data (Dee et al 2011) have been used to assess the cloud-cci level 3 data. The results for phase 1 show that the regional and seasonal variations in the cloud-cci data correlate well with GEWEX Cloud Assessment data base and with ERA-Interim reanalysis data. However, there are problems detecting high cirrus and some issues over challenging surface property regions (mountain and sea-ice), similar problem regions exit for many other satellite datasets (Stubenrauch et al 2013). These issues still need to be addressed in phase 2.

The Cloud CCI data sets have also been used for evaluating regional and global climate model simulations. The regional climate model, COSMO, have optically thicker clouds compared to cloud-cci data. Sensitivity experiments changing the ice sedimentation in COMSO reduced this bias. However, due to the Cloud CCI phase 1 problem detecting high clouds, these experiments should be repeated for the re-processed data and using several satellite datasets. For the global climate model EC-Earth, sensitivity experiments were made w.r.t. mixed-phase clouds. The results were compared to

Cloud CCI data including the uncertainties, which were useful in showing larger values over problem surfaces, thereby guiding the user on the quality of the data.

The exercises in CAR summarised above should be repeated once longer data-series are available in phase 2 and in addition, extended use of the uncertainties are encouraged. To have an impact on the climate modelling community outside CCI, longer data-series are a requisite. The processes for achieving the climate-based goals are identified, the products partly meet the GOCS requirements but due to the short time-series we judge the maturity of the Cloud-CCI data to be partly fulfilled (Table 2).

2.6 Aerosols

The Product Validation Plan (PVP v1.5) and the Uncertainty Characterisation Report (UCR v 1.2) were assessed in Phase 1 of CMUG. The Product Validation and Inter-comparison Report (PVIR) and the Aerosol Climate Assessment Report (CAR) were released by the Aerosol_cci team in March 2014 at the time when the CMUG Phase 1 Technical note on Validation and User Assessment was submitted. The new versions of PVIR and CAR are assessed here, together with the Comprehensive Error Characterization Report (CERC), which represents an update to the UCR.

Quality

In Phase 1 of ESA CCI, the Aerosol CCI project focused on algorithm improvement and comparisons of different algorithms. Several algorithms for aerosol optical depth (AOD) went through a three-step process: algorithm experiments (Holzer-Popp, et al., 2013), round robin exercise (de Leeuw, et al., 2013), and ECV production and validation (Popp et al., 2016). The extensive efforts on algorithm development, testing, and comparison have been a major step forward in Phase 1. In Phase 2, the round robin exercise was repeated for ATSR (3 algorithms, full mission time period 1997-2011, global) and firstly applied to IASI (4 algorithms, year 2013, dust belt region).

The PVIR summarizes the validation results of ATSR and IASI products produced with the various algorithms participating in the ESA Aerosol_cci. Ground-based AERONET and MAN station data were used as reference for the validation of the satellite retrievals. These were complemented with model climatologies to characterize regional biases. Furthermore, an advanced retrieval algorithm based on POLDER data has also been developed within the project, but applied only to selected regions given its high costs for operational use.

The validation was conducted using several metrics on both Level 2 and Level 3 data, and distinguishing between continents an ocean. The performance of ATSR was further assessed in comparison to the NASA retrievals from MODIS, MISR and SeaWiFS. A special attention was devoted to the China region, were a further validation was performed based on CARSNET data. This is valuable given the importance of this region for anthropogenic aerosol and aerosol trends. For ATSR, the SU 4.21 algorithm was found to perform slightly better in all examined cases (especially in its Angtröm exponent and estiamted uncertainties over land), with the exception of high AOD ocean regions, where the ADV 2.30 algorithm performs better. This basically confirms the conclusion of Phase 1. For IASI, the ULB and LMD algorithms were found to perform best among the 4 tested, but further analyses are still ongoing. The PVIR also reports on the GOMOS data on stratospheric AOD, which gained some importance recently due to several major volcanic eruptions occurring in the last years, and on the 35-year AAI dataset constructed by combining 5 individual sensor dataset after a considerable calibration effort.

At the end of Phase 1 the Climate Research Data Package (CRDP) was produced which contains the latest / best and most complete and validated datasets of Aerosol_cci. The CRDP is openly available at http://www.esa-aerosol-cci.org/, with datasets covering at least the golden year 2008.

Uncertainty

Uncertainty estimates are summarized in the CECR, which provides a very valuable overview of the different types of uncertainty in the datasets and the way to characterize them.

The uncertainties are estimated at Level 2 (pixel level). For AOD (ATRS) the CECR accurately describes the different techniques used by the 3 algorithms (selected in the round-robin exercise) to quantify the uncertainty for this instrument. An attempt to homogenise the techniques is planned within the project and will be reported in future versions of the CECR. For dust AOD (IASI), the error estimation methods are given for the 4 algorithms, but users shall be aware that the estimates are not mature yet since the algorithm are still in an early development stage. Errors are also briefly estimated for the stratospheric aerosol data of GOMOS.

The issues related to the uncertainty estimate for L3 data are briefly summarized and five metrics are proposed. A preliminary evaluation of such metrics is presented based on AERONET, but a more robust technique is currently being developed to refine the results. The general results of the product validation against AERONET are also extensively discussed in the PVIR.

Finally, a valuable list of guidelines is provided to the user on how to make best use of the products.

Maturity of data

Aerosol_cci data for AOD generally meet the GCOS target requirements. The horizontal resolution of 10x10 km^2 is in line with the requirements, although the provided daily temporal resolution is coarser than the requested 4 hours. The validation of satellite AOD products against the ground-based AERONET data shows a bias up to 0.01 and rmse/noise up to 0.08 over land: the GCOS requirement of 0.03 (10%) is met by 62% of pixels. The accuracy over oceans is lower, around 20%, but the validation in this case is more difficult due to the limited amount of ground stations over the oceans. Additional products were provided for the Angstrom parameter, the aerosol absorption index, aerosol types, mixing fractions and effective radius, and stratospheric extinction. The latter is provided at 0.5 km vertical resolution (well within the GCOS requirements of 1-2 km) and 2.5°x10° (lat, lon) horizontal resolution, which is close to the 100-500 km requirement. The accuracy has been quantified by validation against OSIRIS and SAGE-II, resulting in a height-dependent bias of 10-25%, again close to the required target of 10%. Currently, there is no effort on aerosol precursors (e.g., SO_2). This would be a valuable addition for climate studies, in particular concerning the estimate of emissions.

More recently, a almost 17-year (7/1995-3/2012) ATSR aerosol climate data record has been provided by the Aerosol_cci. This new data record for mid-visible AOD (550nm) and Angstrom parameter is competitive to commonly used data records of NASA sensors (e.g. MODIS, MISR, SeaWIFs) but goes further back in time until 1996 (note that Angstrom parameters from all ESA and NASA datasets show major weaknesses). The data are released in netCDF format and are freely available through the Aerosol_cci website (http://www.esa-aerosol-cci.org/). The data record involves data from two similar sensors on two consecutive platforms (ATSR2 1996-2003 on ERS2 and AATSR 2002-2012 on ENVISAT). The global annual average AOD (at 550nm) is estimated at 0.15 and annual Angstrom parameter is near 0.77. Global averages of both parameters have not changed significantly over the entire record, despite regional shifts of maxima. Most significant are AOD increases over Arabia (dust) and both southern and eastern Asia (pollution) and AOD decreases over South America (biomass) and dust outflow off Northern Africa into the Atlantic (Kinne et al., 2014).

The CAR demonstrates the maturity of the data, which can be used for several applications, including assessment of AOD trends (given the long time coverage of the AOD record), input to global models for simulating stratospheric aerosol extinction, investigations of aerosol-clouds interactions and their representation in global models (since ATSR can also retrieve cloud properties), and of aerosol direct radiative forcing effects.

In addition, the consistency between the cloud masks used by the Aerosol_cci and the Cloud_cci projects has been analysed in the ESA Climate Change Initiative Aerosol_cci / Cloud_cci Cloud Mask Consistency Analysis Report: "*Overall 0.3% of observations have been found to be inconsistent while about 21% are not used for aerosol or cloud retrievals at all owing to missing reliability. Over land 1% of observations are inconsistent while inconsistency is practically absent over ocean. On the other hand over ocean only 5% of observations are definitely cloud-free and thus used for aerosol retrieval, strongly impacting on the coverage of aerosol data over ocean. Remaining open issues are especially the 1% inconsistent observations over land, where thick aerosol plumes and broken cloud fields have been identified as possible reasons. Moreover the analysis so far contains five days in September 2008, which have been selected mainly to especially cover difficult scenes (with high aerosol loads or complicated mixtures between aerosol and clouds). Thus it has not yet been proven that the results are representative globally for all seasons.*"

2.7 Greenhouse gases

Since the publication of the last CMUG Report on Product Assessment, the GHG-CCI has released Climate Assessment report (CAR, version 3), and Product Validation and Inter-comparison Report (PVIR, version 4). All other documents had already been reviewed.

Quality

The PVIR describes the validation of the XCO2 and XCH4 products that are part of the Climate Research Data Package version 3 (CRDP#3) generated by the end of year 2 of phase 2. Compared to the v3.2 PVIR, the assessments of GOSAT OCFP and of the ensemble-based (EMMA) XCH4 products have been added.

The validation made use of in-situ data from the TCCON network and modelled fields. The assessments are designed to characterize the achieved quality with respect to both the GCOS requirements and those specified in the URD, and the effort in summarizing the amount of information derived from the various assessments.

The document presents many methods to assess the same product (e.g. the random error) and shows in some cases comparable results that indicate a robustness in the derived product.

The rather sparse coverage and limited number of the available validating observations still represent a limitation.

Compared with the previous data package (CRDP#2), the temporal coverage of both XCO2 and XCH4 has been extended forward in time using the GOSAT observations and now cover the period 2014. Also compared to CRDP#2 results, CRDP#3 shows some improvements in the error characterization. The achieved quality show that the random error of all products but the SCHIAMACHY XCH4 are well within the requirements. Good improvements have also been achieved in reducing the systematic error component so that the requirements are met or close to be met for a larger number of products than in the previous version.

Attention was paid in extending the stability characterisation to all products. Results show that the stability requirements are met for all products. As mentioned in the previous report, it would help if some of the PVIR plots were presented in terms of the TCCON – GHG CCI residuals rather than in absolute terms overlapping results for different algorithms, this would facilitate the comparisons and understand the differences. Figure 6.2.1.4 does not appear in the pdf file.

The v3 CAR became available in May 2016. It is a comprehensive assessment of the GHG CCI products. Compared to the v2 CAR, the assessment of the aerosol impact on the XCO2 retrievals was further extended. The study suggests that a better the aerosol size distribution can improve the quality of the total column retrievals by reducing their bias against in-situ measurements. This result shows potential to further improve the XCO2 products and would suggest possible synergies with Aerosol CCI that could be perhaps exploited.

Differences between available models in estimating the regional carbon budgets are still acknowledged for both XCO2 and XCH4 products, although comparisons suggest some convergence, at least for the former.

Uncertainty

This aspect is covered in other documents (e.g. in the PVIR) and summarized above.

Maturity of data

The documents, particularly the PVIR, clearly show the level of maturity achieved by the data. In particular:

- Most of the data requirements on error characterization, bias, and data stability have been achieved with the CRDP#3 products, at least with respect to the threshold value. Although, it is noted that some of the requirements cannot be achieved with the current EO missions (e.g. the 4h temporal resolution required by GCOS)
- The products from different algorithms have generally achieved a good level of consistency between them.
- Some problems are still found in reconciling regional estimates from different models, though some convergence could be seen especially compared to CRDP#2.

The continuous investigation into possible factors that could explain the differences between the CCI GHG retrievals and TCCON data is very valuable to determine incremental corrections and future algorithm modifications.

2.8 Ozone

Since the publication of the last CMUG Report on Product Assessment (v1 May 2015), the O3-CCI has released a Comprehensive Error Characterization Report (CECR, version 2.01.01) and version 2.0 of the PVIR, which are both discussed below.

Quality

The Product Validation and Inter-comparison Report (PVIR) - published in June 2016 - presents an in depth summary of the assessment of the various core O3 products (total column ozone, and nadir and limb-instrument based ozone profiles). The assessment was performed bearing in mind the O3 User Requirements (URs) that concerned with three main points: 1) accuracy of the error bars (consisting in the random and systematic components); 2) temporal and spatial domain and resolutions met by the data, including long-term stability; and 3) other user requirements (e.g. on format, metadata, and visualization tools). These URs account for the GCOS requirements and those independently characterized by the O3-CCI and CMUG.

The latest PVIR document refers to an extended data package that includes for instance the GOME-2B and OMI total column ozone (TCO3) data among the nadir instruments. Tropospheric O3 product based on the usage of nadir and limb data is also available. Compared to the previous PVIR document the GOME-2A TCO3 agreement with in-situ observations seems to be deteriorated in places (fig. 5 pag 27 vs. fig 5 pag 24 of PVIR_1.0). It is not clear what drives this result as no data version is given while the referred algorithm version appears to be the same.

One limitation is that the assessment heavily relies on comparisons with in-situ observations, which are generally sparse and can offer good coverage generally over Europe and the USA. It is also shown in some cases that depending on the in-situ instrument (Brewer vs Dobson) different levels of agreement are achieved, posing some questions on the validating dataset.

Uncertainty

The CECR v2.01.01 document was released in May 2016. It provides a summary of the error characterization of Level 2 and Level 3 ozone products mostly based on information already given in the ATBD, which was reviewed in previous versions of the present document, and a number of papers on the validation of the error budget in the Level 2 ozone retrievals.

The uncertainty characterization is presented for the three main lines of production (total column, ozone profiles from nadir and limb sensors) and for the tropospheric ozone column. The description of the error characterization is the same presented in the previous version of the CERC report (v1.01.01) and discussed in the 2015 CMUG D2.3 document, thus not repeated here. The current report includes a new section discussing Observing System Simulation Experiment (OSSE) that is used as a new, alternative method to assess the random component of the uncertainty. The OSSE method was applied to the validation of the CCI total column ozone product and suggested that while the systematic errors appear negligible, the previous results presented in the literature were too conservative, proposing random error levels of about 0.7% in the equatorial regions and about 1% in the extra-tropics.

It would help the reader if the data version was clearly mentioned at the start of each section. Please note that the ECMWF web-site is www.ecmwf.int and not www.ecmwf.eu as written in page 25. In addition to a few typos, please also note the following:

- The reference to figure 20 in page 23 should point to figure 19.
- The final sentence of page 31 misses the subject, which I assume refers to the table that follows.

Maturity of data

The URs considered in the ozone products assessments take into account the GCOS requirements and those independently characterized by the O3-CCI and CMUG. It is important to notice that the GCOS requirements do not distinguish between nadir and limb profiles as done by O3-CCI, and it should be appreciated that meeting some of the requirements might be limited by instrument design and characteristics.

The assessment of the long-term stability is interesting. The TCO3 products shows high level of consistency and high inter-sensor stability, and generally meet the UR of a long-term stability better than 1%/decade. For the nadir products, insignificant long-term drift has been achieved only for the ERS-2 GOME product. The limb retrievals all show a drift (values depending on the product) in the lower stratosphere, and a negligible to small drift in the middle and upper stratosphere.

The URs for TCO3 are generally met in terms of the error characterization and data stability but for criteria like temporal coverage and horizontal resolution they are met for some users only. Meeting the URs is still largely work in progress for the nadir profile products, particularly in the troposphere and middle atmosphere. For the limb data, the URs concerning spatial resolution, and time coverage and frequency, as well as those on the uncertainty in the middle atmosphere are normally met, at least for some users. Improvements are still needed to meet the URs on the uncertainty in the UTLS region for most products.

2.9 Land cover

There are two CCI Land Cover products:

- The global surface reflectance (SR) time series
- The CCI Global Land Cover map V1

The CCI Land Cover products validation plan was reviewed in CMUG Phase 1. The following sub-sections will thus not elaborate further on this validation plan. However, it is worth noting that even if the validation plan of the SR refers to the validation of both the SR composites and the global SR time series, only the latter has been achieved for the time being. More specifically only the computation of the temporal variance at the pixel level for the various spectral reflectance values has been achieved for the time being.
The obtained values were compared to other reflectance values originating from other sensors (e.g., SPOT-VGT).
For the CCI Global Land Cover Map V1, both qualitative and quantitative assessments were performed.

Quality

Global surface reflectance (SR) time series
Following an analysis, the authors of the report draw the following conclusions:

- the number of pixels which contribute to the analysis of the time series is very variable whereby this is caused by the data availability or by the cloud coverage;
- the impact of undetected clouds is visible in the time series and influences the statistical parameter estimate;
- the standard deviation values reach an order of magnitude from 1.6 through 65 % (mean 22%)

In addition to the quality assessment described above, a visual quality assessment of the LC CCI global SR-7day composites has been performed. 5 different issues were identified:

- Issue 1: striped pattern
- Issue 2: blurred inland band along coast lines and lake/river outlines
- Issue 3: pixelized coast lines and lake/ river outlines
- Issue 4: missing lakes and islands
- Issue 5: NoData (NaN value) in the desert over bright areas

Overall the quality of the global surface reflectance (SR) 7-day composite time series product was considered satisfactory for use in current applications. The issues identified do not constitute a critical road block on the path forward.

CCI Global Land Cover map V1

The Land Cover map results from a processing chain which uses the MERIS Full Resolution (FR) and Reduced Resolution (RR) multispectral SR 7-day composites as inputs. The map is a Level 4 product according to the CEOS definition. More specifically, the MERIS RR and SPOT-VGT data were used when needed to compensate for the lack of MERIS FR acquisitions.

CCI Global Land Cover Map V1						
Parameters	Sensors	Spatial coverage	Spatial grid	Temporal coverage	Temporal resolution	Total data volume
Land cover	MERIS SPOT-VGT	Global	300m*300m	2008-2012	1 product over 5 years	~350MB

The qualitative assessment enabled to identify a certain number of issues. However for all of them the potential solutions for their resolution have also been identified.

- Two types of spatial inconsistencies were identified: tiles limits and low resolution patterns in landscapes. The former mostly pertains to the classification process and algorithm while the latter mostly pertains to the use of lower resolution data (MERIS RR or SPOT-VGT).
- The classification of urban areas is sometimes incorrect.
- Occurrences of misclassification of the larger land cover classes exist. Similarly water is sometimes misclassified as another LC class.
- Also, certain small islands appeared to have been classified as water.

Uncertainty

No Uncertainty Characterisation Report is available for assessment.

Maturity of data

The CAR was assessed in CMUG Phase 1.

In Phase 1 the climate users evaluated the impact of the new Land Cover maps on the carbon, water and energy budgets using three different land surface models (LSMs). In addition, a new procedure for converting the land cover class fractions to the Plant Functional Types (PFTs) used in the LSMs was implemented following Poulter et al. (2015). However, an inter-comparison between these models was difficult due to the fact that different reference maps were used for each model. In Phase 2 this evaluation is being extended further by assessing the impact of the uncertainty in the land cover map, as well as in the LC to PFT conversion. This will also facilitate the model inter-comparison, as the same reference map from Phase 1 of the project will be used. To that end the LC_CCI team has provided the climate users with two alternative maps of the maximum and minimum "likely" biomass based on likelihood information from the classification procedure.

The individual climate user groups have performed more specific assessments based on research priorities at each laboratory. At LSCE the NDVI condition was used to evaluate the leaf phenology (seasonal cycle of vegetation) before and after the ORCHIDEE LSM parameters had been calibrated with another satellite NDVI dataset. This enabled an independent verification of the optimisation performance, as the posterior simulations match the LC CCI data more closely than the prior. This demonstrates an alternative, and very beneficial, use of the NDVI condition derived by the LC CCI project.

The Max Planck Institute for Meteorology (MPI-M) have used the LC_CCI Phase 1 maps to facilitate the development and implementation of new processes into the land surface scheme of the JSBACH LSM and its evaluation. The plant functional types (PFT) definition has been changed according to the new LC product and a new wetland extend dynamics (WEED) scheme has been implemented into JSBACH. The LC water body (WB) product is used as a prescribed boundary of wetland extent (JSBACH-ESA), but also for the evaluation of JSBACH-WEED.

A further improvement in Phase 2 will be the provision of yearly land cover maps, and an improved characterization of land cover changes. This will enable more robust climate change simulations that include land cover and land use changes over the past 15-20 years than are possible at present. However, this work will only start in the 2nd year of Phase 2.

2.10 Fire

Quality

Fire CCI summarizes the validation protocol to test the performance of the algorithms including the round-robin protocol for inter-comparison of improved and data merging algorithms in the **Product Validation Plan** (PVP). The validation approach follows the guidelines of the CEOS Land Product Validation subgroup (LPVS). The BA validation builds on cross-tabulation, regression analysis and the number of burned land patches detected, or variations on these. Most reference data for validation within the fire_cci project is generated by the consortium itself, but information is also gathered from other sources of data. The PVP defines therefore the requirement for reference data. The validation exercise is separated into two parts (i) use pre-selected study areas to test the performance of the BA algorithm (ii) validation of the global BA product. For the validation of the global BA product a probability sampling procedure is introduced based on spatial stratification to ensure sufficient sampling in each biomes, with a focus on regions with high BA.

The study sited are validated in the Product Validation Report I (PVR I). The product validation for the global product is still ongoing. The product validation addresses requirements defined in the user requirement document (URD):

- Global accuracy, understood as the agreement between the global product and reference datasets.
- Error balance of BA estimates, understood as the level of over or underestimation of BA.
- Temporal stability, understood as the homogeneity of accuracy over time.

Special emphasis is given to the temporal stability of the product, a parameter which hasn't been assessed for a BA product so far. Validated are the single fire CCI products for ten selected study sites for the different sensors applied (ATSR, AATSR, VGT, MERIS) and the final merged product. In addition previously released BA products were applied in the validation exercise to compare the performance of the fire cci products with those that are being already used by different communities. The validation report concludes that the BA products derived from MERIS show significantly better results in terms of overall accuracy than those derived from VGT and ATSR. The merging product performs significantly better than others in terms of omission errors and with similar commission errors than other European BA products. In terms of temporal stability all products showed stable validation parameters with highest values for AATSR. The fire CCI products showed medium to high values for VGT and MERIS in terms of overall accuracy, omission error and dice coefficient, with lower values for the commission error.

Two measures are used to validate the temporal stability. Both test were applied in the PVR. It is not clear from the PVP or the PVR which measures should be the preferred one to assess the temporal stability of the BA product, nor are the different results discussed in the PVR. The temporal stability is assessed for the composite of study sites. This could be further separated into temporal stabilities for different land cover classes, similar to the analysis of the other accuracy measures.

Overall, the PVP provides a good framework for the validation of the fire CCI product. The PVR applies this framework in its part I for the validation of the study sites. Measures and metrics are defined and explained in the PVP. The PVR could do, however, a much better job in linking to those, citing specific paragraphs and equations. What is missing is a description of the reference data used in the validation exercise. While the PVP states the requirements for reference data, none of the documents gives details on the actual reference dataset used.

The Product Validation Plan misses to link to the user requirement document (URD), in which specific requirements for a validation combining different user interests are combined. The Product Validation Report (PVR) does actually link to the URD. This should, however, also be part of the PVP. In addition, it is unclear how the PVP or the PVR links to the uncertainty assessment for the BA estimates (Fire_cci_Ph2_UAH_D1_2_2_CECR_v2).

The PVR Part II (Fire_cci Ph3_UAH_D4_1_1_PVR_II_v1_3) validates the global fire CCI products (MERIS_cci, VGT_cci, MERGED_cci) for the year 2008 based on 103 non-overlapping Thiessen Scene Areas (TSA) with reference fire perimeters determined from two multi-temporal Landsat TM/ETM+ images. The PVR assesses the accuracy and the error balance using six accuracy measures. In addition previously released BA products were applied in the validation exercise to compare the performance of the fire cci products with those that are being already used by different communities. In line with the local study site assessment MERIS_cci is the most accurate product of those produced by the fire_cci project (overall accuracy: 99.6 %, a Dice of Coefficient (DC) for the burned category of 28 % and errors of commission (Ce) and Omission (Oe) of 64 % and 77 % respectively). VGT_cci and the MERGED_cci product accuracy levels are lower. In comparison the MCD64 product results in: overall accuracy: 99.6 %, a Dice of Coefficient (DC) for the burned category of 42 % and errors of commission (Ce) and Omission (Oe) of 42 % and 68 % respectively. The PVR does not find any biome specific differences in terms of accuracy between the different assessed products. The PVR misses to intercompare the single global products beyond the single test sites. To be used in climate application users of the product will require a comparison of the product to the ones already used in various applications in climate research. At this stage no uncertainty characteristics of the fire cci products were used in the validation exercise and as the validation is limited to the year 2008 no stability measure was assessed.

The Product Validation Plan misses to link to the user requirement document (URD), in which specific requirements for a validation combining different user interests are combined. The Product Validation Report (PVR) does actually link to the URD. This should, however, also be part of the PVP. In addition, it is unclear how the PVP or the PVR links to the uncertainty assessment for the BA estimates (Fire_cci_Ph2_UAH_D1_2_2_CECR_v2).

Uncertainty

The CECR summarizes potential errors related to data acquisition, data processing, and landscape characteristics in the pre-processing phase as well as errors in the burned area detection algorithm and errors related to the merging algorithm. It describes how uncertainties will be documented for the pre-processing of the data and the burned area detection algorithm. This documentation is, however, missing for the merging algorithm. The CECR announces a guideline, that will be prepared for the users both for the pixel based and gridded product on how to use the reported quality levels of the product.

The latest version of the Comprehensive Error Characterisation Report (Fire_cci_Ph3_UAH_D_1_2_3_CECR_v2_1.pdf) describes the methodology used to quantify the uncertainty of BA estimates in the fire_cci pixel as well as grid products. In line with user needs of the climate research groups, the uncertainty of the gridded product is expressed as standard error of the total burned area, while in the pixel based product the uncertainty is given as the probability that a pixel is really burned. This is similar to other burned area products. The errors were measured using reference data for multi-temporal pairs of Landsat images derived within the fire cci project.

The pixel product uncertainty is based on the number of burned pixel in a 9x9 moving window. The r2 of the uncertainty model is thereby with 0.127 very low. For the gridded product the r2 for the regression model is 0.136. In line with the validation of the MERIS pixel product the uncertainty model reveals an underestimation of the MERIS fire_cci grid product that is proportional to the estimated burned area. The document states several reasons for the low performance in the uncertainty models, including that factors potentially important for the BA uncertainty were not included. The uncertainty models might improve if more data will be available in the future. Currently they are limited to three years for each study site. The analysis of uncertainties is planned to be implemented in the fire_cci processing chain. In terms of error characterization the commission errors were slightly related to the confidence level of the burned pixel, but no relationships were found for the land cover variables (biome type, land cover type, tree coverage).

The current version of the CECR does not provide a guideline on how to use the reported uncertainties. Given that the performance of the uncertainty models is rather weak this has to be clearly stated with the product. Upcoming uncertainty assessments will improve as more data will be processed. Here it should be also explored if more sample sites could be taken into account.

Maturity of data

The climate assessment report (CAR, Fire_cci_Ph3_LSCE_D4_2_CAR_v2_1) summarizes first applications of the fire_cci products and gives an idea about the maturity of the data for climate applications. Assessed are the pixel-based and gridded product for the year 2006-2008 for cci_meris and cci_merged. The cci_veg product was not applied. Compared to regional and national statistics the fire cci products underestimate burned area in line with other remote sensing based burned area products such as GFEDv4. This underestimation is related in the CAR to the omission of small fires. The pixel based product was successfully applied to derive fire patch size information. This will provide valuable information for fire model development. The gridded fire cci products were used to benchmark the fire model in ORCHIDEE. The information on landcover type burned provided by fire_cci was used to further disentangle the discrepancies. In addition, the fire_cci data were compared to other commonly used burned area products. This comparison was limited to the spatial distribution and did not address seasonal or inter-annual behavior. Furthermore, the fire_cci data was used as boundary condition in the ORCHIDEE model to derive fire CO and CO_2 emissions, which compared reasonable well with earlier estimates based on a similar approach reported in GFEDv3. Overall the CAR concludes that the cci_meris product is in close agreement with the MODIS based GFEDv4 product and outperforms previous burned area data based on ESA products (GLOBCARBON, L3JRC). To further assess the maturity of the data fire_cci will have to combine their uncertainty characteristic as well as their detailed site level validation with the applications performed in the climate research group to ensure that the full potential of the new product will be used.

2.11 Soil moisture

Quality

The overall validation approach for the CCI soil moisture product is outlined in the Product Validation Plan (PVP) which was assessed by CMUG in Phase 1. Further details on the validation results from the Round Robin Exercise are provided in the Product Validation and Algorithm Selection Report (PVASR) as well as in the Product Validation and Intercomparison Report (PVIR, v1.0).

The reports provide a comprehensive overview about the product validation approaches. The PVP proposes a hierarchy of different product validation approaches with the following major components:

- Validation at point scale using in situ station data
- Validation at the regional to global using cross comparisons of the ECV SM product with other soil moisture products (e.g. SAR), precipitation records, terrestrial water storage records, data assimilation experiments and the so called R-metric. The latter quantifies the skill of a soil moisture product to compensate for known precipitation errors and provides a measure on the representation of short term precipitation anomalies by means of soil moisture anomalies.
- Analysis of long term trends of various variables (NDVI, precipitation, tree rings, runoff) and their consistency with CCI SM trends.

Different metrics for the comparison of the CCI SM product with other reference data are introduced in the PVP (e.g. correlation, bias, triple collocation error) and successfully applied in the evaluation of the round robin exercise. The PVP also provides a comprehensive overview about the required datasets for the validation.

As the provided documents are overall of excellent information content, CMUG has only a few comments on aspects that were not fully covered in the available documents so far:

Validation of long term product stability: The CCI SM team envisages to compare their product against a variety of different other long term EO datasets and compare the consistency of trends among these different datasets. Results published by the ECV SM team (e.g. Dorigo et al, 2012 doi: 10.1029/2012GL052988) have shown large potential of this approach. However, these kind of cross-comparison studies do not provide a means for the validation of the actual temporal stability of the data record. It is recommended that the ECV SM adopt similar measures of temporal stability like defined in the CCI SST Uncertainty characterization document (SST_CCI-UCR-UOE-001) and to develop more stringent approaches to quantify the temporal stability of the ECV SM record if possible.
Unclear temporal scale for validation: The PVP remains unclear in many cases on which temporal scales the comparison with reference data will be made. For cross-comparison with precipitation

datasets, the ECV SM team plans for instance cross comparison with the Global Precipitation Climatology Center (GPCC) data products. These products are currently only available on monthly timescales and it remains unclear if this would be suitable for the evaluation of the CCI soil moisture product. On climatological timescales such a comparison could be useful and it has been shown by Loew et al, 2013 that the first version of the CCI SM data product shows good skill in capturing monthly precipitation anomalies at the global scale.
Validation of uncertainties: the available documents lack a description of how the ECV SM team plans to validate the accuracy of their product uncertainty estimates (if planned and possible).

Uncertainty & Quality

The ESA CCI soil moisture project has produced a comprehensive update of the Comprehensive Error Characterization Report (v. 1.0). This provides an excellent reference document describing uncertainties in the generation of an ECV soil moisture record from either passive or active microwave observations. It also deals in details with additional uncertainty sources related to the validation of the coarse scale soil moisture products and the merging of soil moisture information from different data sources. The document provides a thorough analysis of the uncertainty terms in the ECV SM production chain. Different metrics for validation of the data products like triple collocation, R.metric and data assimilation are discussed.

Major scientific advances between v0.7 and v1.0 of the CECR have been made in particular through novel developments and insight for the theoretical framework of Triple collocation (TC) which is used as a basis for the quantification of the uncertainties of the CCI soil moisture product. An excellent summary about the recent advances in TC analysis can be found in Gruber et al. (2016). A summary and its implication for the CCI soil moisture dataset validation is provided in chapter 5.2.4.3 of the CECR. The relationship between TC measures and standard skill scores like e.g. RMSE or correlation measures is shown. Based on these novel insights, the ESA CCI SM team has also proposed to use the Signal to Noise Ratio (SNR) as an additional metric for the quantification of the uncertainties of the data product (see Gruber et al., 2016 for technical details).

An assessment of an initial version of the ECV SM dataset has been made by Brocca et al. (2011) showing good correlation of the soil moisture data with in situ observations. Dorigo et al. (2014) provide a very comprehensive analysis of the uncertainties of the first version of the ESA CCI SM dataset compared against a global network of in situ observations from the International Soil Moisture Network. In general good agreement between ECV SM and in situ observations is found when using correlation as a measure for the product skill.

The ESA CCI SM team and collaborators have been also very active in developing independent means for the validation of soil moisture products. A novel and indirect method for assessing the quality of remote sensing datasets has been developed by Brocca et al. (2013). They use the soil moisture dynamics to invert precipitation information using a simple surface water balance model (SM2RAIN). The such obtained precipitation datasets are then compared against reference precipitation measurements. Several datasets have been produced based on different soil moisture records (Brocca et al., 2014). A recent study analysed the quality of the novel ESA CCI SM dataset at the regional scale using the SM2RAIN method (Abera et al., 2016).

Maturity of data

No CAR is available for assessment at the time of writing.

It is expected that the next version(s) of the ECV SM product will further increase the accuracy of the data. However it will be in general difficult to quantify if the final product will meet the GCOS accuracy and stability requirements. The reason is that GCOS defines an absolute accuracy level of 0.04 [m^3/m^3] with the rationale of an accuracy of 10% of the typical soil moisture dynamics. However, remote sensing as well as in situ measurements are often biased compared to the soil moisture of a larger area (representativeness error). Thus, while the absolute values might differ, the relative dynamics is often quite well captured by in situ and remote sensing techniques. Thus using correlation and anomaly correlations as a measure for skill is a good approach, but not implemented in the GCOS requirements. CMUG therefore recommends to revise the definition of the GCOS soil moisture requirements by including also relative accuracy measures, like it is done also for other ECV's (e.g. albedo). The ECV soil moisture team should be encouraged to take an active role in this revision.

3. Summary of CMUG assessment of Quality, Uncertainty and Maturity in ECVs

It is challenging to summarise the assessments made in this report on the Quality, Uncertainty and Maturity in ECVs as taken from the CCI ECV project documents and other sources (GCOS 2016, 2011) for many reasons. Firstly each CCI project is addressing its own set of issues unique to its ECV in remote Earth system observations and data processing, secondly not all project reports are available for assessment, and lastly the assessment is made against a wide base of potential users and applications. However, adopting the 'maturity index' approach shown in Table 2 allows an overview and summary. Table 2 shows clearly that although there are many CCI projects which have successfully tackled the issues around producing high quality mature data with good uncertainty characterisation there are still some that have yet to do so.

ECV	Quality	Uncertainty	Maturity	GCOS requirements
SST				
Ocean Colour				
SSH				
Sea Ice				
Clouds				
Aerosols				
GHG				
Ozone				
Land Cover				
Fire				
Soil Moisture				
Key				
AMBER =	PVP / PVIR partly describes quality process or quality goals	UCR / CECR partly describes process for uncertainty characterisation or goals	CAR partly describes process for achieving climate-based goals	Partly meets GCOS requirements
GREEN =	Meets quality requirements	Meets uncertainty characterization requirements	Meets climate data user requirements	Meets GCOS requirements
GREY =	No information	No information	No information	No information
WHITE =	No assessment made in this table	No assessment made in this table	No assessment made in this table	No assessment made in this table

Table 2. Summary of the CMUG assessments on how well reports published by the ECV dataset teams address Quality, Uncertainty, Maturity and GCOS requirements.

4. References

Abera et al., 2016, Comparative evaluation of different satellite rainfall estimation products and bias correction in the Upper Blue Nile (UBN) basin. DOI: 10.1016/j.atmosres.2016.04.017

Brocca, L., L. Ciabatta, C. Massari, T. Moramarco, S. Hahn, S. Hasenauer, R. Kidd, W. Dorigo, W. Wagner, and V. Levizzani (2014), Soil as a natural rain gauge: Estimating global rainfall from satellite soil moisture data, J. Geophys. Res. Atmos., 119, 5128–5141, doi:10.1002/2014JD021489.

Brocca, L., T. Moramarco, F. Melone, and W. Wagner (2013), A new method for rainfall estimation through soil moisture observations, Geophys. Res. Lett., 40, 853–858, doi:10.1002/grl.50173.

Brocca, L., S. Hasenauer, et al. (2011). "Soil moisture estimation through ASCAT and AMSR-E sensors: An intercomparison and validation study across Europe." Remote Sensing of Environment

CMUG, 2013: Demonstration of CMF functionality for the assessment of ozone, aerosol and soil moisture. Technical Note/Deliverable 3.1v_1d, September 2013.

Dorigo, W.A., Gruber, A., De Jeu, R.A.M., Wagner, W., Stacke, T. Loew, A., Albergel, C., Brocca, L., Chung, D., Parinussa, R., Kidd, R. (2014): Evaluation of the ESA CCI soil moisture product using ground-based observations. *Remote Sensing of Environment*. doi: 10.1016/j.rse.2014.07.023

GCOS, 2016: The Global Observing System for Climate: Implementation needs. Draft.

GCOS, 2011: Systematic observation requirements for satellite-based data products for climate (2011 Update). Supplemental details to the satellite-based component of the "Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC (2010 Update)" GCOS – 154.

Gruber, A., Su, C.-H., Zwieback, S., Crow, W., Dorigo, W., Wagner, W., 2016a: Recentadvances in (soil moisture) triple collocation analysis. Int. J. Appl. Earth Obs.Geoinf. 45, 200–211

Rodgers, C. D., 2000: Inverse methods for atmospheric sounding – Theory and practice, World Scientific Series on Atmospheric, Ocean and Planetary Physics Vol. 2.

de Leeuw, G., T. Holzer-Popp, S. Bevan, W. Davies, J. Descloitres, R.G. Grainger, J. Griesfeller, A. Heckel, S. Kinne, L. Klüser, P. Kolmonen, P. Litvinov, D. Martynenko, P.J.R. North, B. Ovigneur, N. Pascal, C. Poulsen, D. Ramon, M. Schulz, R.Siddans, L. Sogacheva, D. Tanré, G.E. Thomas, T.H. Virtanen, W. von Hoyningen Huene, M.Vountas, S. Pinnock (2013). Evaluation of seven European aerosol optical depth retrieval algorithms for climate analysis, Remote Sensing of Environment (2013), http://dx.doi.org/10.1016/j.rse.2013.04.023

Holzer-Popp, T., de Leeuw, G., Martynenko, D., Klüser, L., Bevan, S., Davies, W., Ducos, F., Deuzé, J. L., Graigner, R. G., Heckel, A., von Hoyningen-Hüne, W., Kolmonen, P., Litvinov, P., North, P., Poulsen, C. A., Ramon, D., Siddans, R., Sogacheva, L., Tanre, D., Thomas, G. E., Vountas, M., Descloitres, J., Griesfeller, J., Kinne, S., Schulz, M., and Pinnock, S., Aerosol retrieval experiments in the ESA Aerosol_cci project, Atmos. Meas. Tech., 6, 1919 - 1957, doi:10.5194/amt-6-1919-2013, 2013.

Kinne, S. & the entire aerosol CCI-team, ESA's new 17 year ATSR climate data record, ESA internal document, 2014.