CMUG Deliverable

Number: D2.3
Due date: January 2014
Submission date: March 2014
Version: 0.4

Climate Modelling User Group

Deliverable 2.3

Technical note on Validation and User Assessment

Centres providing input: MOHC, MPI-M, ECMWF, MétéoFrance

Version nr.	Date	Status
0.1	Dec 13	Initial draft R. Saunders
0.2	15 Jan 14	Update R. Saunders
0.3	26 Feb 14	Added other contributions except SSH
0.4	27 Feb 14	Added SSH

CMUG Deliverable

Number:	D2.3
Due date:	January 2014
Submission date:	March 2014
Version:	0.4

Deliverable 2.3

Technical note

Validation and User Assessment Report

Table of Contents

CMUG Deliverable
Number: D2.3
Due date: January 2014
Submission date: March 2014
Version: 0.4

Validation and User Assessment Report

1. Purpose and scope of the Technical note

The purpose of this document is to review the product validation plans of each CCI team and give feedback to ESA and the CCI teams. It shall provide comments, technical advice on some of the CCI Projects Task 4 deliverables, including the "Product Validation Protocol", "Product Validation and inter-comparison Report", and "Climate Assessment Report". This document is intended as user feedback to CCI Projects (but is not part of a formal review).

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 programme for the climate research community?
- Is the user assessment complete and representative of the wider community?

At the time of writing this report there were only a few climate assessment reports available and so the latter two questions could not be addressed from most of the CCI teams perspective. The CMUG make their own assessment of this in their Scientific Exploitation Report (D4.2).

2. Comments on CCI validation and user assessment reports

The validation plans were not all in a coherent report and some aspects were covered in other reports. The comments below refer to the relevant parts of the reports which were available to CMUG at the end of phase 1 (Jan 2014). The documents reviewed are listed in Table 1.

A general comment applicable to all ECVs is although there are good ATBDs for all ECVs *before* the round robin comparisons there isn't an ATBD of the finally selected product. It may of course be the same as the original ATBD, but not necessarily, and it would help the users if there is a clear ATBD to reference with only the chosen product. It is recommended by CMUG to address this before the end of phase 2.

ECV	Version of PVP	Climate Assessment Report	Other Docs
SST	SST_CCI-PVP-UOL-001 Dated 30 Jan 2012	SST_CCI-CAR-UKMO-001 24 Jan 2014	SST_CCI-UCR-UOE-001 15 July 13 SST_CCI-CRDP-UKMO-001 30 Sept 13
Ocean Colour	PVASR Part 1 Atmos Corrn D2.5 Dated: 26/11/12 v1.3; Part 2 In-Water Algorithms D2.5 Dated: 13/11/12 v1.4 Part 3 Pixel Identification D2.5 Dated: 26/11/12 v1	Not available	Uncertainty Characterisation Document (UCD) Ref: D2.9 Dated: 30/11/11 v1.1
Sea Level	SL_CCI-PVP-005-1-1 v1.1 11 Oct 2011	SLCCI_CAR-035-1-1 15 Dec 2013	SLCCI_ErrorReport-030 v1.1 9 April 13 SLCCI-PVIR-31 v1.0 18 Sept 2013
Clouds	PVP v1.1 (28.03.11)	Not available	Comprehensive Error Characterisation Document v1.1 11 Sep 12
Ozone	Ozone_cci_PVP_1.0 Dated 03/01/2012	CAR_Ozone_CCI_030 22014_final (3 Feb 2014)	GODFIT_ATBD_v0 (25/05/2011) Ozone_cci_KIT_ATBD_0_01_0 (18/04/2013) ACC-9_Validation_Protocol (8 Mar 2013)
Greenhouse Gases	PVPv1_GHG-CCI_final Dated 20 May 2011	GHGCCI_CAR_V1.1 (18 Nov 2013)	PVIR-GHG-CCI-v2 (4 Nov 2013) GHG-CCI_ATBD_v1.0_final (15 Mar 2012)
Aerosol	PVP v1.5 13/4/2010	Not available	Aerosol UCR v3.0 26 Nov 2013
Land cover	Land_cover_CCI_PVP_1.3, Dated: 04/07/2011	Not available	
Fire	Fire_cci_Ph1_UAH_D2_1_PVP_v4_0 Dated: 03/02/2014	Not available	Fire_cci_Ph3_UAH_D4_1_1_PVR_I_v2_0 Dated: 22/01/2014
Soil moisture	Product validation plan (PVP, v1.1, 30.08.2012)	Not available	Product validation and algorithm selection report (PVASR v1.0, 27. June 2013)

Table 1. Version of documents reviewed.

2.1 Sea surface temperature

Most of the relevant material is in chapters 7 to 9 in the product validation plan. There is also relevant material in the 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 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 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.

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 the 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. The CMUG have provided an assessment of the uncertainties (see D3.1v2).

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 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.

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.

2.2 Ocean colour

The two 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.

/ 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 be 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.

- Equation (3) in the PVARS-Part1 should be corrected – the second line represents "standard error of relative RMSE (not absolute RMSE)".

- In the atmospheric correction round robin, SeaDAS is showing advantage over POLYMER for SeaWiFS and MODIS processing. Could this be attributed to the use of different auxiliary meteorology and use of different in-situ dataset in the matchup study?

- PVARS-Part 2, Page-14 "*comply waters with lw(665)<0.02 (higher scattering waters)*" should read lw(665)>0.02.

- PVARS-Part 2, Page-26 "*Chi-square test*" needs clarification as to why this is defined as a chi-square test in order to avoid the ambiguity. Is the idea to test the Null Hypothesis here?

- 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.

- The units of the chlor_a_rms_uncertainty and chlor_a_bias_uncertainty variables in the OC-CCI NetCDF files are not clear so the documentation should be improved.

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.

No climate assessment report was available at the time of writing.

2.3 Sea surface height

Two documents are addressing the quality and uncertainties of the SL CCI products. In the PVIR the authors certify the end-to-end quality of ECVs and analyse the total contribution of improvements done in the final products 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 the 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.

But the error report document allows to have 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 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 allow to go further in the understanding of the uncertainty evaluation process. The hypothesis that are done to combine the different sources of uncertainties are also well stated. The authors highlight the fact that the resulting errors, at the global and at the regional scales, are often higher than the user requirements identified at the beginning of the project. However, it appears that this was not a limitation in the analysis done by CMUG over the Mediterranean basin at least on average over the whole basin (the spatial variability within the basin is under study).

The final version of the SL PVP was already examined in the CMUG D2.2 deliverable (“Technical note on CCI System Requirements”). 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 in the CAR that 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, …).

The user assessment presented in 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. It also covers the emerging use of the data to evaluate the ability of climate system models to simulate the observed sea level trends over the last decades. This last application is of particular interest to make this data more widely used by the climate modelling community.

The comparison between the SL-CCI dataset and the SL AVISO dataset is done all along the document. 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 with closer agreement between the two products over this specific region.

2.4 Clouds

The Product Validation Plan gives a summary of the CCI's activities involving validation of the cloud products against satellite, ground-based and aircraft data. A detailed summary of these validation data sets is provided (Sections 2-4), together with a general outline of the comparisons (Section 5). The Methodology section itself (Section 6) is rather short, however the CCI will in general follow already-established methods – developed as part of CREW (Level 2) and the GEWEX cloud assessment (Level 3) – so this may not be too important as these are documented elsewhere.

The Comprehensive Error Characterisation Document provides a description of how the errors will/could be determined (e.g. it outlines the relevant equations from the ORAC algorithm) but does not currently provide any quantitative information on the errors of the CCI cloud products themselves.

No climate assessment report was available at the time of writing.

2.5 Aerosols

The Product Validation Plan (PVP) briefly addresses the validation required for the assessment of the products using different algorithms rather than validation of the final product. The documents referred to in the PVP which described the specific validation sources could not be accessed from the aerosol CCI web site but a very brief description is given in an appendix. The AEROCOM network is used to provide the standard reference. It is surprising that advanced infrared sounder data (IASI, AIRS) is not used as these hyperspectral data can give some information on the aerosol height profile, especially for dust and volcanic aerosols, and would improve the validation.

The Product Validation Report (PVR) summarizes the validation/comparison of the CCI aerosol products against independent measurements. The validation is focused mainly on AOD at 550 nm and to a lesser degree on the Angstrom coefficient. There is absence of validation at other channels, which would provide information about the goodness of the retrieval and the suitability of the chosen aerosol model (although some information about these can be provided from the Angstrom coefficient). The wavelengths from which the Angstrom coefficient is calculated are not mentioned in the document. It seems to be the pair 555-870 nm in general, but in the case of the ADV algorithm over land the pair 555-1610 nm seems to be used. It is advised that this point should be clarified. In the Section 3.1 there is inconsistency about the values reported in the text and the respective ones found in the Tables. For the same validation (LOA/ICARE), the collocation criteria should to be verified as they are not the same to ones found in the appendix PVR_L2_ICARE.pdf. As the collocation criteria applied for the validation of L2 between LOA/ICARE and NILU are not very different a paragraph could be added in order to compare/discuss the results from the two validations. In the Section 3.3 (validation realised by NILU), it will be better to provide the comparison for the same number of points not only between ATSRs algorithms but also between ATSRs and MODIS or MISR. This will further show the good quality of the CCI aerosol products against these widely used satellite retrievals. There is lack of validation for L3 Angstrom coefficient (MPI provides only its scoring about it). Regarding the validation of the stratospheric extinction profiles, the average results (Figure 3.7-1) should be given for the extinction profile and not for AOD (which is the integral of the extinction coefficient), as it has been realised for the individual months in the appendix. This will provide information about discrepancies in the vertical. The absorbing aerosol index (AAI) from OMI (Section 3.8) can be also compared to the respective ones from GOME2 and SCIAMACHY (at least for L3 on a monthly scale). This way of comparison may be more relevant than using a model. Concerning the synthetic study (Section 3.9), the main results should be included in the PVR, as they provide an indication about the uncertainty in the retrieval process for the algorithms (even if this is restricted to a limited number of cases). Finally, it is not clear why validation against the AERONET-MAN measurements has not been realised, as there are at least 20 cruises for 2008. Thus, in an annual scale some match-ups should exist above oceans at least for PARASOL and MERIS.

The Uncertainty Characterisation Report (UCR) gives a comprehensive list of the potential sources of error in the aerosol products and also on the different propagation of the errors for each of the three products. The uncertainty value provided with the data is for 550 nm but the aerosol optical depths are also provided at three other wavelengths and so there is a question how appropriate the aerosol optical depth at 550 nm is for these other wavelengths. It is very likely the error characteristics will be different for the longer wavelengths. CMUG therefore recommend that uncertainty estimates be provided for all wavelengths not just 550 nm and the Angstrom coefficient. It should be noted that the uncertainty calculation for every retrieval algorithm is treated in a different way (although harmonization will be the focus of phase 2 of Aerosol CCI).

At the time of writing there is no Climate Assessment Report (CAR) provided by the CCI aerosol team. Thus, there is no comparison with some model, as it will be in the CAR.

2.6 Greenhouse gases

The GHG Product Validation Plan (PVP) document recognizes that, when they exist, standard, traceable, and approved (by the CEOS-WGCV) validation protocols should be adopted. It is also noted that common centralized infrastructure and protocols, e.g. for data validation, do not exist across CCI, and that the possibility of establishing one could be examined. This is a good point, and it would be useful to define a common standard protocol as general guidelines.

Two products, XCH4 and XCO2, are retrieved from two satellite instruments (SCIAMACHY and GOSAT). The data validation described in the PVP follows the standard method of confronting the retrieved products with independent observations with traceable quality, namely ground-based data from the TCCON and NDACC networks, as well as any other reliable in-situ station. Thorough descriptions of these independent data, of the differences between the satellite and the in-situ instruments, and more generally of how the comparisons should be performed are all provided in the PVP. This part is a really nice summary for users to understand how the data should be treated and used, that could perhaps be encouraged across all teams as appropriate, if not yet adopted.

The temporal matching criterion for these comparisons is two hours, while three spatial criteria are selected using observations falling within radii of 100, 350, and 500 km from the station location. This seems a very good idea as it allows one to understand how to compromise between having tight matching requirements and statistically significant number of data for the validation.

One initial concern was that the validating data described in the PVP only consist of ground-based instruments, and thus the obvious question on the validity of the assessment conclusions to periods and locations where no in-situ observations are available. Other (more recent) documents discuss the methods and the results obtained by the team. The impression is that the PVP document neither reflects the extent and amount of work performed by the team, nor provide information on the protocol to assess the observation errors. Regarding the latter, information is also available in the GHG ATBD, which provides an in-depth description of all used algorithms. A more concise, high level document focusing on the uncertainty characterization would have been very useful. The understanding is that the retrieval algorithms and the corresponding error characterization follow Rodgers (2000) and that the observation errors account for the observation noise, and errors in the forward model parameters.

The Product Validation and Inter-comparison Report (PVIR) and the GHG CRG Climate Assessment Report (CAR) - both published in November 2013 - present much more in depth assessment of the various GHG products. The PVIR includes comparisons of the core GHG products with similar collocated products retrieved from other satellite sensors that are used as Additional Constraints Algorithm (ACA) Products (e.g. ACE-FTS, MIPAS, IASI, and so on). These comparisons are used to provide both an assessment of the observation quality and a validation of the error characterization.

Various aspects were discussed by the GHG CRG in the GHG CAR for XCO2 and XCH4. For instance, the impact of XCO2 is assessed for the retrievals obtained from two of the available algorithms using two independent approaches: a data assimilation approach and a direct flux inversion method also based on a variational method. The impact is assessed in terms of the reduction of the analysis error with respect to the background error. The two approaches lead to different impact levels. These differences could partly be due to the different hypotheses applied to the two methods. Additionally, they are applied to datasets retrieved using different algorithms. It would be interesting, in order to consolidate the outcome, to verify if and how the results changed if these two methods were applied to the same dataset.

The data assimilation approach described in the CAR was also used to characterize the presence of systematic errors in the BESD retrievals. This was done by applying the procedure three times using three different assumptions on the observation errors (i.e. fully correlated, uncorrelated, and partly correlated with the TM3 model). The results showed that the largest analysis error reduction was obtained when the observation and the TM3 model errors are assumed uncorrelated, suggesting that the data are essentially unbiased with respect to the TM3 model.

2.7 Ozone

The Ozone Product Validation Plan document describes in details the method used to validate the ozone products from the three operation lines (total column ozone, nadir ozone profiles and limb ozone profiles). The general strategy was to define a common validating procedure to be applied to all products to guarantee a standardized approach, though it is rightly recognized that differences may exist depending on the specific characteristics of each data product. In the PVP document, it is clearly indicated that the various products have to meet the requirements established within the QA4EO framework, particularly in terms of well documented and traceable quality indicators. This is a very important aspect as it represents a way to make these products user-friendly. It is noted that such a quality indicator or quality flag would be useful for all data levels, including level 3 monthly mean data.

The quality assessment is performed in two ways: by verifying that all the production steps are sound and correct, and then by comparing the retrieved products with reference data with accredited quality. This is a standard practice for data quality assessment. Clearly, limitations on the assessment result validity may exists depending on the used reference data, e.g. spatial and temporal coverage. In this sense, attention has been paid to using different types of reference data sets that include ground-based and satellite observations (as available and appropriate to the retrieved product), as well as model outputs in regions where no other independent observation was available. In PVP, it is recognized that the conclusions on the quality assessment may be affected by comparison uncertainties, i.e. limitations with the method. Matching criteria with independent should be carefully defined depending on both the retrieved and the reference dataset coverage and availability. This is recognized in the PVP but no indication of the actual matching criteria is given. The ACC-9_Validation_Protocol provides this information for the total column. In this case, daily averaged data within a 150km radius from the ground station location are used. Here, the use of daily averaged data seems a too relaxed criterion, though the number of coincidence could not be statistically significant if a more strict temporal criterion was used. No reference is given in the ACC-9_Validation_Protocol for the criteria applied to the nadir and limb profiles, nor was made in the PVP document, and in other documents, for instance the PVASR, DARD, and the web-site, that were browsed. It would have been useful if this information (even as a preliminary choice) was given in these documents.

High level information on the validation procedure is included in the observation metadata. This is a very useful practice for users to understand the level of quality assessment performed that could be extended to all ECVs when not already adopted.

The PVP reports that the uncertainty estimates can include estimates of the bias and precision of the data with respect to reference data, and identification of the temporal and spatial domains over which those estimates are valid. Different components (measurement noise, a priori uncertainty, smoothing error) characterize the retrieval errors provided at pixel level. A thorough description of the error budget is provided in the two ATBDs (one for the GODFIT total column ozone, and one for both the nadir and limb profiles). As for the error assessment, neither the PVP nor the ATBDs provide a detailed discussion of the protocol to be followed, but in the PVP a remark is made according to which this "*has to be assessed and expressed as the percent relative ozone difference with respect to correlative measurements of reference*".

Results from the CRG assessment were published in the O3 CCI Climate Assessment Report (CAR) on 3 Feb 2014. The O3 CCI CRG includes user groups working with Climate-Chemistry and Chemistry Transport Models, but no user within the reanalysis community, which could have been not represented. This limitation was overcome in the final CAR by including a summary of the preliminary CMUG comparisons of the L3 Ozone products with several ozone reanalyses. A more detailed discussion on the latter was published in CMUG (2013).

Focusing on the CRG assessment, monthly mean zonal mean modelled ozone datasets from two Climate-Chemistry Models and a Chemistry Transport Model show the same spread of the data. Although this can give some information about the quality of the data and the models, a proper analysis of the quality of each dataset, based for instance on a deep knowledge on the models' ability and limitations, is missing. Such an analysis would have been useful for instance to appreciate the differences between different instruments/observation types/viewing geometries and understand more about potential inter-instrumental biases.

A trend analysis is performed for two observed and two modelled ozone datasets. It shows clearly shortcomings in the models, but also, despite an overall good agreement, some regional differences in the two retrieved datasets.

There is an attempt to assess the observation standard deviations of the merged total column ozone with two examples shown for March and November 2008.

2.8 Land cover

The CCI Landcover product validation plan reviewed here is already from the year 2011 and summarizes therefore the plans of the project to validate the CCI landcover product. No written information about the actual validation performed for the product validation was available to CMUG at the time of writing. The comments below can therefore only refer to the planned validation.

The product validation approach of CCI landcover is based on pillars, namely:

1) **A confidence building procedure**, which is a continuous, qualitative assessment of the accuracy of the land cover product. The product accuracy is assessed by a quick screening procedure that allows to identify mainly artefacts and misclassifications from the data processing and class labeling procedures.
2) **An independent statistical validation** is based on a network of local landcover/land uses experts from different regions all over the world. These experts are supposed to review the land cover product for their particular (spatial) region of expertise and provide critical feedback on the CCI landcover by comparing the landcover product against high resolution remote sensing datasets as well as the field experience/knowledge of the experts.
3) **Comparison with other datasets:** Further, validation on larger spatial scales will be performed on the basis of high resolution satellite imagery from various sources as well as other existing land cover products.
4) **Temporal consistency analysis:** It is proposed to perform a temporal consistency analysis of the different CCI LC epochs by (i) comparing the individual epochs with each other and (ii) reflect observed differences in the light of the validation of the individual CCI LC epochs.

A comprehensive probabilistic sampling design is proposed to perform a representative sampling of the locations to validate the CCI LC product. The required tools have been developed at the prime contractor of the ESA CCI LC project (UCL). The document also provides an overview about the employed accuracy metrics.

Overall, the suggested approach provides a very convincing framework for the evaluation of the CCI LC product. It is nevertheless not clear how traceable and reproducible the final product evaluation will be. The incorporation of local experts is indispensable, but adds a major subjective component to the entire evaluation process.

- It is therefore questionable, if the product skill scores finally estimated might be reproducible my a different panel of experts.

The PVP remains unclear about the expert knowledge on historical land cover changes. Currently, the ESA CCI LC project is dealing with data from roughly a single decade.

- How solid is the local expert knowledge on these timescales?

The detailed approach how the temporal consistency analysis between the different epochs will be performed remains very unclear from the document. The different epochs of the CCI landcover product are based on 5 years of data each. Differences between these different products might be related either to (i) a change of the landcover over time or (ii) an artefact of the classification procedure. While a user might be interested in the former change as an information, an artefact should be avoided from a user's perspective.

It is a very difficult task to assess the temporal consistency between the different data products and the CCI LC team has always been clear about the fact, that a difference between two land cover maps is expected to be different from a land cover change product.

- However, if the ESA CCI LC products shall be used as boundary conditions in climate models over multiple decades, a realistic representation of land cover dynamics is considered as important.

A proper assessment and flagging of areas with dubious changes over the different epochs is therefore considered as essential. The PVP remain unclear how the temporal consistency analysis will be made exactly and how dubious areas are identified and marked.

2.9 Fire

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).

2.10 Soil moisture

The overall validation approach for the CCI soil moisture product is outlined in the Product Validation Plan (PVP). Further details on the validation results from the Round Robin Exercise are provided in the Product Validation and Algorithm Selection Report (PVASR).

Both 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)

3. References

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

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