CMUG Deliverable

Number:	**D2.2 Technical note on CCI system requirements**
Due date:	**March 2013**
Submission date:	**August 2013**
Version:	**0.7**

Climate Modelling User Group

Deliverable 2.2

Technical note on CCI System Requirements

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

Version nr.	Date	Status
0.1	8 Nov 2012	Initial Template RS
0.2	15 Feb 2013	Added SST
0.3	25 Feb 2013	Added Landcover and soil moisture AL/Ocean Colour/Aerosol RS and added GHG RD, Added Fire SK
0.4	25 Mar 2013	Added final ECVs except Cloud and section 3
0.5	8 April 2013	ESA Cloud now included
0.6	17 April 2013	Added more on European initiatives in sec. 3
0.7	5 Aug 2013	Response to comments from ESA

CMUG Deliverable
Number: D2.2 Technical note on CCI system requirements
Due date: March 2013
Submission date: August 2013
Version: 0.7

Deliverable 2.2

Technical note on CCI System Requirements

Table of Contents

1. PURPOSE AND SCOPE OF THE TECHNICAL NOTE 3

2. HOW CCI DATASETS ARE MADE AVAILABLE TO THE CLIMATE MODELLING COMMUNITY 3

3. REFERENCES 8

ANNEX 1. REVIEW OF CCI SYSTEM REQUIREMENT DOCUMENTS 9

2.1 SST 10

2.2 Ocean Colour 10

2.3 Sea-level 11

2.4 Landcover 12

2.5 Soil moisture 13

2.6 Greenhouse gases 13

2.7 Aerosols 14

2.8 Fire 15

2.9 Ozone 16

2.10 Clouds 17

CMUG Deliverable
Number: D2.2 Technical note on CCI system requirements
Due date: March 2013
Submission date: August 2013
Version: 0.7

CCI System Requirements

1. Purpose and scope of the Technical note

This report is to document how the CCI data should be made readily available to the climate modelling community. It also provides comments from the CCI Climate Modelling Users Group (CMUG) on the CCI team system specification document (SSD), product validation plan (PVP) and algorithm theoretical basis document (ATBD) in the annex. The main aim of this document is to provide comments on any issues relevant to the end users, particularly climate modellers and users for reanalyses. This is not a detailed review of all the system specific documentation.

2. How CCI datasets are made available to the climate modelling community

The architecture for delivering climate datasets to the user community is still very much in its infancy. There are several initiatives which are starting to influence this. The overall infrastructure is outline in Global Observing Systems Information Center (GOSIC) http://gosic.org/ which provides one-stop access to data and information identified by the Global Climate Observing System (GCOS), the Global Ocean Observing System (GOOS) and the Global Terrestrial Observing System (GTOS) and their partner programs, such as the Global Atmosphere Watch (GAW) and regional observing systems, such as the GOOS Regional Alliances (GRA). Datasets can be searched by Essential Climate Variable and other methods. Satellite datasets are not well covered in this framework however.

The reanalysis communities have by necessity developed their own archives of satellite datasets mostly at level 1 and 2 for assimilation into the models. In Europe ECMWF has a large archive of historical satellite data mainly at level 1. At present this dataset is very hard to access for external users but there is a project as part of ERACLIM underway to make access to these data easier for climate model and other climate research applications. CORE-CLIMAX will also facilitate the integration of ECVs into the reanalysis chain by proposing a feedback mechanism ensuring that the results of the re-analysis process get appropriately reflected into updates of the ECVs.

The IS-ENES project is attempting to co-ordinate the infrastructure for climate modelling in Europe to facilitate model comparisons, exchange of results etc. One element of their work is to develop a common system to confront models with observations and the CCI datasets should be able to feed into this system.

In general climate modellers like to have easy access to observational datasets in a standard format which they can easily read and compare with their model fields. They also have to be clearly documented. The Obs4MIPS programme, which was started by NASA/JPL, is an initiative to make satellite data more easily available in a standard format (NetCDF-CF) and averaged on to a 1 deg grid with monthly means. This does not meet all modeller's requirements but it is a good basis from which they can investigate the basic characteristics of the observations. Currently the Obs4MIPS project is only 'repackaging' existing satellite datasets which are not always CDRs but it does provide the

framework for satellite CDRs to be made available to the modelling community. It uses the Earth System Grid Federation (ESGF) infrastructure (www.esgf.org) as modellers already use this for transferring the CMIP5 model fields to their institutes. The ESGF grew out of the larger Global Organization for Earth System Science Portals community. Collaborating partners in the federation include Lawrence Livermore Labs and a host of other organizations such as the German Climate Computing Centre (DKRZ), British Atmospheric Data Centre (BADC) and the University of Tokyo Centre for Climate System Research. The ESG has recently been upgraded to allow the transfer of large datasets. Figure 1 shows the overall concept of the processes in the ESG to access datasets for modelling applications. Regional nodes provide convenient access to the ESG (e.g. in Europe nodes are BADC in UK and DKRZ in Germany). The ESGF portals are gateways to scientific data collections hosted at sites around the globe. Gateways are Web portals that allow the user to register and potentially access the entire ESGF network of data and services. Currently more than 24 portals are in use by the scientific community. All the data transfer is using standard internet links using http or ftp protocols although GridFTP is being trialled at some nodes to speed up the data transfers.

Another similar European initiative being developed at the DKRZ is the C3Grid which is being used as a German national climate data resource. It offers several processing options with different workflows. The aim is to provide a system based on a distributed "e-science" infrastructure composed of middleware services interacting with distributed compute and storage resources.

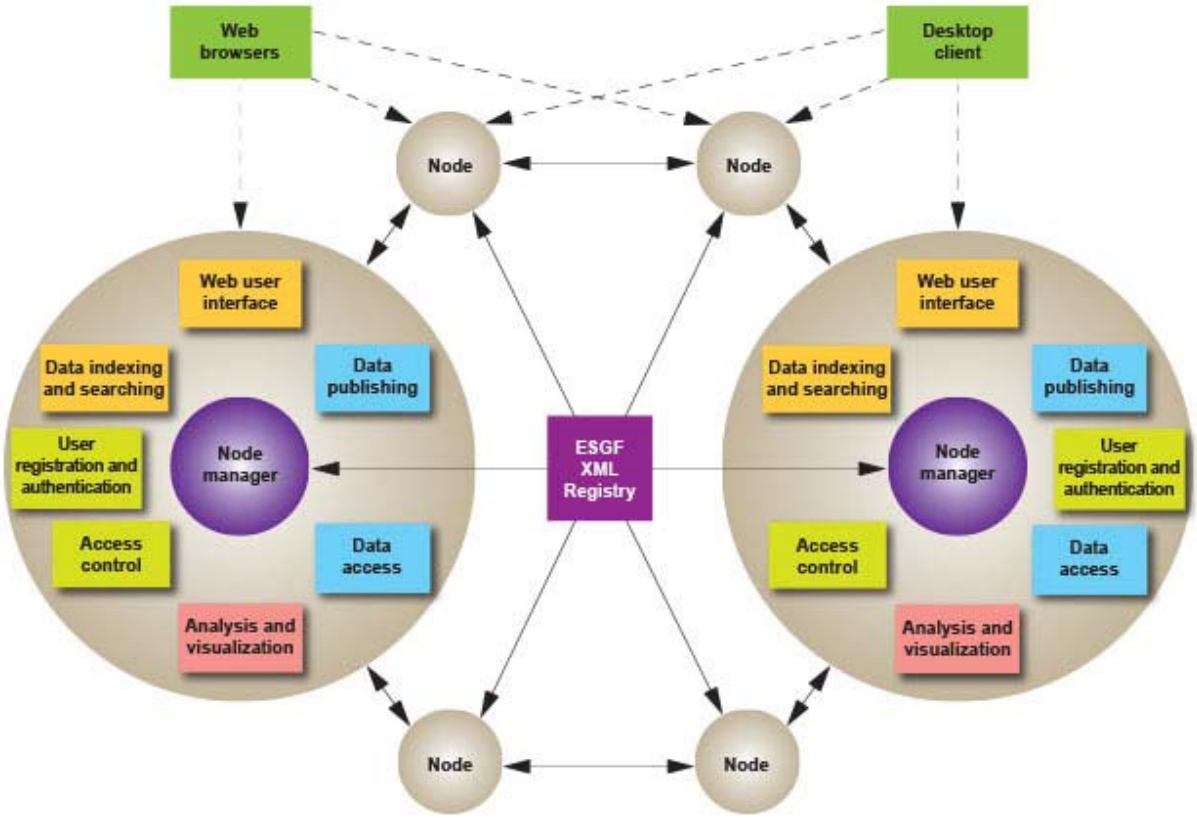

Figure 1. Diagram of the Earth System Grid Federation architecture.

Instead of consolidating at one location all the data from numerous sites, as was done for CMIP3 the precursor project, the ESGF uses distributed storage in conjunction with software that harvests and integrates the data. This approach was necessary as the data volumes were to high to ship round the world. Climate modelling centres download the ESGF software from the central servers at Lawrence Livermore. The modellers then use this software to publish their data needs to the ESGF for harvesting. The users are able to access all the data as if they are on one single archive system. Updates to model runs are also managed with interested users advised of the update. It is important to note that the ESGF is an extensible and scalable system which supports geospatial and temporal searches and includes a dashboard that shows system metrics, a user interface for notifications, and a set of analysis tools to help manipulate the data.

The majority of the international climate modelling community have participated in this project since its inception, to collect and exchange climate model data as part of the CMIP5 project. By having access to the ESGF the modellers in the CMIP5 project have been able to analyse a range of different climate models with ease in a systematic fashion.

An example is shown in Figure 2 where deep convective clouds for January and July are compared from 5 different models in CMIP5 together with the ISCCP cloud observations Tsushima et. al. (2012). Many of these kinds of analyses are informing the AR5 report.

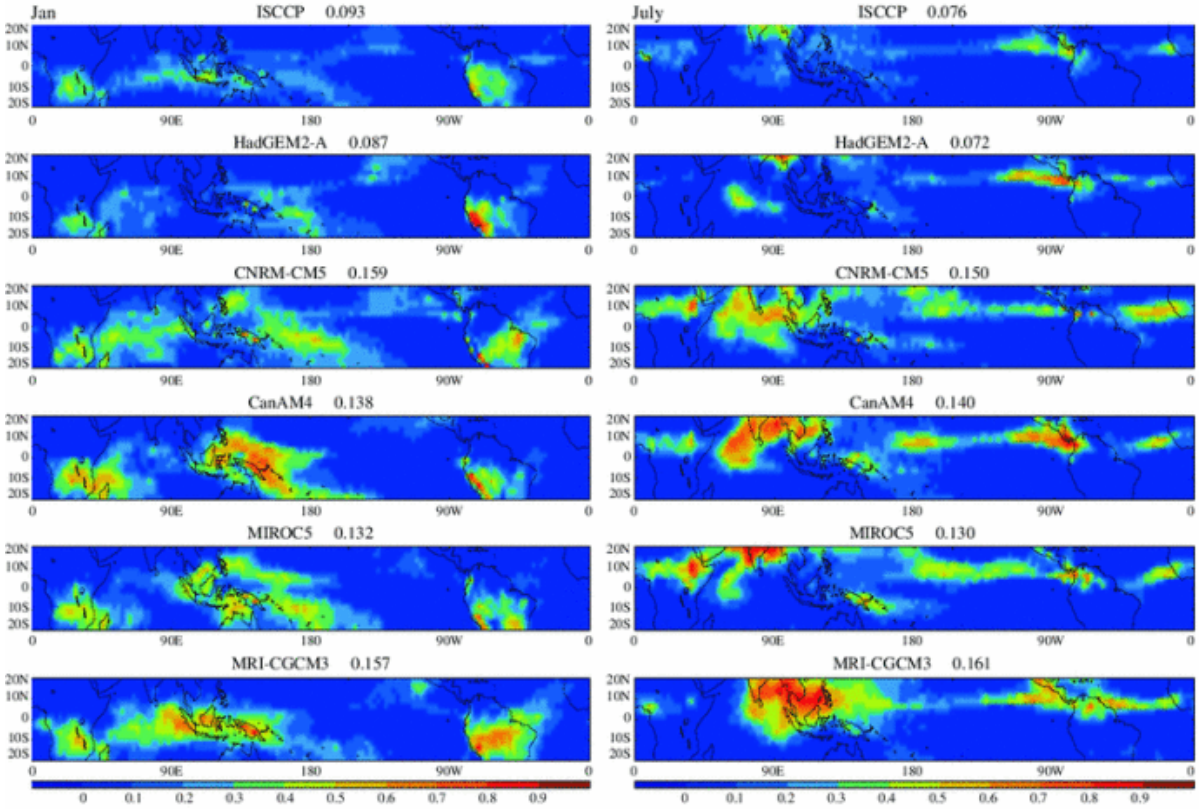

Figure 2. ISCCP and model-simulated deep convective cloud cluster in the tropics for January (left) and July (right) from Tsushima et. al. (2012).

For CMIP5, data volumes are estimated eventually to be over 3 petabytes (10^{15} bytes), two orders of magnitude more than in CMIP3. To date, ESGF has made 60 CMIP5 simulation model runs (more

than 1.8 petabytes) from 25 climate research centres, which are available to users worldwide. ESGF is providing access to an increasing number of climate data products.

It is proposed that for the CCI datasets which are converted to a form for comparison with climate model fields (i.e. primarily level 3 gridded products, of monthly means) this ESGF architecture might be appropriate to provide climate modellers access to the data. The Obs4MIPS project set up at NASA/JPL is providing a collection of well-established and documented datasets (many of them satellite) that have been prepared according to the CMIP5 model output requirements and made available on the ESGF. A list of the available datasets are on the Obs4MIPs site is here and some that were used in CMIP5 are illustrated in Figure 3. Simple documentation (no more than 5 pages) is also provided to guide modellers in the use of these satellite datasets avoiding the use of the 'satellite world' jargon. All the details on how to submit a dataset to the Obs4MIPs archive is available on the Obs4MIPS web site.

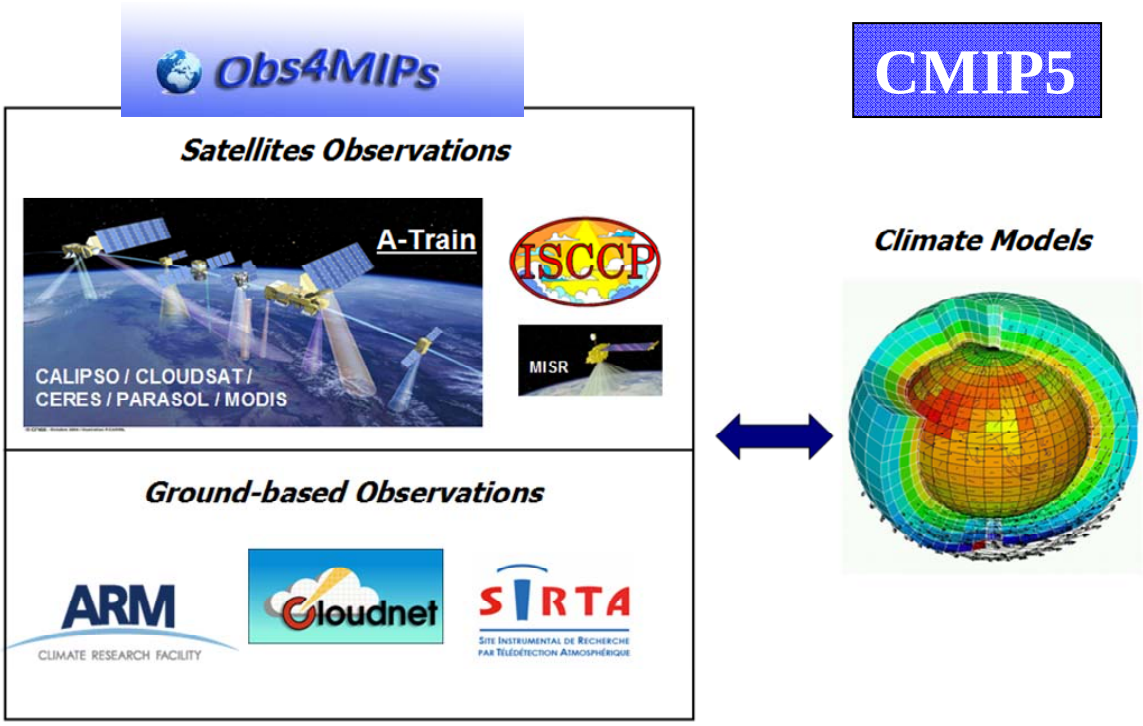

Figure 3. Observations on the Obs4MIPs site used in the CMIP5 evaluations

It also planned to provide analysis fields (e.g. JRA-55) in the same format in the related Anal4MIPs project. For example the MERRA data for AR5 are already available at the NASA Center for Climate Simulation on the ESGF. In the future the WCRP GDAC will oversee these projects through an international steering committee.

Metadata is also an important prerequisite associated with each dataset and the CHARMe project (Characterisation of metadata to enable high-quality climate applications and services) is helping to define a framework for this so that Metadata for each dataset is provided alongside the access to the dataset itself. Examples of the Metadata which can be included are:

- Post-fact annotations, e.g. citations, ad-hoc comments and notes

- Results of assessments, e.g. validation campaigns, intercomparisons with models or other observations, reanalysis, quantitative error assessments
- Provenance, e.g. dependencies on other datasets, processing algorithms and chain, data source
- Properties of data distribution, e.g. data policy and licensing, timeliness (is the data delivered in real time?), reliability
- External events that may affect the data, e.g. volcanic eruptions, El-Nino index, satellite or instrument failure, operational changes to the orbit calculations.

This project is underway to complete in Dec 2014.

Some CCI ECV datasets may not lend themselves to the Obs4MIPs interface and Table 2 provides a guide as to which CCI ECVs would be suitable for inclusion in the Obs4MIPs infrastructure. There are two European pre-cursor datasets which are already being prepared for input to the Obs4MIPs portal. These include the ATSR Reprocessing for Climate (ARC) SST dataset which has already been accepted on the site and the GlobVAPOUR dataset of SSM/I and MERIS total column water vapour. In preparing these datasets for the Obs4MIPs care has to be taken in a few areas. Firstly consideration has to be given to the spatial and temporal averaging if you have associated uncertainties with error covariances. Secondly how the averaging is managed at boundaries such as coast or ice edge is important. The CCI should have guidance on these aspects so that not every ECV team is making different decisions on this. There is also experience from the modelling community who have been putting their model fields into the CMIP5 archive in the same format for several years now.

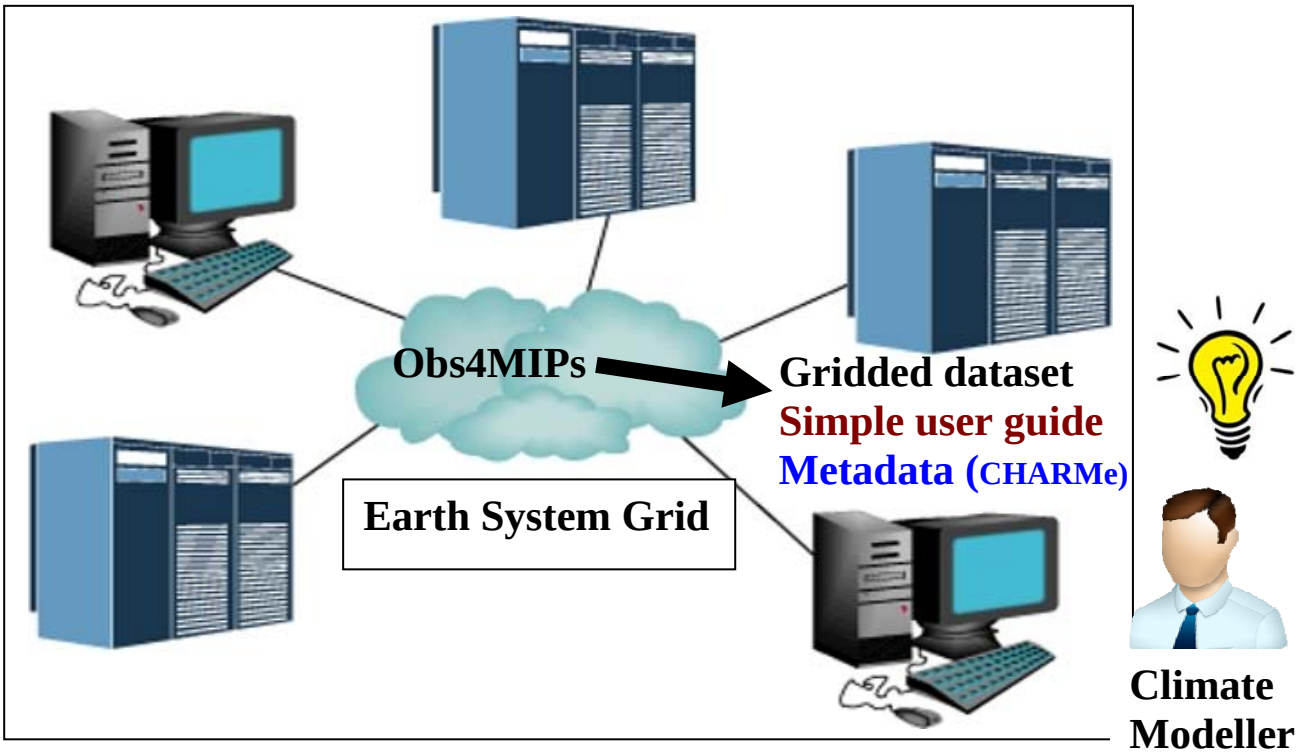

Figure 4. Schematic representation of climate modeller using the Obs4MIPS/ESGF.

ECV	Suitable for Obs4Mips?	2D or 3D field?
SST	Y	2D
Ocean colour	Y	2D
Sea level	Y	2D
Sea-Ice	Y	2D
Land cover	N	
Fire	N	
Soil moisture	Y	2D
Glaciers	N	
Ice Sheets	N	
Clouds	Y	2D
Aerosol	Y	2D
Ozone	Y	2D & 3D
GHG	Y	2D

Table 2 List of ECVs with recommendations of which are suitable for inclusion in the Obs4MIPs site

Once the satellite datasets are on the Obs4MIPs/ESGF climate modellers should be able to access all the ingredients of satellite climate data records they need for model evaluation as illustrated in Figure 4 and start producing results such as those presented in Figure 2. Assuming there will be a follow-on CMIP6 project the CCI datasets should become an integral part of the observational archive for climate model evaluation and getting the CCI data included in the Obs4MIPs project will be an important step to achieve this aim.

3. References

Tsushima, Y., M.A. Ringer, M.J. Webb, K.D. Williams, 2012, Quantitative evaluation of the seasonal variations in climate model cloud regimes, *Climate Dynamics*, doi: 10.1007/ s00382-012-1609-4.

CMUG Deliverable
Number: D2.2 Technical note on CCI system requirements
Due date: March 2013
Submission date: August 2013
Version: 0.7

Annex 1. Review of CCI System Requirement Documents

The documents reviewed and their versions are given in Table 1. Not all the documents were available to CMUG as indicated in the table and in some cases other documents were reviewed in place of these.

A general comment is that there is no real uniformity in the documents between ECV teams. In some cases there are additional documents (some of which are useful) and in many cases there are missing documents as is clear from Table A1. This makes the job of reviewer difficult.

ECV	Version of SSD	Version of PVP	Version of ATBD	Uncertainty characterisation
SST	v1.2 (07.05.12)	v1.0 (30.01.12)	v1.0 (08.11.11)	v1.0 (27.01.11)
Ocean Colour	V0.6 (27.11.12)	v1.3 (23.11.12)		v1.1(30.11.11)
Sea Level	v1.0 (02.08.12)	v1.1 (11.10.11)	v2.0 (19.12.12)	
Clouds	v1.0 (03.05.12)	v1.1 (28.03.11)	v1.3 (06.05.11)	
Ozone		V1.0 (03.01.12)	KIT_ATBD 0.01.04_small (10.01.12)	
Greenhouse Gases	Not available	v1.0 (20.05.2011)	v1.0 (15.03.2012)	
Aerosol	v1.5 (09.03.2102)	v1.5 (13.04.2010)	v1.0 (09.03.12)	v1.2 (13.12.11)
Glaciers and ice caps	Not addressed	Not addressed	Not addressed	Not addressed
Land cover	v1.0 (11.12.2012)	v1.3 (04.07.2011)	v2.1 (30.11.2012)	
Fire	v0.1 (01.03.12)	v3.1 (05.07.11)	Part 1 v1.2 (18.07.12) Part 2 v1.1 (14.01.13) Part 3 v1.1 (04.09.12)	
Sea-ice	Not addressed	Not addressed	Not addressed	
Ice sheets	Not addressed	Not addressed	Not addressed	
Soil moisture	not available	v1.1. (30.08.2012)	v0 (30.04.2012)	

Table 1. Version of documents reviewed

Also documentation on how the uncertainties will be derived is useful and some ECV teams have provided this in the uncertainty characterisation documents (e.g. ocean colour) which were reviewed when available. It would be better overall to include the uncertainty characterisation in the ATBDs as they are related to the retrievals. Another general comment is that there should be a version 2 of the ATBDs which only describes the chosen algorithm in more detail.

2.1 SST

System Requirements Document

Overall a useful summary of what is required to provide continuous improvements in the SST product by allowing several parallel streams to be run and by ensuring that the entire time period can be rerun from L1 to L2/L3/L4 in less than 2 months. This modus operandi was found to be invaluable during the precursor ARC project. This would be a good model for many of the other ECV teams and Fig 2.5 in the document illustrates the principle well.

The document was obviously written before the demise of ENVISAT but it does anticipate a gap between AATSR and SLSTR and how it might be filled using METOP AVHRR and IASI. Studies are underway to demonstrate the validity of this approach.

The L4 processing is done remotely at the Met Office using the OSTIA SST analysis. It would make sense to try and combine this with the core processing.

Product Validation Plan

The multi-sensor matchup database described in this document is a good model which provides the necessary data for evaluating various datasets. Again other ECV teams could learn from this. For SST the challenge is to relate the SST measured at depth (~20cm) by the buoys and that measured from the satellite which is the skin temperature.

There is also a good discussion on the various common metrics used for deriving the SST bias, non-systematic uncertainty and stability. Independence from in-situ measurements and the SST sensitivity are also quantities derived. The latter is defined as $\frac{\partial \hat{x}}{\partial x}$ where $\hat{x}$ is the estimated SST and x is the true SST. Finally the PVP contains an appendix which assesses how the metrics defined will show they will meet the user requirements.

Algorithm Theoretical Basis Document

It would be useful if a second version of the ATBD is issued by all ECV teams after the round robin comparison and the decision on which algorithms to use is made. The alternative methods would be removed and more details given on the chosen method.

Uncertainty characterisation Report

The SST ECV team seem to be taking the lead in this area which is good and section 2 I would suggest is mandatory reading for all ECV teams. Something like this should be incorporated within future GCOS requirements documentation.

2.2 Ocean Colour

System specification document

It is stated in the ocean colour SSD that:

"To get the full dataset or to get more than a few OC products the user may request for bulk delivery of a dataset. For the merged products the actual version is available in a cloud service for immediate download. For other products offline delivery is agreed between the user and the development team."

It is not clear this will meet the users needs as users may not want merged products for Ocean Colour but still require FTP access for the single sensor products. It depends what is meant by *bulk delivery*. This dataset would be a good candidate for the Obs4MIPS portal (see sec. 3) which is not mentioned.

Product Validation and Algorithm Selection Report
There is no comment on the report itself but it is understood that the ECMWF operational fields are used for the atmospheric correction not the reanalyses. This will introduce step functions into the products when there are changes in the ECMWF model which affect the total column water vapour. This will make it very difficult to use the ocean colour datasets for long term trend analysis. It is recommended that the ocean colour products move to using ERA-Interim as soon as possible.

Uncertainty Characterisation Document
This document is a useful discussion of the uncertainties and how they are derived for the ocean colour product chlorophyll alpha which could be adopted by other ECVs. It may be best placed as part of the ATBD. The uncertainties in areas of sunglint are not addressed and should be, as the part of the sunglint affected swath which can be used needs to be clear in the data.

For the assimilation of chlorophyll concentrations the observation errors used to date are based on a combination of O-B values (Hollingsworth-Lönnberg method) and forecast differences (NMC method). The values provided with the dataset, partly because these don't include representativity error are not used by assimilation centres. Nevertheless they are a useful input into the estimation of the observations errors. Having horizontal error correlations would be of interest for assimilation which is not proposed here. The uncertainties provided are also useful for validation.

2.3 Sea-level

System specification document
The SSD takes advantage of the existing DUACS system in particular for the definition of the architecture of the system. The document provides output of pre-designed activities and a preliminary design of the system. The system description and the strategy to define the architectural options are based on standards of system development and modelling. From a user point of view, this preliminary description takes appropriately account of the requests on data access, visualisation, metadata and documentation. The role of CMUG is also appropriately identified. One concern is how the chosen options (like THREDDS technology for access and visualisation) will be accommodated with a Pan-ECV system for the access to the whole set of CCI products or for the access to selected products through an existing infrastructure like the one put in place within the context of Obs4MIPs project. However, the document shows that at least for the pan-ECV system, this concern is taken into account (p 47-49, p 91-95 and appendix C). It should be extended to the link with Obs4MIPs.

Product Validation and Algorithm Selection Report
The PVP has the objective to define the validation protocol for both the round-robin intercomparison and the final product validation. It gives an appropriate focus on the interaction with the scientific expert team and on the diagnoses concerning long temporal scales. The input data take advantage of existing in-situ observations (tide gauges and ARGO T/S profiles). The appendix is a very useful description of validation diagnoses in synthetic sheets with illustrations. It appears from this list that 7 diagnoses over 24 are adapted to the final product evaluation. These are mainly statistics aggregated over the globe. A more regional (basins and latitudinal bands) evaluation and ocean-process based evaluation should complement these very useful, but very integrated, diagnostics.

Algorithm Theoretical Basis Document
The ATBD gives the theoretical basis for the algorithms that are used to calculate the components of the sea surface height calculation (orbit, distance between the satellite and the sea surface, mean sea surface height and corrections due to atmospheric effects of geophysical phenomena). For each of

these components, the document indeed provides a brief description of the chosen algorithm including input and output data, an estimate of the accuracy of the calculations, references and some illustrations of the results. It is mainly a technical document addressed to the specialists of the generation of sea surface height (SSH) from satellite altimetry. From an external user point of view it is not however clear if the document covers all the algorithms that are involved to compute SSH or only those that have been improved within the frame of the CCI project. A synthetic presentation of how the different sources of errors (precision, accuracy and stability) will be combined to infer the uncertainty estimate on the final product, maybe accompanied with an illustration on a case study, is missing (or a reference to a document including this kind of information).

Product User Guide Document
The PUG contains the description for the FDCR and ECV data products. After a very brief overview on the determination of SSH from altimetry, the document gives a list of references by mission for the different components of the SSH calculation. Since the document mentions the availability of two versions of the products (v0 and v1), it could be of interest to have the references for the two versions (here only given for the version 1). The FCDR and ECV products are then defined, the file nomenclature and format are described and a detailed list of metadata and variables are given. The information on product uncertainties is not currently available.

2.4 Landcover

System specification document
There are no specific comments regarding the SSD.

Product validation plan
The PVP describes in detail the validation strategy for the CCI landcover product. The CCI landcover validation strategy is based on three basic components:

1. Validation using independent data
2. Comparison against already existing data products
3. Validation of temporal consistency of the product

The validation relies on expert knowledge which is supported by sophisticated analysis tools for validation. The document contains a comprehensive list of local experts for different continents who will be responsible for the validation. A random sampling strategy is applied to perform an independent validation of randomly chosen control points using a GUI based interface that supports the experts. Further comparison will be done against already existing land cover data products.

The document also contains a brief description on the validation of the temporal stability of the dataset.

There are two major criticisms regarding the current version of the PVP:

1) The CCI land cover team employs a new approach to describe the land surface state. The PVP does not address the validation of this land surface state.
2) For users who are interested in using CCI land cover in combination with CCI fire products, it would be important that the CCI land cover and CCI fire datasets are consistent. This would mean a) that they match geometrically and b) that the CCI land cover states reflect the fires identified in the CCI fire product. The PVP does not provide any information on achieving and validating this inter-product consistency.

Algorithm Theoretical Basis Document

The ATBD provides a comprehensive overview and insight into the algorithms to generate the CCI land cover product. Major criticisms are that a reproduction of the CCI dataset would not be possible from this document. Most critical is the labelling of the clusters obtained from the clustering of the input data. The labelling procedure acts like a kind of a black box and the document does not provide sufficient information to reproduce the labelling procedure at a later stage.

2.5 Soil moisture

System specification document
Not applicable as document not available so far.

Product validation plan
The PVP provides detailed insights into the planned validation of the CCI soil moisture product. The validation will be performed using a multitude of approaches, ranging from comparisons with point like observations up to continental scale validation. The following general approaches will be applied:

- Comparison against in situ dat (ISMN) (local scale)
- Validation using runoff data (catchment scale)
- Validation using Terrestrial water fluxes and evaporation estiamtes (continental scale)
- Validation using precipitation observations (R-metric; regional scale)
- Validation using SAR based soil moisture estimates

Random errors of the data products will be quantified using triple collocation analysis techniques.

The PVP clearly defines the different datasets and metrics that are planned to be applied for the CCI SM validation. The different quantities to be investigated (e.g. definition of soil moisture anomalies) are described and various score skills are defined.

The document also discussed longterm trend analysis validation. In particular it is planned to ensure that trends are preserved during the data merging procedure employed in the product generation. It is however not discussed, how it is planned to avoid spurious trends in the data which might be caused due to changes in the observation system.

Algorithm Theoretical Basis Document
The ATBD provide details on the envisaged CCI soil moisture ECV production chain. It is basically a consolidated overview about the different techniques developed to generate the ECVSM precursor dataset, which was generated in the frame of the ESA WACMOS project. Some of the techniques presented in the document have been already published in the literature.

The documents describe in detail the following major aspects of the ECV production system:

1) Generation of a harmonized L1 dataset from passive and active microwave observations
2) Retrieval of soil moisture from active data
3) Retrieval of soil moisture from passive data
4) Error propagation for active data
5) Error propagation for passive data
6) Blending approach

2.6 Greenhouse gases

System Specification Document
This document is not yet available.

Product Validation Plan
The PVP clearly presents the overall plan for the validation of the ECV GHG. The document also provides an overview of the Round Robin (RR) evaluation protocol.

The ECV GHG validation will be performed in three stages during Phase 1, roughly once every year. By the end of stage 3 of the validation procedure (i.e. validation of the final ECV GHG), ground-based measurements from the TCCON and NDACC archives will have been considered. The matching criteria of 2 hour maximum time-lag between observations and a selection of three maximum distances (100, 350 and 500 km) in space seem sensible. Additional in-situ surface data will also be used together with models that convert the surface information into a total column dataset. There is also a good discussion on the characteristics and shortcomings of the used datasets, as well as on the validation of the NDACC data based on TCCON measurements. The latter offers the advantage of using the same measurement technique of the satellite data but does not consist of long data records, while the former can provide long time series of measurements but obtained at a slightly different spectral range that only partly overlap the TCCON one.

One problem is that altogether the ground-based datasets are rather sparse and that entire regions of the globe are not at all covered. This could have consequences when generalizing the results of the validation study.

Algorithm Theoretical Basis Document

The current version of the ATBD was issued before the RR exercise. As it stands it is a collection of five ATBDs, one per algorithm used in the RR phase. It is mentioned that a version 2 focussing only on the selected algorithm would have been issued at the end of the RR, but it is not yet available. The current document is mostly based on material already published in the literature.

When reading the current document, comparing and understanding the differences between the various algorithms is not always straightforward. It is appreciated that the authors have tried to summarise the main characteristics of each scheme in the first fifty pages or so of the document. Though, a table (summarizing e.g. retrieval method, state vector composition, a priori source, aux data, output data, etc...) would also have helped and eased the reading.

Auxiliary data: It seems that most algorithms make use of the ECMWF meteorological data. However, apart from a couple of them (one clearly using ERA-Interim and the other likely using the operational high resolution data) it is not clear which ECMWF archive the meteorological datasets are taken from. It is recommended to always specify the name of the used archive. Furthermore, it is reminded that for the CCI ECV data production the ERA-Interim data should be used as aux information instead of the data from the weather archive.

2.7 Aerosols

System Specification Document
It was unclear reading this document how the routine validation using measurements such as Aerocom would be easily carried out. The generation of match-up datasets for example would help to facilitate easy validation of a new algorithm with in situ observations. Comparisons with model fields and other satellite products (CALIOP) should also be considered as part of the assessment mechanism to determine if a new product is better than the previous one. Modellers will need information about how to distribute the aerosol concentrations in the vertical.

Product Validation Plan and Product Validation and Algorithm Selection Report
In the PVP there is a good description of observational datasets to compare with the CCI products but not much consideration of comparison with models which include aerosol as a prognostic variable (e.g. MACC).

The PVASR describes the round robin protocol and lessons learnt. The assessment of the uncertainties in the different products does not appear to have been evaluated.

Uncertainty Characterization Report
Section 2 contains a good description of the definition of the different components of uncertainty. In comparisons with models representativity error also needs to be included which is not referred to here. As there are several different measurements made of aerosol optical depth which could be co-located the 3-way estimation of observation errors could be tried for example with Aeronet, AATSR and Parasol.

2.8 Fire

System specification document
There are no specific comments for the climate modelling perspective regarding the SSD.

Product Validation Plan
The product validation plan reviews existing validation methods for burned area products. Two steps in the validation procedure were identified (1) selection of study areas to test the performance of the burned area algorithm (2) validation of the global BA products. 10 study sites were selected and described in detail in the DARD. For the global burned area product 100 validation sites were defined dependent on average fire density and vegetation type representativeness. Primary source of reference information is multi-temporal analysis of Landsat TM/ETM+ images, which will be processed within the project. External reference data will be used as well when the data meets the same quality criterions as the fire CCI generated ones. A validation document will be made available after the round robin exercise. The validation will be done only for the pixel based product. The PVP does not clearly state for which time periods the validation will be performed. The global product will be only made available for selected years. Will the validation be performed for these years? The spatial coverage of the test sites is derived in a way to reflect a large variety of fire densities and vegetation types. The BA product will be a result of a merging of different satellite BA products. It is not clear if the validation sites are suitable selected to evaluate the merging procedure.

Algorithm Theoretical Basis Document
Volume 1: Pre – processing
Describes the pre-processing chain of (A)ATSR, MERIS and SPOT VEGETATION for the test sites and the global processing. The next version of the document will include more detail on how cross ECV consistency will be assessed in data pre-processing. Currently, there are no specific comments from the climate modelling perspective regarding the ATBD–Volume I.

Volume II: Algorithm Development
Presents the primary algorithm developed by the (A)ATSR, SPOT VEGETATION and MERIS teams, for which (A)ATSR and SPOT VEGETATION apply a similar algorithm. There are not specific comments from the climate modelling perspective regarding the ATBD – Volume II.

Volume III: Burned area merging

The fire CCI BA product will be a result of the merging of two different burned area algorithm processing six satellite data sources. The document summarizes possible merging strategies. It is, however, not clearly stated which merging strategy will be used. The merging strategy "Land Cover mosaic" may benefit from an assessment of the CCI landcover data, which will be available for several years.

Regarding the data output for the global gridded product: "Dominant vegetation burned" should be replaced with relative vegetation type burned for each landcover class (see also CMUG D2.4). This information is available in the pixel based product and should be aggregated accordingly into the gridded product. This information is essential when integrating BA products into global vegetation models to simulate the fire emissions into the atmosphere.

2.9 Ozone

System specification document
Not available.

Product Validation Plan
The document is very well structured and gives a comprehensive view of the validation plan . It first recalls the user requirements with respect to each product: Total Ozone column, Nadir profile and limb profile before describing the Product evaluation protocol inspired by QA4EO and other valuable principles. Validation includes confrontation with ground-based references from WMO's global atmosphere watch programme WOUDC and NDACC. Other satellite data are also used. The error budget in this comparison takes into consideration the difference in time, horizontal and vertical sampling. For profiles, errors are based on information content theory. Overall, validation also incorporates validation of individual components to eliminate any compensation in various modules, validation against services, against user requirements, quality control and validation of updates. Specific details for each of the 3 FCDRs are given.

Prior to the validation of final products, products from an interim baseline algorithm (for total column of ozone), or products from various algorithms or assembling various modules are evaluated in the round robin. In this case, the method is the same as the one described for the validation but with confrontation of a limited set of products to correlative measurements. It addresses the compliance with the ESA criteria which are all met but three which are of importance. External experts are not involved in the reviewing of the PVP (V-6). The selection criteria for the pre-selected algorithm dedicated to total ozone must be described in a document, e.g. the ATBD (no reference given here) (RR-3). The algorithm selection was not performed by an independent team (RR-6). The quality flag would also need to be validated, as it could be used by end users to screen good quality data.

Algorithm Theoretical Basis Document
This document is only available in a preliminary version. A very good description of errors in the inversion process is given and terminology is revisited. The impact of level 1 processing especially calibration, pointing accuracy or cloud screening is not addressed sufficiently while it can be a major source of error: calibration, stability of sensitivity could lead to wrong trend, cloud screening to artefacts, pointing stability of limb sounders to errors in profiles. If possible, cross correlation with other sensors (e.g intercomparison of GOME2-A with GOME2-B) could allow better characterization of level-1 data quality and corrections of the products at this level before inversion. Update of the ATBD with the actual algorithms is requested.

Product User Guide

Not available.

Uncertainty Characterization Documents
Not yet available.

2.10 Clouds

System specification document
This appears to be a comprehensive summary of the CCI cloud system and there is little of direct relevance to the modelling community – as would be expected for such a technical document. Two points which may be worth noting are: (i) Reference is made to the development of a product visualization interface but there is no clear indication of who the potential users of this will be or how it fits in with the other ECVs; (ii) There is no mention of any link to Obs4MIPs/ESGF.

Product Validation Plan
This document provides a summary of the CCI's activities which will involve 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.

Algorithm Theoretical Basis Document
This is essentially a collection of separate documents, each of which describes the different candidate algorithms which participated in the round robin exercise. This is includes the ORAC algorithm, which was the one that was eventually chosen by the CCI. There is currently no specific CCI ATBD, which would include, for example, modifications to the original ORAC algorithm made as part of CCI cloud product development process.

Product User Guide
Not currently available.

Uncertainty Characterization Documents
Not currently available.