Document Ref.: D1.1

CMUG Phase 2 Deliverable

Number:	**D1.1: Requirements Baseline Document**
Due date:	**October 2014**
Submission date:	**April 2015**
Version:	**0.6**

Climate Modelling User Group

Deliverable 1.1

Requirements Baseline Document

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

Version nr.	Date	Status
0.22	October 2014	Added partner ECV contributions
0.3	November 2014	Consolidated version for final input and check by partners
0.5	November 2014	Submission to ESA
0.6	April 2015	Updated Table 19

Deliverable 1.1

Meeting the needs of the Climate Community – Requirements

Contents

Requirements Document

1. Purpose, scope and construction of the Requirement Baseline Document

The purpose of this document is to assist the CCI projects in focussing on the needs of the Climate Modelling Community (CMC) and other expert users of climate data. It aims to do the following:

1) present an analysis of the satellite climate observation data requirements of the CMC and other expert users of climate data. The requirements were captured by CMUG through interviews with 75 experts. The responses given by climate modellers are representative of the full range of models and the applications operated by them. Experts outside the CMC responded with information from the areas of climate services (including the Copernicus Climate Change Services and national Climate Service Centers), detection and attribution of climate change, climate process studies and climate/environmental monitoring.

2) cover both the requirements for the 13 ECVs in terms of parameters, resolution and errors/uncertainties and also where appropriate cover the requirement for observation operators for each of the ECVs.

3) address the requirements for CCI datasets to be included in the Copernicus Climate Change Service and the Obs4MIPs interface.

4) cover overarching technical requirements and scientific linkages for the datasets produced.

This document confirms and builds upon the user requirements inventoried by CMUG in Phase 1[1]. The new information found here, compared with the Phase 1 survey results, is of greater detail in describing user needs, from a wider base of users interviewed (Copernicus requirements were not part of Phase 1) and, from users experienced in usage of CCI data. It is acknowledged that the climate data needs of the climate research community are evolving and that the CMUG, through its interactions with this community, will ensure that its knowledge of user requirements is up to date and relevant.

A key example of user requirements across the CMC applies to the Obs4MIPs[2] initiative (Teixeira et al. 2014) that provides an archive of gridded Earth system observations to facilitate model evaluation in the recently started CMIP6 project (Meehl et al., 2014).

[1] CMUG Phase 1 Deliverable 1.2: User Requirement Document (v2.0), available at http://www.esa-cmug-cci.org.
[2] http://obs4mips.llnl.gov

Individual CCI datasets in CMIP compliant format for Obs4MIPs and supporting technical notes need to be created and submitted by the individual CCI teams. The CCI data submitted to the Obs4MIPs will sit alongside other observational data used for model evaluation and will be implemented into the ESMValTool as part of CMUG to routinely benchmark the models against ESA CCI data. Four Work Packages in CMUG Phase 2 are concerned with the application of CCI data to evaluate climate models through the ESMValTool.

2. Introduction

The climate is continuously changing, so climate researchers need to measure its changes globally and regionally, and to model the climate system to understand and attribute the causes of the changes. Given their global and temporal coverage and spatial resolution, satellite data, which now for some variables span more than 30 years, can potentially be used for both climate monitoring, and model initialisation and evaluation provided certain requirements can be met.

The uncertainties of the satellite datasets must be understood and quantified; otherwise little confidence can be placed in the derived climate data records. Because most of the measurements were not taken with climate applications in mind, the data need careful preparation for climate monitoring. Also, satellites do not make localised 'conventional' *in situ* measurements of e.g. temperature or moisture as represented by climate models, but measurements of indirect parameters e.g. upwelling radiance or GPS signal refraction angles. For some parameters climate models can deal with this by including 'observation simulators' to compute the variable measured by the satellite from the model fields, thus avoiding the uncertainties in the retrieval of conventional variables from satellite data. However it is important that these simulations can be interpreted in terms of standard geophysical variables, or physical properties such as humidity, cloud drop size or crystal shape, as model parameterisations are often framed in terms of these physical quantities.

Climate researchers usually confront models with observations with the following aims:

- To interpret the observations and explain the causes of observed variability and change
- To evaluate, constrain and improve climate models, thus gaining confidence in their projections of future change
- To initialise models for reanalyses, seasonal and decadal timescale predictability (data assimilation) and to provide representative initial conditions for climate model simulations
- To prescribe boundary conditions of quantities that are not prognostic variables in climate models

Accordingly, the generic requirements for satellite data are:

- to provide long term monitoring datasets of particular parameters with or without *in situ* data to ascertain decadal and longer-term changes Models can then be used to attribute the observed variations to natural and anthropogenic forcings and internal variability (IPCC WG1 AR5, 2013);

- to provide long term sets of validated, high quality climate data, with good uncertainty characterisation and documentation for Earth system model evaluation.
- to compare measured parameters, or combinations of observed and/or reanalysed parameters, with model equivalents on hourly up to decadal timescales, to assess the processes and biases in the models and if necessary to constrain, the processes.
- to initialise seasonal forecasting models, for example with realistic estimates of soil moisture and sea surface temperature.
- to help evaluate the skill of seasonal to decadal forecasts.
- to interpret short term variations of the climate in the long term context, as in the recent hiatus in observed surface warming.
- to help identify biases in the current and past *in situ* observing network. Comparisons of Microwave Sounding Unit (MSU) retrievals to "families" of radiosondes for identifying shortcomings both in the raw radiosonde data and the satellite datasets.
- to provide homogeneous data, with good estimates of random errors and bias-correction uncertainties, for reanalyses. Existing reanalyses are already very useful for model evaluation, especially in combination with independent satellite data; but the next generation of reanalyses also needs to be sufficiently homogeneous to allow the estimation of long-term trends (Simmons et. al. 2014). In addition, especially in areas with sparse sampling like the Polar regions, different reanalysis products differ significantly from one another.
- to provide long term sets of validated, high quality climate data, with good uncertainty characterisation for use by climate service providers.

Now that satellite climate data records are reaching 35 years in length they have become an important source of data for use in climate research, hence the CMC and climate research community need to make best strategic use of the emerging opportunities provided by satellite data. Only after quality assurance is demonstrated, can high quality climate datasets be produced that are fit for onward use in an operational or wider societal application. Opportunities for exploitation of the CCI datasets now exist in various emerging activities related to climate services in both national and international arena. Also the improved interface to climate modellers provided by the planned ESA portal and Obs4MIPS project are other channels to lead to the uptake of CCI data. Providing the CCI data to these interfaces imposes certain requirements on the datasets which are given here.

Section 3 identifies in more detail those generic application areas where satellite datasets are required for climate modelling. Section 4 outlines the specific requirements for the satellite CDRs for the 13 ESA ECVs and section 5 lists cross-ECV requirements. Section 6 lists the requirements for other ECVs. Sections 7 and 8 cover the requirements for climate services and Obs4MIPS. Section 9 gives the requirements for observation simulators and other tools required by climate modellers to exploit the datasets. Section 10 outlines the technical requirements for data formats, projection, access etc. Section 11 summarises the key point of this report. A list of acronyms and definitions of various terms is in section 13. Finally Annex

A gives a definition of the error characteristics and Annex B the datasheets which were used as input to this requirements analysis.

3. Generic requirements for climate applications

Table 1 summarises the generic requirements for the 13 ESA CCI ECVs from a recent survey (August to September 2014) of experts from climate modelling centres that was conducted by the CMUG. It also lists in the bottom row the responses from the CMUG questionnaire as to what the CCI CDRs will be used for. All application areas are mentioned but the comparison with models for model evaluation and development dominated the uses. It should be noted that the high number of experts who are using, or intending to use, CCI datasets for model development and validation will be well served by CMUG work on developing an ESMVal tool. Although CMUG asked a broad range of experts, not many users questioned are engaged in long term climate monitoring and attribution studies.

An important requirement for all the CDRs is to include their associated errors for each observation where possible and to document the dataset and its uncertainties well. For many applications it is crucial to have an associated precision for each observation. Also the error correlations between variables are important to consider.

GCOS ECV	Model Initialisation	Prescribe Boundary Conditions	Re-analyses	Data Assimilation	Model Development and Validation	Climate Monitoring/ Attribution	Q/C in situ data	Climate process study
Atmospheric								
Cloud properties	X	X					X	X
Ozone	X	X	X	X	X	X	X	
Greenhouse gases	X	X	X	X	X	X	X	
Aerosols	X	X	X	X	X	X		
Oceanic								
SST	X	X	X		X	X		
Sea level	X	X	X	X	X	X		
Sea-ice	X	X	X		X	X		
Ocean colour				X	X	X		
Terrestrial								
Glaciers and ice caps	X	X			X	X		X
Ice sheets	X	X			X	X		X
Land cover	X	X	X		X	X	X	
Fire	X	X		X	X	X	X	
Soil Moisture	X	X	X	X	X	X	X	X
Users responses								
Declared uses	36	34	23	22	71	39	11	7

Table 1. Use of CCI ECVs for different climate applications

3.1 Climate monitoring and attribution

Satellite datasets need to span at least several decades in order to meaningfully monitor climate change. Some satellite datasets already approach 35 years in length, but many are shorter than 20 years although continually expanding.

Climate monitoring implies the most stringent requirements for satellite data both in terms of stability of the measurement and in the minimum time period of the dataset. In addition significant overlap periods between successive sensors as recommended by the GCOS monitoring principles (See Annex 2 of GCOS, 2010) is also a crucial requirement to ensure the fidelity of the time series.

Time series of greenhouse gas, ozone and aerosol concentration profiles and total column amounts are important for trend analyses to assess if there are significant increases or decreases in these atmospheric variables which will affect the atmospheric radiative heat balance. The global coverage allows regional and/or temporal variations to be investigated and potentially attribute them to natural or anthropogenic causes.

For the ocean ECVs sea level, sea-ice coverage and thickness are critical parameters that must be monitored as key indicators of climate change. Sea surface temperature similarly is an ECV which has been monitored by in-situ observations since the mid 1800's and so is an excellent indicator of climate change. The complication with satellite measurements of SST is that they measure the skin not the bulk SST and so a "correction" has to be made to the satellite CDRs of SST to obtain a "bulk" SST as would be measured by ships and buoys. This is an example of the need for an observation simulator (see section 9.1). The record for ocean colour measurements is relatively short but when the length of the time series reaches >20 years this will provide another important indicator of climate change.

For the land surface, fires are important to help monitor and understand the carbon cycle. Records of fire numbers and burnt area help to show the amount of deforestation occurring in the last 2 decades. The extent of ice sheets, glaciers and ice caps is also an important indicator of climate change and the satellite data can complement the ground based observations. Land cover type is an ECV required as a model surface field as it can affect the local radiation and provide sources and sinks of various atmospheric variables (e.g. aerosols, CO_2, CH_4 etc). All NWP and climate models use land cover to initialise their land surface models. Information on soil moisture dynamics is of major importance as soil moisture has a primary effect on the land surface memory and the partitioning of surface turbulent heat fluxes. Because of its importance to e.g. better monitor and predict heat waves, soil moisture is also assimilated in NWP models and used for the initialization of seasonal to decadal climate prediction systems.

A new area of concern in climate monitoring is the assessment of rapid climate changes which requires confidence in the prediction of the thermohaline circulation and carbon cycle/sea ice non-linear feedbacks. Close monitoring of greenhouse gas concentrations and sea-ice coverage/thickness from satellites is important to provide early warning of any sudden changes. Fire and vegetation changes are also examples of variables that can change rapidly and have significant impacts.

Finally there are some satellite derived metrics, which are not ECVs as defined by GCOS, but nevertheless are of interest. Severe weather events such as the annual number of tropical cyclones in each ocean basin, frequency of intense extra tropical storms, severe drought episodes and heat waves are all of interest for climate change and applications studies and can be inferred from satellite data with some effort. There is a need from policy makers and other users for a better understanding of the risk of current extreme weather events and the extent to which this risk has changed as a result of human influence. Some of the ESA ECVs may contribute to these metrics and the requirements will need to reflect this.

The requirements for climate monitoring measurements are stringent. For example, an SST decadal trend of 0.2°C per decade requires the satellite CDRs to have a stability of <0.05K. It is important to distinguish between stability and accuracy here (see definition of these terms in Appendix A). For climate trends the measurements have to be stable over long time periods and any changes must be understood and be able to be accurately modelled. Requirements on the bias (accuracy) can be less stringent so long as there are other complementary measurements to compare with. The GSICS project is putting in place an infrastructure to provide these measurements to estimate and monitor biases in different sensors. Therefore one of the requirements on some of the ECVs is that they make use of the GSICS measurements to ensure their accuracy can be traced back to International Standards as addressed by the WMO QA4EO project[3]. Traceability from satellite measurements through bias correction to ECV data is essential for the integrity of any Fundamental Climate Data Record. The GSICS initiative[4] is therefore crucial to improve the quality of the global satellite datasets. Another initiative, the QA4ECV EC FP7 project, is also developing a system for quality controlling ECV datasets so that they have ‘climate quality’ with respect to both observed long term trends and variability. This project is driven by the user needs of the Climate Services community. It is noted that quality assurance in CCI data production chains is a requirement for production of long term climate quality data

3.2 Model initialisation and definition of boundary conditions

A major requirement for satellite data to date has been to help define the initial state of the atmosphere/surface for NWP models and decadal prediction systems along with conventional *in situ* data. The ECMWF Reanalyses are important examples of this. An example of this is shown in Figure 1 from the ECMWF ERA-40 reanalysis where the link between total column water vapour and sea surface temperature becomes closer once satellite data are available (from 1972 onwards).

3 Quality Assurance Framework for Earth Observation (QA4EO) [http://www.qa4eo.org/]
4 http://www.star.nesdis.noaa.gov/smcd/spb/calibration/icvs/GSICS/

For initialisation of 'present-day' coupled climate control experiments the atmospheric state (provided the latter is in reasonable balance) is not so crucial as in principle, the model should equilibrate to its own climate state no matter what the initial state is, but it is still preferable to start from accurate initial conditions in order to avoid big adjustments that take a lot of computational time to settle down, and to be able to judge the growth of errors without massive drifts.

Some of the 13 ESA CCI ECVs have potential for model initialisation (see Table 1) primarily through improving the representation of the surface fields. The stability and accuracy requirements for initialisation are more relaxed than for climate monitoring as the initial uncertainties in the model fields without the observations are often far greater than the measurement uncertainty. Ozone is a good example of an ECV where the model uncertainty easily exceeds the measurement uncertainty.

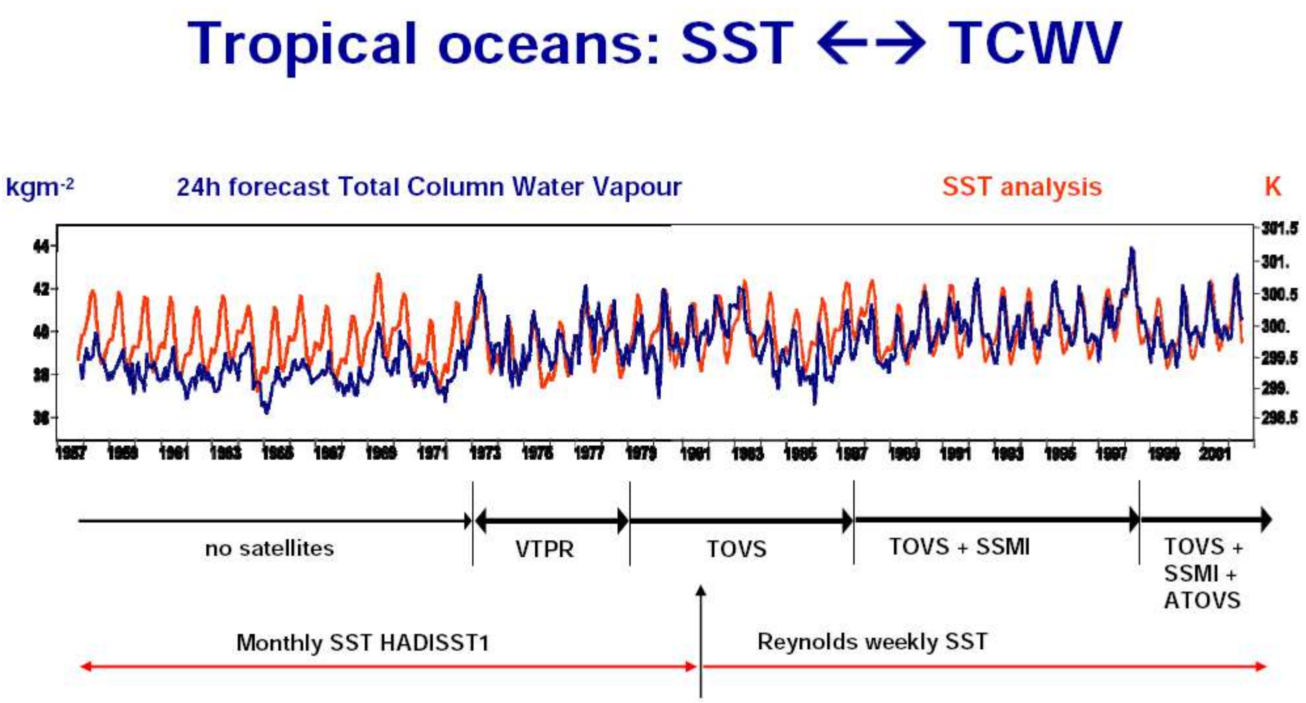

Figure 1. Correlation between total column water vapour and SST in ERA-40 before and after satellite data were introduced.

3.3 Model Development and Evaluation

Satellite observations are a key part of the development and evaluation of climate models. Banks et al. (2008) present assessment criteria for the Hadley Centre model, HadGEM3, where components of HadGEM3 were found to be sensitive to atmospheric and ocean fluxes, e.g. land surface temperature (particularly northern continental summer temperature), rainfall over land (particularly Indian sub-continental rainfall in northern summer), soil moisture, and dust concentrations over both land and ocean (Banks et al (2008)).

Coupling the various components is a priority. For instance, the coupling between atmospheric chemistry (air quality, oxidation, stratosphere-troposphere processes, ozone hole, etc) and climate is important. Although the current generation of tropospheric ozone models is generally successful in describing the principal features of the present-day global ozone distribution, there is less confidence in the ability to reproduce the changes in ozone associated with perturbations of emissions or climate. There are discrepancies with observed long-term trends in ozone concentrations over the 20th century (Shindell et al., 2003; Lamarque et al., 2005, Parrish et al, 2014, Cooper et al, 2014), including after 1970 when the reliability of observed ozone trends is high (Fusco and Logan, 2003). Resolving these discrepancies is needed to establish confidence in the models. Consistency between the processes described in the models has to be checked. The observations of the various ECVs allow to check this consistency and if appropriate help to improve the bio-geo-physical-chemical schemes used in the models.

Long term vertically resolved data sets of constituent observations are required to assess Chemistry Climate Models (CCM). This includes ozone, but also other species that are used to diagnose processes involved in CCM: transport, chemistry, radiation, and dynamics. Such observations are required by CCM validation exercises like CCMVal-2 (see overall recommendations in executive summary, http://www.atmosp.physics.utoronto.ca/SPARC/CCMVAL_FINAL/index.php).

For the ESA ECVs clouds, aerosols and trace gas concentrations are important to validate the model fields. For example the accurate representation of clouds in climate models is important to reduce the range of uncertainty in climate sensitivity studies. Datasets of cloud properties (i.e. fractional cover, top height, phase, microphysical properties etc) provide an important constraint for climate models. Cloud droplet size and drop number concentration are also variables of specific interest. Regional estimates of all these parameters will be important for detection/attribution studies. In addition instantaneous estimates of cloudiness are also important to monitor the diurnal to annual cycles of cloud. In order to compare satellite clouds (e.g. from ISCCP) with model clouds a cloud simulator (sec 9.1) is needed. The MOHC has developed the COSP (CFMIP Observational Simulator Package; http://cfmip.metoffice.com/COSP.html) to enable such comparisons.

The oceanic ECVs also provide important inside into model quality. For example, some of the longest-standing biases in most large-scale model simulations relates to sea-surface temperature biases in the low-latitude ocean around South America. Another bias relates to the trend in Antarctic sea ice, where observations show a slight regional increase but model simulations project a decreasing sea-ice cover. For better understanding and eventually reducing these biases, reliable satellite observations of oceanic variables is crucial. Also possible uncertainties of satellite retrievals must be known, but these have usually not been provided by products before the ESA CCIs.

The requirements on accuracy for model evaluation are less stringent. It depends on the magnitude of the model error but in all cases the requirement should be more relaxed than for climate monitoring.

Evaluation of climate models through the systematic application of community tools with agreed benchmark datasets was in common use by the time of the CMIP5 evaluation of climate models for the IPCC AR5 (IPCC, 2013). The next generation of tools for evaluating Earth System Models continue their development under the ESMVal initiative, which is a collaborative research activity of several research centres and universities, and will support CMIP6 studies. The evaluation of climate models operated by CMUG partners using CCI datasets and the latest ESMVal tool will help to (a) validate the models used by CMUG partners, (b) evaluate the CCI data sets used, and (c) better develop the ESMVal tool.

3.4 Input to reanalyses

Global and regional atmospheric and ocean reanalyses are now being undertaken in a number of centres to provide a consistent analysis of the atmosphere over a long time period, typically 40-100 years using an NWP model as a constraint for the variables. Increasingly these reanalysis datasets are being used for climate applications. A key requirement for the data to be assimilated into these reanalyses is that they are uniformly processed without the discontinuities often seen in operational real time processed datasets caused by changes to operational real time processing of the instrument data.

Accordingly, satellite climate data records are well suited for reanalyses provided they come from a stable processing environment and provide associated error estimates. For the recent ECMWF reanalysis (ERA-40) satellite agencies did make an effort to provide some homogenous datasets for example the atmospheric motion wind vectors provided by EUMETSAT where the products from the early years were much improved with reprocessing.

In general, re-analysis applications require single-sensor products rather than merged products. Furthermore, these applications often ingest Level-1 satellite data rather than Level-2 retrievals and thus there is a strong interest in uniformly processed fundamental climate data records. Should such records be generated during the ECV projects, it would be desirable to make them available to the user community as well.

It is worth noting that comprehensive multi-decadal reanalyses are substantial computational projects with demanding production schedules. Uptake of CCI ECV products would be increased if the ECV production timelines can be coordinated with such activities, and CMUG is in a position to keep the ECV projects informed of relevant reanalysis plans. ERA5 is the latest global reanalysis planned by ECMWF in early 2015 and the UERRA European regional reanalysis project is also planned to start production in 2015.

3.5 Data assimilation for seasonal and decadal forecasts

Recently the need for better initialisation of seasonal and decadal hindcast or forecast models in the operational forecasting centres has become apparent. The oceanic variables with sufficient inertia to act as forcing for seasonal time scales include sea surface temperature, salinity and sea-ice thickness and concentration. Proper initialisation of land surface temperature, soil moisture, snow cover and depth, and aerosol concentration can also increase prediction skill. Vegetation type is also of interest particularly if coupled with a vegetation model though a good high resolution dataset of recent vegetation distribution and its conditions (e.g. albedo, LAI) is valuable in its own right.

Interactions between the polar stratosphere and the mid-latitude troposphere occur on the timescale of a few weeks, and the initialisation of the former could aid the prediction of the latter especially in the first few weeks of seasonal forecasts (Scaife et al, 2005). Stratospheric temperature, winds and gas concentrations are therefore of interest to define in the model initial state. These parameters can now be measured by satellites to a reasonable degree of accuracy.

The experience of satellite data assimilation at NWP centres, which now provides the major impact on forecast skill, can be applied to these longer range model initialisation problems in particular from seasonal to decadal forecasts. The atmosphere is now represented by at least 70 levels from the surface to 0.1hPa with a horizontal grid size approaching 50km. Only satellite data can provide truly global coverage at this horizontal scale although radiosondes will still have better vertical resolution. In contrast for reanalyses the satellite climate data records are assimilated to affect the short range forecasts. In order for models to be able to assimilate a particular ECV it must be represented within the model as a prognostic variable. Table 1 shows those variables where data assimilation will be a possibility in the next 5 years.

3.6 Climate Services

The Copernicus Climate Change Service (C3S) is a European programme initiated in April 2014 as part of Copernicus, the environmental and emergency monitoring service for Europe. Its purpose is to provide accurate and independent information for climate security in Europe. C3S will be an operational service managed by ECMWF, together with other modelling centres and climate data providers subcontracted to them to provide operational services. It is anticipated that it will be fully operational by 2019. The intention is for C3S to use climate quality data produced mainly from satellite observations. For C3S there are about ten EC FP7 research projects (CLIPC, ERACLIM2, EUCLEIA, QA4ECV, UERRA, ECLISE, SPECS, EUPORIAS, CLIM-RUN) currently acting as precursors for components of the eventual operational C3S, and one of these is also a user of CCI data products (SPECS). CMUG engages directly with these projects for information and feedback about their data requirements. The interviews conducted for this requirements analysis included experts currently working on some of these C3S precursor projects as well as from a broader user community.

The integrity of CCI data is first fundamental requirement to its adoption by C3S, and a part of this is the 'line of sight' back to documented user requirements (such as this). Climate quality for data is a second requirement for users, which goes beyond provenance and scientific/technical specifications - the data must include user-friendly information about its usability and a rich description of the uncertainty. All these quality aspects should be captured in both maturity indices and metadata commentary.

The details for input of data, including ESA CCI data, to the C3S will be developed over the next few years. On the assumption that CCI data meets required standards it could be entrained directly in to the C3S (for combination with other data) before being made available to users as data products and services. It is almost certain that CCI data will continue to be an input to the climate research arena which is then used for further developing climate services. These possible data flows are shown in Figure 2, which also shows the direct application by users of CCI data. These parallel and serial chains of data processing and application create a multi-faceted set of user requirements.

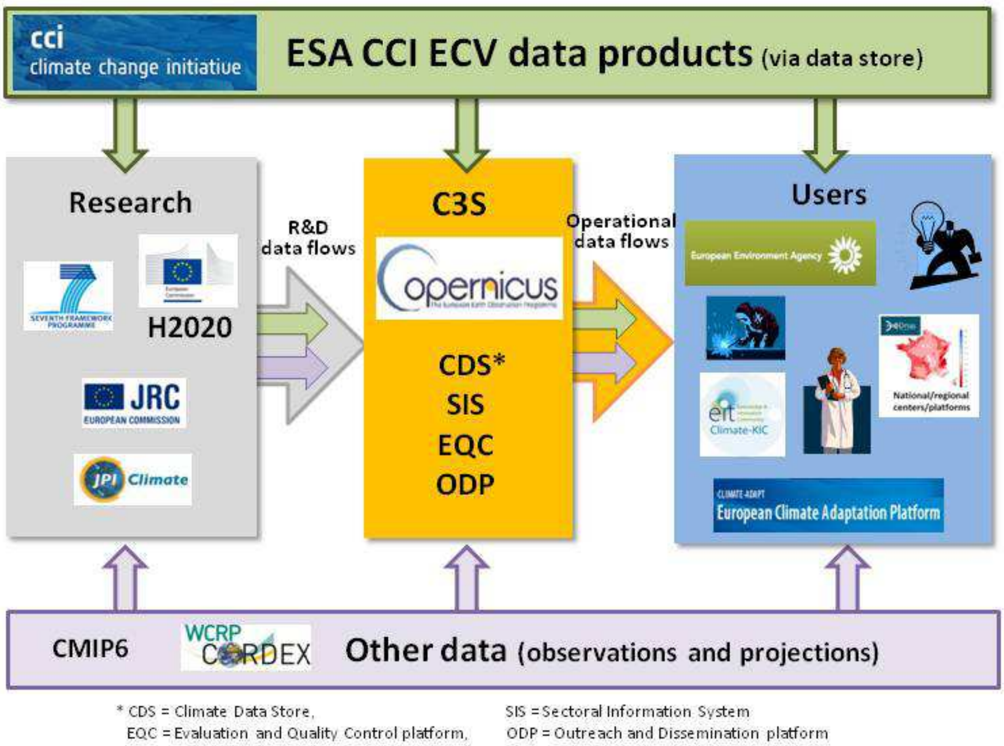

Figure 2: Shows the likely data flows of CCI data to the research community, C3S and direct to other users. It also illustrates the combination of CCI data with data from other sources in the provision of C3S.

Climate services also exist at a national level and have similar, almost identical, requirements to the C3S but on a finer scale. For example in the UK a scoping study started in 2014 for the Climate Change Risk Assessment which will define user requirements for the UK with respect to addressing climate change on different systems and sectors (e.g. agriculture, industry, health, infrastructure and the natural environment). The report will set out the main priorities for tackling the risk including environmental monitoring and climate services.

In some areas the data requirements for users of climate services are not fully clear and will evolve in future as these services develop. This can be explained with the following example - for applications such as monitoring of climate hazards and extreme climate events it is essential that the datasets produced can be used to a) calculate anomalies and b) are available within a short time period (two weeks) after acquisition so as to be useful. A concept similar to reanalysis data would be useful, where a consistent data processing is done for a long time period to generate a climatology and process more recent data with the same algorithms to enable anomaly calculations. In parallel, some final datasets with improved algorithms could be generated over shorter time spans (of up to a year).

3.7 Climate Studies (non-modelling)

There are a number of other research areas that use, or have the potential to use, CCI data. These are often specialised, with a core of experts at the centre of the community who often act as both data provider and user. Such communities are:

1. **World glacier inventory** (serviced by the Glacier CCI) which is a unique resource for glacier monitoring and research. Climate modellers are not using this very high resolution information yet except for regional studies.

2. **Environmental monitoring** - The aerosol, ozone, and GHG ECVs are input to the MACC/Copernicus Atmosphere Service to provide forecast products of atmospheric composition and air quality in addition to reanalyses of atmospheric composition.

3. **Many land use studies** – The Land Cover, Soil Moisture and Fire ECVs have the potential to be used by the GeoLand/Copernicus Land Service to support the provision of products for agriculture, forest, hydrology, etc to users.

4. **Quality control of in-situ data** - Satellite data can be used to validate in-situ measurements by using the large scale attributes of the satellite data if it can be assumed that any bias is stable over large spatial (>1000km) and temporal (>1hr) scales. The requirement is for the stability of the satellite CDR to be more stable than the in situ measurement errors being validated and so this depends on a case by case basis. If the in situ measurements are accurate and only have small drifts then the accuracy (stability and bias) requirements on the satellite data can be high. An example of this might be the use of AATSR brightness temperatures to validate drifting buoy sea surface temperature measurements. The latter can often be in error by several degrees and so an accuracy requirement on AATSR for this application need only be 0.5K to still show useful results. This is a much lower accuracy than the requirement for climate monitoring.

5. **As input data for adaptation research** – adaptation to climate change is an area of research which frequently combines data from the natural world and managed systems with socio-economic data to understand vulnerabilities to and the risks from climate change. CCI data can be of use to this group if the data meets there often high spatial and temporal requirements.

6. **As input data for other studies** – there are an increasing number of environmentally focused research areas which combine data on the natural world from different sources to better understand natural phenomena. Such an example is phenology where the timing of natural events is recorded and analysed. CCI datasets of SST, Soil Moisture, Sea Ice and Ocean Colour can support this user community.

4. Synthesis of requirements for CCI ECVs

The CMUG has undertaken a review of the requirements for the 13 CCI ECVs through direct interactions with expert users and responses to a questionnaire. This report builds on the user requirements made by CMUG in Phase 1[5] and presents an analysis of the user input together with the GCOS requirements
The requirements from the CMC and other expert users given here are in addition to those made by the CCI projects. An underlying assumption in this requirements definition process is that the CCI datasets produced will be *better* than any existing satellite CDRs, The complete datasheets containing the CMC and expert user requirements are given in Annex-B but a summary table for each ECV are listed in the sub-sections below. Note that it is difficult to be too prescriptive for accuracies as this depends on the horizontal scale chosen to represent the parameters so for example a SST at a 50km scale may be more accurate than at a 1km scale.

In addition to the consistent presentation of the requirements a consistent description of the errors also needs to be used. This is outlined in Annex A of this document. There are different requirements for errors for different applications. Table 2 gives those type of errors which are considered here.

Types of error
Single sensor uncertainty estimates for every observation (SSEOB)
Single sensor accuracy estimates for every observation (SSAOB)
Single sensor uncertainty estimates for TCDR (SSECDR)
Single sensor accuracy estimates for TCDR (SSACDR)
Error covariance matrix for TCDR (ERRCOV)
L3 merged product accuracy (ERRMERG)

Table 2. Types of errors for inclusion with TCDR datasets.

[5] CMUG Deliverable 1.2: User Requirement Document (v2.0), available at http://www.esa-cmug-cci.org.

The acronyms are used in the tables below.

4.1 Sea surface temperature

Sea surface temperature (SST) is an important variable to monitor over many timescales as a key indicator of climate change. Satellite SST data are crucial to obtaining globally complete SST analyses and in particular the high temporal and spatial resolution that is increasingly needed for understanding processes such as ENSO, NAO, PDO etc.

The IPCC AR5 report states *"Since the AR4, major improvements in availability of metadata and data completeness have been made, and a number of new global SST records have been produced. Intercomparisons of new SST data records obtained by different measurement methods, including satellite data, have resulted in better understanding of uncertainties and biases in the records.."* and so removal of the biases and understanding biases is clearly a critical need for climate monitoring. It is also important for climate change to monitor the SSTs over the Arctic Ocean which has become ice-free during the summer months as there is a lack of conventional air temperature measurements in the Arctic.

To be able to use an SST data set as a boundary condition for atmospheric reanalyses or in atmosphere-only climate simulations gridded data sets with complete coverage over the global ocean are needed. These are based on a special form of Optimal Interpolation that retains large-scale correlation structures and can accommodate sparse data coverage. The OSTIA SST analysis is used by the Met Office and other NWP centres for both operational forecasting (NWP and Ocean) and an OSTIA reanalysis has been run using the historical observations available. This complements the HadISST climate quality data analysis produced in the MOHC which makes use of the CCI SST climate data records. These high resolution analyses are linked to the longer term climate record of SST. The intention is to use the HadISST analysis for the next ECMWF reanalysis ERA5 which uses satellite data (AVHRR and ATSR) from 1979 onwards along with in-situ data.

The requirements for satellite SST are given in Table 3 for a number of applications related to climate modelling. An important consideration is whether sea surface skin temperature or sea surface subskin temperature (also known as a foundation temperature) is required (the latter requires an observation operator). The requirements are the same for both. For long term trend monitoring both parameters are of interest with foundation temperature used more in the past but for the satellite era skin temperature could also be used and models are being developed to use skin temperature or even radiances. Long term trend monitoring and attribution is the most challenging application with high demands on the accuracy and stability of the product especially if regional trends are required.

There are a number of requirements for initialising the initial state of seasonal, decadal and coupled climate model runs which all have similar requirements on accuracy. The deep ocean temperatures are more important for these longer range forecasts. For reanalysis the requirement is to provide a 3 hourly update to the SST field as a boundary condition for the assimilation of the atmospheric and other oceanic variables.

SST is also an important dataset for Climate Services with many applications (fisheries, military, tourism, transport, etc). Here a range of horizontal and temporal sampling options will be required for delivery to the diverse list of users. There is a requirement to have the reprocessed SST data within a month of real time to be able to put severe weather events into context for Government or media requests.

Application	Horizontal resolution	Temporal sampling	Precision	Accuracy	Stability	Error Type (see Table 3)	Source
Trend monitoring (global/regional)	10km/1km	1 week	0.1K	0.1K	0.01K/ decade	ERRMERG	GCOS JNT/SS
Seasonal f/c	30km	12h	0.1K	0.1K	0.05K/ decade	ERRCOV	JNT
Decadal f/c	30km	1 month	0.1K	0.1K	0.1K/ decade	ERRCOV	KW
Climate quality analysis	30km	12h	0.1K	0.1K	0.05K/ decade	ERRCOV	JNT DL KW
Global Reanalysis	30km	12h	0.1K	0.1K	0.01K/ decade	ERRCOV	JNT
Regional Reanalysis	5km	3h	0.1K	0.1K	0.01K/ decade	ERRCOV	RR
Climate services	1km	3h	0.1K	0.1K	0.01K/ decade	ERRMERG	

Table 3. Requirements for satellite SST observations. The accuracy and stability values assume global coverage for 100km spatial scales.

4.2 Ocean Colour

The impact of climate change on marine ecosystems and the ocean carbon cycle, from global to regional scales, can only be quantified by using long-term data sets, including satellite ocean colour. Synoptic fields of ocean colour (derived chlorophyll pigment), are used as an index for phytoplankton biomass, which is the single most important property of the marine ecosystem. Ocean colour is also the basis to infer primary production (CO_2 uptake by algae) and is currently the only source of observational data offering complete global coverage. This offers a wide scope of ocean colour CDRs applications, which include:

- initialisation and verification of coupled ocean-biogeochemical models and potentially ocean-atmosphere-biogeochemical models.
- data assimilation for state, as well as parameter estimation in ocean forecasting models.

The patterns of ocean phytoplankton concentration provided by the ocean colour data, combined with models, are an important source of information to physical-biogeochemical process studies, such as primary production, respiration and interactions at the air-sea interface.

Ocean colour is also an important dataset for Climate and Marine Services with numerous applications (e.g. fish stock assessments, carbon sequestration, ecosystem health monitoring and integrated ecosystem assessment to name a few).

Parameter	Application	Horizontal Resolution	Observing Cycle	Precision	Accuracy	Stability	Error Type (see Table 3)
Derived chlorophyll *a*	Trend monitoring Global/Regional	4km	1month	30% or under	30% or under	2%/decade	ERRMERG
	Decadal forecasting	25km	1 month	30% or under	30% or under	2%/decade	SSEOB/ERRMERG
	Seasonal forecasting	25 km	1 month	30% or under	30% or under	2%/decade	SSEOB/ERRMERG
	Global reanalyses	25 km	1 day	30% or under	30% or under	2%/decade	SSEOB/ERRMERG
	Regional reanalyses	7 km	1 day	30% or under	30% or under	2%/decade	SSEOB/ERRMERG
	Shelf (tidal) seas	4 km to 200 m	1 day	30% or under	30% or under	2%/decade	SSEOB/ERRMERG
	Assimilation	4km	1 day	30% or under	30% or under	N/A	SSEOB/ERRMERG

Table 4. Requirements for satellite ocean colour observations

The CMC requirements for satellite ocean colour observations are given in Table 4. Compared to the GCOS requirements these are close to the goals of GCOS in terms of resolution and observing cycle. The accuracy and precision requirements are well below the GCOS requirements not even approaching the threshold value of 25% (which calls in to question the GCOS value though for 100km grid scale it may be realistic) but modellers input stated that even 30% accuracies in derived chlorophyll alpha would provide some benefits. The requirements could also be sub-divided into CASE-1, CASE-2 and coastal waters where the first is the easiest case to achieve the stated requirements. There are a range of other possible products which could be considered for example in carbon budget assessments but modellers to date have not expressed any firm requirements for these.

4.3 Sea level

Sea level increase is one of the clearer indirect impacts of global warming and its potential effects justify a careful study of the sea level trends at the global and regional scales. It is also a key parameter to monitor some important features of climate variability such as the ENSO. Satellite observations with altimetry from the early 90's has demonstrated their great potential for monitoring sea level at scales extending from global to the mid-latitude ocean eddies. They have also provided an incentive to the development of ocean data assimilation schemes through the constraint they bring to ocean dynamics and thus to the initialization of seasonal, decadal and climate prediction models.

For the CMC a first interest is to run historical realisations of the climate and to compare the modelled regional variability of sea level with that observed. Getting models to match the observed variability should improve the value of their predictions. It is also important to ensure the overall sea level rise due to rising temperatures and melting of ice sheets that is modelled is consistent with the observations.

Another interest of the CMC for sea level data concerns data assimilation in ocean models. These data indeed provide invaluable information to complement in-situ observation in order to constrain the simulated ocean circulation. Ocean data assimilation can either be used by the CMC to initialize the ocean component of the coupled models used for climate prediction or through the use of ocean-reanalyses for a wide range of applications (for example to force atmospheric stand-alone models, to evaluate ocean models, and to analyse climate variability).

Sea-level trend analysis and detection/attribution studies also require sea-level ECVs with specific requirements in particular for stability.

The CMC requirements for satellite sea level observations are given in Table 5. They correspond to target requirements based on GCOS but updated with responses to CMUG 2010 and 2014 questionnaires and discussions with CLS scientists. A more stringent requirement on resolution in the most recent set of responses to the CMUG questionnaire (a factor 2) reflects the progress in model resolution as well in ocean data assimilation systems as in climate modelling. The request on precision, accuracy and stability remain the same between the two sets of enquiries.

Parameter	***Application***	***Horizontal Resolution***	***Observing Cycle***	***Precision***	***Accuracy***	***Stability***	***Types of error***
Ocean dynamic topography	*Model Development and Evaluation*	*25 km*	*5 days*	*1 cm*	*1 cm*	*2mm/ decade*	*SSEOB*
	Reanalyses and data assimilation	*25 km*	*2 days*	*1 cm*	*1 cm*	*2mm/ decade*	*SSEOB*
	Long Term Trend Monitoring and Attribution	*25 km*	*2 days*	*1 cm*	*1 cm*	*2mm/ decade*	*SSEOB*
Coastal sea level change	*Model Development and Evaluation*	*12 km*	*5 days*	*1 cm*	*1 cm*	*2mm/ decade*	*SSEOB*
	Reanalyses and data assimilation	*12 km*	*2 days*	*1 cm*	*1 cm*	*2mm/ decade*	*SSEOB*
	Long Term Trend Monitoring and Attribution	*12 km*	*2 days*	*1 cm*	*1 cm*	*2mm/ decade*	*SSEOB*

Table 5. Requirements for satellite sea level observations note all global datasets should go to the ice edge and not be limited to a latitude of 66S.

4.4 Sea-Ice

Changes in the polar sea ice cover are one of the most direct indicator of climate changes, since the ice cover integrates changes in both the atmospheric and oceanic forcing on time scales that are indicative of climatic changes. The strong decrease of the Arctic sea-ice cover and the concurrent slight increase of the Antarctic sea-ice cover have therefore sparked significant societal and scientific interest to better understand the ongoing changes. For this purpose, a reliable observational record of sea-ice properties is crucial that covers the entire polar regions. Such record is only available from satellites, which is why scientists that work on the large-scale evolution and predictability of the sea-ice cover all use satellite data as part of their daily work.

The most important sea-ice properties that can be obtained from satellites are the gridded sea-ice concentration as derived from passive-microwave retrievals from 1979 onwards, and the sea-ice thickness as can be derived from laser or radar altimetry and, for thin ice, from SMOS during the recent years. Additionally, but much less wide spread, albedo and ice drift products are used to understand the evolution of sea ice. From ice concentration, integrative quantities such as sea-ice area or sea-ice extent can be derived, while sea-ice volume can only be derived from the combined data set of sea-ice thickness and sea-ice concentration. The main specifications for data parameters under different applications are shown in Table 6.

A main issue for the usage of sea-ice data in a climate-research context relates to the poor description of uncertainties for both thickness and concentration retrievals, which are both important for model-evaluation and for model-initialisation purposes. Also the short length of all records hinders some of the scientific work regarding the long-term sea-ice evolution, as does the sometimes poor consistency between records based on different sensors.

Most scientists use level 3 data, with level 1 or level 2 data being primarily used for algorithm development. This focus on level 3 data might, however, change with the ongoing development of satellite simulators of sea ice that aim at directly providing level 2 fields from the model simulations.

The merging of several products into a single field is not strictly necessary, but might be useful for sea-ice thickness where Cryosat provides information on thicker ice and SMOS provides information on thin ice. In any case, it should always be possible to trace back the underlying data source at each grid point.

Data is ideally distributed as NetCDF, which is the most common format used by climate-modelling centres. The format does, however, not necessarily need to follow CMIP5 guidelines that are used for model output.

Parameter	Application	Horizontal Resolution	Observing Cycle	Precision	Accuracy	Stability	Types of error
Sea-ice cover (first year & multi-year	trend monitoring Global/Regional	12.5km / 12.5km	1 day	5%	5%	1%/decade	SSAOB
	decadal f/c	50km	1 month	5%	5%	1%/decade	SSEOB

ice)	Initialise	5km	1 day	5%	5%	1%/decade	SSEOB
	Reanalysis	12.5km	1 day	5%	5%	1%/decade	SSEOB
Sea-ice thickness	trend monitoring	20km	1 month	10cm or 10%	10cm or 10%	2 mm/decade	SSAOB
	decadal f/c	50km	1 month	10cm or 10%	10cm or 10%	2 mm/decade	SSEOB
	Initialise	20km	1 day	10cm or 10%	10cm or 10%	2 mm/decade	SSEOB
	Reanalysis	20km	1 day	10cm or 10%	10cm or 10%	2 mm/decade	SSEOB
Sea-ice drift	trend monitoring	12.5km	1-2-7 dy	0.01 m/s	0.01 m/s	0.01 m/s/decade	SSAOB
	Initialise	5km	1 day	0.01 m/s	0.01 m/s		SSEOB
	Reanalysis	12.5km	1 day	0.01 m/s	0.01 m/s		SSEOB
Melt pond fraction	trend monitoring	12.5km	1-2-7 dy	2%	5%	1%/decade	SSAOB
	Initialise	5km	1 day	2%	5%		SSEOB
	Reanalysis	12.5km	1 day	2%	5%		SSEOB

Table 6. Requirements for satellite observations of sea-ice

4.5 Clouds

The latest IPCC AR5 report state that clouds and aerosols continue to contribute the largest uncertainty to estimates and interpretations of the Earth's changing energy budget. Progress has been made in the understanding of how cloudiness and humidity changes simulated by climate model in warmer climates are related to large-scale circulation changes, such as the rising of high clouds and poleward shift of clouds associated with the stormtracks. However, some of the cloud changes vary substantially among models and are likely due to sub-grid scale processes, including the representation of convection and aerosol-cloud interactions in models. The uncertainty in the sign and magnitude of the cloud feedback is due to continuing uncertainty in the impact of warming on low clouds (Boucher et al. 2013).

The use of satellite data has increased since AR4, due to data records have become long enough and there is more available data from passive and active sensors as well as new types of technologies. The WCRP Grand Challenge on Clouds, Circulation and Climate Sensitivity is focused around five main initiatives, the fourth one, "Leveraging the past record", aim to exploit observations of the recent past, or proxies for longer-term changes, to better constrain cloud processes and feedbacks (http://www.wcrp-climate.org/grand-challenges/gc-clouds). The Cloud-CCI data-set planned to cover 30 years can contribute to this challenge by adding a new data set with consistent cloud variables and uncertainty information.

The GEWEX Cloud Assessment coordinated intercomparison of L3 cloud products of 12 global "state of the art" datasets have shown how cloud properties are perceived by different instruments and how cloud property averages and distributions are affected by instrument choice as well as some methodological decisions (Stubenrauch et al. 2013). In the assessment they found that differences in long-term variation in global-mean cloud amount between the datasets were comparable in magnitude to the interannual variability (2.5-3%). Still, these satellite cloud products are very valuable for climate studies or model evaluation, the geographical and seasonal variations in the cloud properties agree very well. They do not agree as well over deserts and snow-covered regions and for high level cloud statistics, due to problems detecting thin cirrus (Stubenrauch et al. 2013).

The AR5 report summarize the status of clouds observation as: "*In summary, surface-based observations show region- and height-specific variations and trends in cloudiness but there remains substantial ambiguity regarding global-scale cloud variations and trends, especially from satellite observations. Although trends of cloud cover are consistent between independent data sets in certain regions, substantial ambiguity and therefore low confidence remains in the observations of global-scale cloud variability and trends*." Therefore, for trend analysis, work remains to be done on quantifying the uncertainties in decadal trends of cloud parameters which should be accounted for by the Cloud-CCI team. For process studies there is also a strong requirement for satellite observations to improve the representation of clouds in climate models and here the long term stability is not an important requirement as the data are used to investigate changes on timescales of hours to seasons.

When comparing to climate models, observation time and view from above as well as retrieval filtering have to be taken into account. This can be achieved either by simple methods or by using observation simulators for the different datasets as in the Cloud Feedback Model Intercomparison Project (CFMIP) Observation Simulator Package (Bodas-Salcedo et al. 2011), which consists of individual simulators, with each corresponding to a specific cloud dataset (e.g., ISCCP, CALIPSO, MODIS, MISR, or CloudSat). Cloud-CCI efforts to develop a Cloud-CCI simulator, as well as testing more simple methods to be used by models without all fields available needed for the full simulator, follow the GEWEX Clouds Assessments recommendations.

The answers to the CMUG phase 2 cloud user survey, from five regional climate modellers and eight global climate modellers, are presented in the rest of this section. The main use of the Cloud-CCI datasets by the CMC phase 2 survey participants range from comparisons with models, for improved process understanding and parameterisations to detecting climate trends on regional and seasonal scales. The major obstacles in current use of satellite data are concerns about drifts and continuity between satellites and platform and lack of documentation. Answers and comments repeated in the survey by different users is that, error estimates must be provided with the data and it must be well described and documented.

Here, we first discuss the precise requirements on horizontal and vertical resolution, observing cycle and the type of usage as summarized in Table 7. Thereafter the more general requirements and comments from the survey are given. The five regional climate modellers only expressed special requirements for the horizontal resolution as mentioned below and marked in Table 7.

Horizontal resolution

Current global climate models are run typically at 100km but model development and faster computers will allow horizontal resolutions of 50 km to 25 km over the next 5 years or so.

Regional climate and NWP models have resolutions from 50km down to km scale. For detailed process studies it is desirable to have information at sub-grid scales, hence the specification of 10 km and for high resolution models km scale. For more general evaluation studies, e.g. comparison of monthly mean geographical distributions, this could be relaxed considerably and horizontal resolutions of around 50km-100 km can still be useful.

Vertical resolution

The distribution of the vertical levels in atmospheric models is highly non-linear with respect to altitude – the layers are typically much more tightly spaced in the boundary compared to the free troposphere, for example. Current global climate models have vertical resolutions of around 200 m in the boundary layer (with even this not being entirely satisfactory to represent stratocumulus cloud), increasing to around 500 m in the middle troposphere – the specification of 100 m is thus again based on the requirement for process studies. This could also be relaxed for other evaluation work and a vertical resolution of 500 m (or more) might be useful, depending on the information content of the particular observations. Vertical resolved clouds from CloudSat and CALIPSO have been used extensively over the last couple of years by the CMC. For the passive sensors used for the Cloud-CCI products there is no vertical information, except cloud top height and cloud top pressure. For validation purposes it would be useful to have these products compared to the CloudSat/Calipso data sets.

Observing cycle

In common with many related processes (e.g. rainfall, convection) the diurnal cycle of cloud remains a common weakness in the majority of current models. Examples of cloud systems with large diurnal cycles are tropical convection over land and marine stratocumulus cloud. Ideally, data with a temporal resolution comparable to the typical model time step (15-30 minutes) would be desirable. Again, however, much useful information could be obtained with 1-hourly data, with the upper limit on utility probably being 3 hours.

Model development/ evaluation

There are various products of interest which range from fields of cloud cover and top pressure/temperature to profiles of water and ice cloud concentration. The CMUG initial proposal to the CCI clouds project was to produce histograms of cloud parameters, which has been fulfilled, additional histograms relating the cloud parameters to other parameters, for example aerosols would also be of interest. The Cloud CCI team's plans to incorporate their data sets into the COSP simulator is largely supported by CMUG. The utility of statistical summaries (e.g. optical depth vs cloud top pressure histograms) when employing the COSP simulator, can be compared to climate model output in a very straightforward manner. This has been recognised by the observational community and ISCCP-like histograms are now produced using both MODIS and MISR data. This approach has several advantages:

- It puts the CCI data into a format that is already familiar to modellers.
- It allows the CCI data to be easily compared to other cloud data sets.
- It allows the CCI data to be easily integrated into pre-existing and tested methods for exploiting satellite cloud data for model evaluation.

Parameter	Application	Horizontal Resolution	Vertical resolution	Observing Cycle	Precision	Accuracy	Stability	Types of error
Cloud cover	model development	50km-1km	N/A	Monthly to 1h	10%	5%	1%/year	SSEOB
	trend monitoring	50km	N/A	Monthly to 3h	10%	5%	1%/decade	SSAOB
	Reanalysis/ Processes	10km-2km	N/A	6h to1h	10%			
Cloud top height	model development	10km	N/A	Monthly to1h	0.1km	0.1km	0.1km/ year	SSEOB
	data assimilation	5km	N/A	1h	0.1km	0.1km	N/A	ERRCOV
	trend monitoring	30km	N/A	Monthly to 3h	0.2km	0.2km	0.1km/ decade	SSAOB
	Reanalysis/ Processes	10km-2km	N/A	6h to 1h	10%			
Cloud top temp	model development	10km	N/A	1h		0.25K		SSEOB
	trend monitoring	30km	N/A	3h	0.25K	0.25K	0.25K/decade	SSAOB
Cloud ice profile	model development	50km-1km	0.2km	1h				SSEOB
Cloud water profile (> 100 µm)	model development	50km-1km	0.2km	1h				SSEOB
Cloud water profile (< 100 µm)	model development	50km-1km	0.2km	1h				SSEOB
Cloud effective radius?	model development	50km-1km	0.2km	1h	1um	1um	1um	SSEOB

Table 7. Requirements for satellite cloud observation. The underlined values for the horizontal resolution are requirements from regional climate modellers.

Trend monitoring

The requirements for trend detection are somewhat more difficult to ascertain. Firstly, there is currently no clear indication from presently-available observations about cloud trends and secondly this may well be too stringent a test for current models, given the known uncertainties in the representation of cloud processes. It certainly the case that the cloud modelling and cloud feedback community is currently much more focused on process studies than on long-term trends. That said, a new data set that was able to determine trends in cloud amount, for example, with the specified level of accuracy/stability would be a major advance and would undoubtedly be of great interest to climate modellers.

The GCOS requirements for the cloud ECV are somewhat relaxed in terms of observing cycle (3-6hr) compared to the CMC requirements which may reflect the needs in terms of long term trend monitoring rather than model process studies. Also the GCOS accuracies for cloud cover and cloud top height are more relaxed than those required for model processes.

Single sensor vs merged sensor products
Another consideration is that the generation of merged products from quite different sensors will be difficult to interpret for most applications. Such merged products are difficult to use: indeed, the rationale behind the simulator approach is precisely to avoid such difficulties by generating model equivalents of single-sensor products. However, the CMC answering the phase 2 survey are interested in both single sensor and merged products. To ensure traceability for merged products it is important with pixel (grid-point) errors and good documentation of the processing for merged products.

Satellite data validation/evaluation, format and access
For validation, the CMC users recommend it should be done for all seasons, for day and night and on regional scale against reference data with known errors, e.g. against station data and in-situ measurements and CloudSat/CALIPSO. Scores could be combined addressing bias, spatial and temporal correlations. The preferred format of the data is netCDF. Many modellers say it would be very useful to follow the CMIP5 format, and some even say that it is a prerequisite for extensive use of the data within the climate community. The preferred means of access to the data is via ftp or via a Web browser. Some strongly recommend that the data is available from a centralised server as ESGF and or that the data is available through Obs4MIP. Already users of ESA-CCI data, ask for a technical note similar to the one distributed for phase 1 to be provided also for phase 2 data. CMUG will contribute to this request, updating the phase 1 table and add information on access of the different ECV datasets before the end of phase 2.

Finally, to summarize, the general view on cloud satellite data from AR5, recent papers and the participants in the user survey lead CMUG to recommend that the cloud ECV datasets are continued to be designed for validating cloud model processes as well as building a long term monitoring datasets, despite difficulties. A simulator should accompany the data and the data should have been validated and include uncertainties and be well documented.

4.6 Ozone

The ozone concentration in the atmosphere (mainly the total ozone column) has been measured for several decades after the discovery of the impact of human activities on the upper stratosphere and lower stratosphere chemical processes, resulting in the high latitude ozone holes. Monitoring the trends of ozone content remains a key issue for the study of the recovery of stratospheric ozone and also for monitoring human induced greenhouse gases as far as tropospheric ozone is concerned. It is also essential to study stratospheric-tropospheric exchange processes and to give a better representation of the dynamics, chemical, transport and radiative processes. Ozone data assimilation is of primary importance for environmental studies including the initialization of air quality prediction (interactions between air quality and climate are deemed increasingly important). Some studies have also revealed the potential of ozone observations in constraining the atmospheric dynamics through data assimilation. Considering available observations, those from satellites are crucial in providing information on the ozone content of the atmospheric column but also, through the development of new sensors, to provide valuable information on partial columns and also the ozone profile.

The requirements for the ozone ECV are given in Table 8. The CMUG initial proposal to the CCI project for ozone is to have different specifications according to the altitude range. CMUG would prefer the same threshold specifications on vertical resolution for higher troposphere and lower stratosphere (2km) for a better validation of chemical-transport models.

Parameter	Application	Horizontal Resolution (km)	Vertical Resolution (km)	Observing Cycle (h)	Precision (%)	Accuracy (%)	Stability (%)	Types of error
Ozone profile								
Higher stratosphere & mesosphere (HS & M)	Model Development and Evaluation	100	3	24	5	10%	2.0 %/decade	SSEOB
	Reanalysis and Data Assimilation	50	1	6	5	10%	2.0 %/decade	SSEOB & SSAOB
Lower stratosphere (LS)	Model Development and Evaluation	50	2	24	3	6%	2.0 %/decade	SSEOB
	Reanalysis and Data Assimilation	20	1	6	3	6%	2.0 %/decade	SSEOB & SSAOB
Higher troposphere (HT)	Model Development and Evaluation	20	2	24	3	8%	2.0 %/decade	SSEOB
	Reanalysis and Data Assimilation	20	1	6	3	6%	2.0 %/decade	SSEOB & SSAOB
Lower troposphere (LT)	Model Development & Evaluation	20	2	24	6	10%	2.0 %/decade	SSEOB
	Reanalysis and Data Assimilation	20	1	4	5	10%	2.0 %/decade	SSEOB & SSAOB
Ozone column								
Troposphere column	Model Development and Evaluation	20		24	6	15	2.0 %/decade	SSEOB
	Reanalysis and Data Assimilation	20		4	5	10	2.0 %/decade	SSEOB & SSAOB
Total column	Model Development and Evaluation	20		24	2	4	1.0 %/decade	SSEOB
	Reanalysis and Data Assimilation	20		6	3	5	1.0 %/decade	SSEOB & SSAOB

Table 8. Requirements for satellite observation of ozone.

For lower troposphere and tropospheric column, CMUG prefer more stringent requirements on the observing cycle to better constrain O_3 pollution episodes and the daily cycle.

As far as ozone assimilation is concerned, in particular within the context of successors to the MACC re-analysis, products from single sensors would be preferred to merged products. Merged products if they are all obtained with the same technique and over a long period span (like the SBUV sensors over 30 years) are useful in a model validation context like CCMVal, aiming at evaluating each process separately. This implies to provide these different products as separate datasets.

Compared with the previous CMUG URD on the error/uncertainty requirements, it seems that these have become generally more stringent for the precision while the requirements in terms of accuracy have been slightly relaxed. This could be a consequence of the fact that many models nowadays include schemes to correct the observations for systematic biases, of which the accuracy is an estimate.

User friendly quality information and traceability have been identified as one of the major obstacles in current satellite data usage. While good documentation, especially on the quality assessment, and history of changes (with appropriate data versioning) are also regarded as important aspects to efficiently use the data.

A homogenous and coherent definition of the tropopause (possibly also included in the dataset) was suggested as being very important and useful for some applications.

4.7 Greenhouse Gases

A comprehensive understanding of greenhouse gases is crucial for informing societal response to climate change. Applications with a need for observations of greenhouse gases such as CO_2 and CH_4 include Model Development, Decadal Forecasting and Regional Source/Sink Determination. As shown in Table 10, each application has somewhat different observational requirements reflecting the particular aspect of greenhouse gases under consideration.

To elaborate on the GHG observational requirements for Regional Source/Sink Determination, the tabulated values are based on the activities undertaken within the frame of the MACC sub-project on greenhouse gases and on feedbacks from the GHG CMC. The principal products from the MACC sub-project on GHG are:

- 4-dimensional gridded fields of CO_2 and CH_4 produced in near-real-time (based on data assimilation of near-real-time data products, typically from operational satellites),
- 4-dimensional gridded fields of CO_2 and CH_4 produced in “delayed mode” (6 months delay, to allow data assimilation of research-mode satellite data products),
- 3-dimensional gridded fluxes of CO_2 and CH_4 produced in “delayed mode”,

- Re-analysed concentration and flux fields of CO_2 and CH_4 for the period 2003-2010.

Flux fields are an important factor for decision-makers at several levels, and need to be estimated with confidence. The fidelity of flux estimates is strongly influenced by accuracy and stability of the observations that are used as input to the data assimilation and re-analysis systems. This drives the requirements given in Table 9 for some of the required parameters. The requirements for full GHG concentration profiles are given in Annex B. In general, differences were found in the user requirements even when the same application was considered. An important element to consider in this regard is the actual target each user focuses on (e.g. cities rather than countries).

Horizontal Resolution and Observing Cycle requirements have become more stringent than what suggested by GCOS and by the CMUG Phase 1 URD. This is because if on one hand they reflect the spatial and temporal variability of important classes of regional sources and sinks on the other they also reflect improvements in the models, especially in terms of increased horizontal resolution. The need for good flux estimates makes the current requirements for accuracy and stability generally more demanding than previous GCOS and the CMUG Phase 1 user requirements.

Parameter	Application	Horizontal Resolution	Vertical Resolution	Observing Cycle	Precision	Accuracy	Stability	Types of error
Trace gas profile CH_4 - Troposphere column	Regional source/sink determination	5/20/50 km	N/A	3/4/6 h	0.1/0.5/1% 2/10/20 ppb	0.1/0.5/2.0% 2/10/20 ppb	0.5/0.7/2.0 %/dec 2/7/35 ppb/dec	SSEOB
Trace gas profile CH_4 - Total column	model development	25km	N/A	6 h	1%	1%	10ppb/dec	SSEOB
	decadal f/c	20km	N/A	Daily	<<10 ppb	<<10 ppb	2%/dec 35 ppb/dec	SSAOB
	Regional source/sink determination	10/50/100 km	N/A	3/4/6 h	0.25/0.5/1% 5/10/20 ppb	0.1/0.5/2.0% 2/10/40 ppb	0.1/0.5/2.0 %/dec 2/10/35 ppb/dec	SSEOB
Trace gas profile CO_2 - Total column	model development	25km		6h	0.5/1ppm	0.5/1ppm	0.1/0.5ppm/dec	SSEOB
	decadal f/c	2/5/20km	N/A	Daily	0.3/0.5/1% 1/1.5/3 ppm	0.3/0.5/1% 1/1.5/3.0 ppm	0.5/1.5/2 %/dec 2/5/8 ppm/dec	SSAOB
	Regional source/sink determination	5/20/50 km	N/A	3/6/24 h	0.25/0.5/0.75% 1/2/3 ppm	0.25/0.5/1% 1/2/4.0 ppm	0.5/1.5/2 %/dec 2/5/8 ppm/dec	SSEOB
Trace gas profile CO_2 - Troposphere column	Regional source/sink determination	5/20/50 km	N/A	3/4/6 h	0.15/0.4/0.5% 0.5/1.5/2 ppm	0.15/0.5/1% 0.5/1.5/4.0 ppm	0.15/0.5/2 %/dec 0.5/1.5/7.5 ppm/dec	SSEOB

Table 9. Requirements for satellite observation of greenhouse gases

The requirements are given for tropospheric and total column only, in recognition that requirements for profile data would be very demanding for existing satellite data. In the event that data providers consider it feasible to provide profile data approaching GCOS requirements, then more refined user requirements could be given in a future update of this document. The user community increasingly asks for horizontal and vertical resolution in the Lower Stratosphere to be the same as that for the Higher Troposphere, in contrast to previous GCOS requirements. As mentioned above, other applications of greenhouse gas observations may have different sets of requirements. For example, the detection of CH_4 emissions from pipelines or similar small sources would require higher horizontal resolution and vertical resolution in the lower troposphere.

Turning now to the GHG observation requirements for decadal forecasting, it is principally the distribution of the trace gases at the start of the forecast that can be important to help define the atmospheric fields. This consideration was translated in the Phase 1 URD in a requirement of long period averages as sufficient for decadal forecasts. The latest consultation seems to indicate that a much higher observing cycle would be useful. Additionally, more stringent requirements have been made for the horizontal resolution that is now comparable with that needed in other applications.

Similar to the ozone section above, it would be important to provide not only merged GHG products but also products from single sensors as separate datasets. Users also pointed out that the harmonisation between the various datasets is a key aspect to efficiently using the data.

4.8 Aerosols

The impact of aerosols on climate is often cited as one of the most uncertain factors governing climate change. Aerosols have offset part of the warming expected from anthropogenic emissions of greenhouse gases. It is very important to decrease the uncertainties on the aerosol forcing because this will contribute to better constrain the climate sensitivity from current observational climate records. As a result measurements of atmospheric aerosols (both tropospheric and stratospheric) are required. There is a further arbitrary split at 3km height to obtain aerosol products below and above the lower troposphere.

Aside from the direct radiative effect it is in particular the impact of indirect radiative effects (mainly through clouds) which needs to be better understood to better estimate the climate sensitivity to aerosols in climate models. Thus, there are two aspects that need to be addressed. Relatively high resolution data with associated environmental data (e.g. clouds) for a better process understanding, as well as long-term monitoring on global scales to address trends in aerosol properties. Precipitation has also been reported by the users as the single most important climate change impact parameter.

The current aerosol climatologies within global models are usually extremely basic and essentially consist of time-invariant two-dimensional fields an aerosol amount. Thus datasets of aerosol properties considering both spatial and temporal as well as compositional variations will be a step forward. The parameters required are listed in Table 10. It includes the aerosol extinction optical depth (AOD) (at the modelling reference wavelength at 550nm) for both the total atmospheric column as well as stratified over four atmospheric altitude sections to distinguish between stratosphere (important after major volcanic eruptions) and tropospheric layers linked to high-, mid- and low level clouds. Upper tropospheric aerosol have enhanced capabilities for long range transport, while lower tropospheric aerosols remain more local and influence the near surface meteorology (e.g. visibility, air quality). In general, tropospheric AOD can be derived as the total AOD minus the stratospheric AOD. In addition to total extinction optical depth (absorption + scattering) the absorption optical depth is also an important parameter to measure and has more stringent accuracy requirements being only part of the total extinction.

Aside from aerosol amount also the aerosol composition is of interest. A very useful property in that sense is data for AOD at different wavelengths. These different AOD data provide information on aerosol size. AODs at two different wavelengths already define the Angstrom parameter, which a more general size-indicator. Even better is the AOD fine mode fraction, which requires AOD data at least four different wavelengths in the visible and the near-IR. Then via the Angstrom parameter spectral dependence the total AODs can be stratified into fractions associated with smaller (radii <0.5um) and larger sizes (radii >0.5um). Thus, aside from the AOD retrieval at 0.55um, additional AOD retrievals at one or even better at three other wavelengths in the visible or near-IR are desirable (e.g. 443nm, 670nm, 870nm). Other useful elements to characterize aerosol type are data on polarization and absorption. Polarization provides information on aerosol shape (e.g. mainly to discriminate dust from other aerosol type). In most retrievals a-priori assumptions on aerosol absorption are made.

One CMC requirement is defined by the assessment of aerosol processes in climate models which requires data on associated environmental properties. Thus such process understanding of processes involves especially the potential interactions with clouds. Thus, data on clouds (from the cloud ECV) are required which match in terms of spatial and temporal) resolution, observing period and if possible satellite platform. The other CMC requirement is the establishment of long time-series for aerosol properties. In that sense, it is also important that the platform/instrument lifetime is at least 10-15 years, in order to detect possible trends.

The CMC also stressed the importance of vertical profiles of aerosol extinction. These would be useful to answer questions such as the injection altitude of aerosol and the stratospheric transport of tropospheric aerosol.

The GCOS requirements for aerosol optical depth match those of the CMUG in terms of horizontal resolution but the observing cycle of 6hr for monitoring and 1hr for process studies is more frequent than the GCOS goal of 1 day.

Depending on the specific satellite products and periods eventually chosen for re-processing by the CCI-aerosol project, further suggestions for improvements to data quality may be provided based on existing experience within the data assimilation community of existing and related data products.

Concerning single sensor datasets or merged product datasets or both, the CMC stressed to that both are required. Merged (single sensor) products are preferred for monthly mean (instantaneous) data. Traceability back to the sensor and documentation are important issues.

Concerning the preferred validation methodology, the CMC stressed cross-validation against in-situ data, ground-based measurement (e.g. AERONET), other instruments (e.g., CALIPSO, CLOUDSAT) and reanalysis data.

Parameter	Application	Horizontal Resolution	Observing cycle	Precision	Accuracy	Stability	Types of error
Total extinction optical depth (at 4 VIS + IR wavelengths)	model development	1km	1hr	0.02	0.02	0.02/decade	SSEOB
	assimilation	2km	1hr	0.02	0.02	0.02/decade	SSEOB
	decadal f/c	2km	Daily	0.01	0.02	0.005/decade	SSEOB
	trend monitoring	2km	3hr	0.005/ 0.01	0.01/ 0.02	0.02/decade	SSAOB
Total aerosol absorption optical depth at 0.55um	model development	1km	1hr	0.004	<0.01	0.005/decade	SSEOB
	trend monitoring	2km	3hr	0.002/ 0.01	0.004/ 0.02	0.002/ decade	SSAOB
Aerosol optical depth in stratosphere (at 4 VIS + IR wavelengths)	model development	1km	1hr	0.02	0.02	0.02/decade	SSEOB
	trend monitoring	2km	6hr	0.02	0.02	0.01/decade	SSAOB
Aerosol optical depth in troposphere (at 4 VIS + IR wavelengths)	model development	1km	1hr	0.004	0.02	0.02/decade	SSEOB
	trend monitoring	2km	6hr	0.002	0.004	0.01/decade	SSAOB
Aerosol optical depth above ~3km (680hPa) (at 4 VIS + IR wavelengths)	model development	1km	1hr	0.01	0.02	0.02/decade	SSEOB
	trend monitoring	2km	6hr	0.005	0.01	0.01/decade	SSAOB
Aerosol optical depth below ~3km (680hPa) (at 4 VIS + IR wavelengths)	model development	1km	1hr	0.01	0.02	0.02/decade	SSEOB
	trend monitoring	2km	6hr	0.005	0.001	0.01/decade	SSAOB
Aerosol depolarisaton ratio (VIS)	model development	1km	1hr	N/A	10%	N/A	SSEOB
	trend monitoring	2km	6hr	N/A	5%	N/A	SSAOB

Table 10. Requirements for satellite aerosol datasets

4.9 Glaciers and Ice caps

Glaciers and ice caps provide a visible indication of the effects of climate change, as the mass balance at the surface of a glacier (the gain or loss of snow and ice over a hydrological cycle) is determined by the climate. It is important to measure and understand the areal and volumetric changes with time, and also how well climate models can represent or parameterise glaciers and icecaps.

According to the tiered strategy of global glacier monitoring in the Global Terrestrial Network for Glaciers (GTN-G), the basic application of satellite data is the generation of repeat glacier inventories at decadal time scales using cost-efficient semi-automatic classification techniques and data processing in Geographic Information Systems (e.g. Paul et al. 2007). This is in line with Product T.2.1 from GCOS (2006) that ultimately requests to obtain a globally complete map of glaciers and icecaps. The global map of glaciers and icecaps would serves several fields of application, including:

- improved modelling of global sea-level rise (e.g. Hock et al., 2009; Hirabayashi et al., 2010),
- a sound basis for change assessment (e.g. Bolch et al., 2010),
- an important input for hydrological (e.g. Viviroli et al., 2009) and glaciological modeling (e.g. Oerlemans et al., 1998).
- a possibility to validate output from RCMs (e.g. Ghan et al., 2006), and
- a data set to initialise the land ice fields in RCMs (Kotlarski et al., 2010).

Apart from the glacier extent, satellite data are used widely to derive further glaciological parameters including snow facies, velocity fields and elevation changes (e.g. Paul et al., 2009). All these products do strongly vary in terms of sensors (resolution), observing period and cycle, or required precision and accuracy. The related list of satellite based observational requirements and capabilities was compiled by IGOS (2007). We have used this list (table B.6) as a base for Table 11 below. The long term stability of the measurements is crucial for this ECV as it is an indicator of climate change.

Parameter	Application	Horizontal Resolution	Observing Cycle	Precision	Accuracy	Stability	Types of error
Glacier Area	Initialisation	30 m	1 year	0.01km^2	<5%		SSEOB
	trend monitoring	30 m	5 years	0.01km^2	<5%	0.01km^2/ decade	SSAOB
Glacier Topography	Initialisation	<100 m	1 year	1 m	5 m		SSEOB
	trend monitoring	<100 m	5-10 years	1 m	5 m	1 m/ decade	SSAOB
Velocity	Initialisation	30 m	1-12 months	1 m/yr	10 m/yr		SSEOB
	trend monitoring	30 m	1 year	1 m/yr	10 m/yr	1 m/ decade	SSAOB

Snowline	Initialisation	30 m	1 year	30 m	100 m		SSEOB
	trend monitoring	30 m	1 week / 1 year	30 m	100 m	30 m / decade	SSAOB

Table 11. Requirements for glacier and ice caps

The two main requirements for the glacier and ice cap datasets for the CMC are trend monitoring and providing initial conditions for climate models. The datasets can also be used for validation of land surface process in climate model predictions which have the same requirements for accuracy as the trend monitoring.

4.10 Land Cover

Land cover describes the distribution of vegetation and man-made features (living space, agriculture and forestry). In the context of the ECV CCI LC Phase I, detailed LC typology is sub-divided in patches of different plant functional types and groups of classes thematically closed. This has been done exploiting the CCI Land Cover products described below:

- a 7-day surface reflectance time series of the MERIS Full & Reduced resolution for the whole archive (2003-2012)
- three 300m global land cover maps (2000, 2005, 2010) including 22 classes for the 1998-2002, 2003-2007 and 2008-2012 epochs derived from a multi-sensor and multi-year strategy
- three global land cover seasonality products about vegetation greenness, snow and burned areas on a 7-day basis, for 1999-2011, 2000-2012, 2005-2010 epochs, respectively
- a 300m global map of open permanent water bodies, derived from the full ASAR dataset between 2005 and 2010

Detailed information about global land cover is an important variable for global and regional climate modelling over many timescales.
Earth system models are the most advanced tools to conduct studies on climate monitoring/attribution since mid-20th, and also to predict future climate. Land cover information is used in climate models for the initialization as well as to prescribe boundary conditions. However it has been stated that in case of not taking into account historical land cover changes, it is impossible to reproduce atmospheric CO_2 concentration growth, as well as the carbon budgets for the present days. Nonetheless, Land Cover information is widely used to help model development and validation.
In the context of Phase 2 of this project the analysis of the requirements expressed by the various experts highlights a set of greater constraints for the defined criteria. More specifically the minimal resolution is expected to decrease by a factor of 10 (from 300m to tens of meters). Similarly, the observing cycle for the LC ECV is required to be shorter than the 2 to

5 yrs previously specified. This is most particularly relevant for the land cover change detection as those changes occur in timeframe shorter than 2 years.
It should also be noted that the ESA CCI Phase I – Land Cover ECV project conducted a survey across the community to gather expectations and requirements regarding land cover and land cover change. A subset of the key requirements can be listed as:

- There is a need for both stable land cover data and a dynamic component in the form of time-series and changes in land cover;
- Consistency among the different model parameters is often more important than accuracy of individual datasets, and it is important to understand the relationship between land cover classifiers with the parameters and the relative importance of different land cover classes;
- Providing information on natural versus anthropogenic vegetation (disturbed fraction), tracking human activities and defining history of disturbance is of increasing relevance; in particular for land use affecting land cover with most details needed to focus areas with large anthropogenic effects;
- Land cover products should provide flexibility to serve different scales and purposes both in terms of spatial and temporal resolution.

The land cover information is translated into surface parameters (e.g. albedo, LAI, fractional vegetation cover), which provide the lower boundary condition for the atmospheric models. On the other hand, detailed regional land cover information provides a very valuable information for process studies like e.g. the assessment of the impact of fires.

Data should be provided under the netCDF–CF format, which should be made accessible via FTP.

The requirements for land cover are given in Table 12.

Parameter	Application	Horizontal Resolution	Observing Cycle	Precision	Accuracy	Stability	Types of error
Land cover type	model development	10s m - 50km (Global) 300m (Regional)	Monthly Yearly 5 yr	should enable for legend refinement, discrimination within a group, and for transitional zones refinement	5%	Consistency should be maintained across several consecutive maps	ERRMERG
Land cover change	trend monitoring	10s m – 1km-50km (Global) 300m (Regional)	Monthly Yearly 5 yr	should be sufficient to detect meaningful changes for changes (deforestation, desertification,reforestation, greening, and drying)	5%	Consistency should be maintained across several consecutive maps	ERRMERG

Table 12. Requirement for satellite land cover parameters

4.11 Fire

Fire disturbances alter vegetation dynamics and impact climate. Climate models that account dynamically for climate induced changes in vegetation simulate fire disturbance within process based fire sub-models. The development and evaluation of such sub-models depend on the availability and quality of satellites based fire disturbance products. Such complex Earth System models are crucial to assess fire climate interactions and the impact of fire on the global carbon cycle.

In addition, global vegetation models can be utilized to diagnostically simulate fire emissions by combining information on burned area, available fuel load and burning conditions. Satellite based burned area products can thereby serve as prescribed boundary conditions. Besides uncertainties in burned area estimates, such an approach is limited by an uncertain quantification of available fuel loads and burning conditions (e.g. combustion completeness, mortality rates, emission factors). Fire disturbance products will therefore be best exploited in models when consistently derived ancillary data products, such as land cover classification or biomass availability, are provided that help to constrain specific burning conditions.

The assessment of fire emissions will be one important application of fire disturbance products. Fire emissions serve as boundary conditions for atmospheric aerosol and chemistry models used to assess air quality. An operational usage of atmospheric composition models will require near real-time availability of the fire disturbance ECV. Other application of the fire CCI product include improvement of fire model parameterisations as well as process studies.

The strong interannual variability of fire activity vegetation models will require data products that cover a multiyear timespan (10-20 years) for the development and evaluation of process based fire models as well as for the application of satellite observed burned area products as boundary condition.

The specific requirements for the fire disturbance ECV are listed in Table 13. In terms of spatial resolution and observing cycle these are close to the GCOS requirements.

Parameter	Application	Horizontal Resolution	Observing Cycle	Accuracy	Stability
Burned fire area	trend monitoring	0.25/1.0/5.0 km	1/1.5/3 d	30/20/10 %(MAX)	5.00%
	Prescribe model boundary condition	0.25/1.0/5.0 km	3h/ 1/1.5/3 d	30/20/10/1 %(MAX)	5.00%

Table 13. Requirements for satellite burned area fire parameters

Users do apply level 1 (direct data assimilation), level 2 (assimilation), as well as level 3 data (verification and climate monitoring). Major problems with current products include non user friendly traceability of the end-product, lacking documentation and immature uncertainty characteristics. The users want single datasets as well as merged datasets. For the merged product traceability to the single sensors as well as a good documentation is a key requirement. The validation should be based on site level data and inter instrument comparison. The preferred data format is netCDF–CF following the CMIP5 format guidelines, which should be made accessible via FTP. None of the expert users interviewed focused particularly on a regional application of the Fire CCI product. For climate applications, however, the product requirements for regional applications should be comparable with the one for global applications.

No fire radiative power product is planned by the fire CCI project. This is not strictly an ECV although it is a requirement of climate modellers. This issue will need to be raised at least with the CCI project and to make this more explicit within the GCOS parameter list.

4.12 Ice Sheets

Climate modellers are interested in ice sheets because of their interactions with other components of the climate system (e.g. freshwater fluxes from ice sheets to modify sea-level or orographic forcing of wind patterns). However, only a few climate modellers responded to Ice Sheet ECV questions in the survey because satellite data of ice sheets is not commonly exploited in models. Due to the relatively low interest from the CMC the information in summary Table 14 is derived partly from the User Requirements Document of the Ice Sheets CCI project[6].

Parameter	Application	Horizontal Resolution	Observing Cycle	Accuracy	Stability	Types of error
Surface elevation change	Initialisation	<5km	annual	0.1m/yr		SSEOB
	trend monitoring	<500 m	monthly	<0.1m/yr	<0.1 m/ decade	SSAOB
Ice Velocity	Initialisation	0.5m/yr	annual	30 m/yr		SSEOB
	trend monitoring	0.1m/yr	monthly	<10m/yr	<10m/yr	SSAOB

Table 14. Requirements for ice sheets for modelling applications

One important requirement for climate modellers is for data on both the Greenland and Antarctic ice sheets.

4.13 Soil Moisture

[6] Available at: http://www.esa-icesheets-cci.org/?q=documents [User Requirements Document v1.5].

Soil moisture is an important variable for all models from NWP to climate time scales. For reference the GCOS requirements are given in Table 15 below along with those assumed in NWP data assimilation systems.

Parameter	Application	Horizontal Resolution	Observing Cycle	Accuracy	Stability	Types of error
Volumetric soil moisture (up to 5cm depth)	Initialisation	50km	Daily	$0.035m^3/m^3$		SSEOB
	trend monitoring	50km	Daily	$0.04m^3/m^3$	$0.01m^3/m^3/yr$	SSAOB

Table 15. GCOS and modelling requirements for soil moisture

Soil moisture is widely used to initialise surface fields in models and is of particular importance for seasonal climate predictions and the monitoring of moisture anomalies on the terrestrial land surface. There is strong need for consistency in this ECV with other ECVs for example temperature, surface humidity, albedo, vegetation and precipitation. No differences between global and regional modelers were expressed in the requirements.

According to the CMUG user survey, soil moisture observations are essential in all application domains for climate modelling. The widest expected use of soil moisture data is in the field of model development (process studies) and model evaluation. 100% of the expert users are interested in using soil moisture data for these applications, while 60% use it also for model initialization and climate monitoring and attribution.

GCOS ECV	Model Initialisation	Prescribe Boundary Conditions	Re-analyses	Data Assimilation	Model Development and Validation	Climate Monitoring/ Attribution	Q/C in situ data
Soil moisture	X	X	X	X	X	X	X
Fraction of expert users	60%	20%	40%	50%	**100%**	60%	10%

Table 16. Use of ECV soil moisture in climate modelling applications and fractions of expert users being interested in using ECV soil moisture dataset for these applications

The detailed requirements for ECV soil moisture collected from the experts interviewed by CMUG are summarized in Tables 16 and 17. As the results are somehow contrary to the current GCOS requirements, a discussion of the individual requirements is made in the following:

- **Horizontal resolution**: The horizontal resolution requirement identified by the experts is more stringent than the current GCOS requirement. As both global and regional climate models have increased in spatial resolution throughout recent years and will continue to go to higher spatial resolutions in the future, the users also expect that observational soil moisture datasets are provided at higher spatial resolutions. The upper limit for the horizontal resolution is 50 km, but many users required information on spatial scales much better than 10 km (most even at the 1km scale). While these high spatial resolutions are not available for the past decades, SENTINEL-1 provides a unique opportunity to provide this information to the user community.

- **Observing cycle**: The requirements for the observation cycle are consistent with the GCOS requirements. Most users and applications require daily data. Monthly data might be sufficient for some applications like e.g. trend monitoring, while even higher (sub-daily) temporal resolution would be desired for special process studies.

- **Data quality**: The requirements for the data quality are separated into the following:

  Precision/Accuracy: The quantification of the desired precision and accuracy for different applications was difficult for most of the expert users. While some of them gave accuracy values similar to the known GCOS criteria (0.04 m3/m³), others argued that the error should be given e.g. as a fraction of the dynamical range of the data which is in the order of 0.5 [m³/m³]. Users expressed their desire to have an accuracy comparable to 5% to 10% relative to the dynamical range, which would correspond to an accuracy of 0.025 to 0.05 [m³/m³]. These numbers of valid for the spatial scales envisaged by the users (~1 … 25x25 km²). More stringent requirements might apply on larger spatial scales due to the spatial aggregation of potential errors. In addition several users emphasized the need for information on the depth where the soil moisture data is sensed.

  Temporal stability: A quantification of the temporal stability criteria for long term soil moisture records was obviously not possible by the interviewed experts. However all interviewed users indicated that they give high priority to a temporally stable long term data record. Overall the most important aspect for the users is that the datasets show a long term stability without sudden jumps or data gaps. A quantitative accuracy measure was not given and is therefore not provided in the summary table.

  Error measures: All users agree that in case of individual sensor measurements, the uncertainty on the single sensor retrievals shall be provided, while for L3 data the uncertainty of the merged product is needed. The latter would require an uncertainty model to quantify adequately uncertainties from spatial upscaling/regridding procedures as well as effects of spatiotemporal sampling patterns on random and systematic error components.

Parameter	Application	Horizontal Resolution	Observing Cycle	Precision	Accuracy	Stability	Types of error
Soil Moisture							
Volumetric SM	trend monitoring	< 1 km² to 25x25 km²	Daily … monthly	0.005 – 0.01 [m³/m³]	0.5vol.%(SH)	No information available	SSECDR SSACDR, ERRMERG
	model initialisation / boundary condition	< 1 km² to 25x25 km²	Daily	0.005 - 0.035 [m³/m³]	1% / 0.5% (SH) Larger deviations are of less concerns than for temporal	No information available	SSECDR SSACDR, ERRMERG ERRCOV

					anomalies (also strong spatial variability)		
	Validation	< 1 km² to 25x25 km²	Daily	0.005 - 0.035 [m³/m³]	-	No information available	SSECDR SSACDR, ERRMERG
	Monitoring/ Attribution	< 1 km² to 25x25 km²	Daily	0.005 - 0.035 [m³/m³]	-	No information available	SSECDR SSACDR, ERRMERG
	Data assimilation	< 1 km² to 25x25 km²	Daily	0.005 - 0.035 [m³/m³]	0.04 [m³/m³]	No information available	SSECDR SSACDR, ERRMERG
	Model development and validation	< 1 km² to 50x50 km²	Daily	0.005 - 0.035 [m³/m³]	0.04 [m³/m³]	No information available	ERRMERG
Volumetric soil moisture (temporal anomalies, i.e. removing long term mean)	trend monitoring	< 1 km² to 25x25 km²	Daily	No information available	Larger deviations are of less concerns than for temporal anomalies (also strong spatial variability)	No information available	SSECDR SSACDR, ERRMERG
	Prescribe model boundary condition	< 1 km² to 25x25 km²	Daily	0.005 - 0.035 [m³/m³]	Known & constant/	No information available	SSECDR SSACDR, ERRMERG
Soil moisture anomalies	trend monitoring	< 1 km² to 25x25 km²	Daily … monthly	min{0.04 [m³/m³]; 10% relative of anomaly}	min{0.04 [m³/m³]; 5% relative of anomaly}	No information available	SSECDR SSACDR, ERRMERG
	Prescribe model boundary condition	< 1 km² to 25x25 km²	Daily	min{0.04 [m³/m³]; 10% relative of anomaly}	min{0.04 [m³/m³]; 10% relative of anomaly}	No information available	SSECDR SSACDR, ERRMERG
Profile soil moisture proxy	trend monitoring	< 1 km² to 25x25 km²	Daily … monthly	1 mm over rooting depth	1 mm	No information available	SSECDR SSACDR, ERRMERG
	Prescribe model boundary condition	< 1 km² to 25x25 km²	Daily … monthly	1 mm over rooting depth	1 mm	No information available	SSECDR SSACDR, ERRMERG

Table 17: Summary of use requirements for ECV soil moisture as collected from the CMUG interviewed experts.

5. Across-ECV requirements

To ensure consistency between ECV datasets which is important for climate modelling and reanalyses there are a number of considerations that should be respected for the CCI projects. Also to facilitate common practices the CCI should converge on terminology as this can be different for each ECV project and will enhance communication across the project.

Firstly the ECV projects should all use the same level 1 datasets as input to their level 2 processing. Some of the ESA FCDRs (e.g. AATSR) have recently been regenerated with improved calibration, geolocation etc. and there needs to be a clear steer from ESA at least for ESA satellites what are the recommended level 1&2 datasets to use. Table 18 shows which sensors are common to which ECVs.

Secondly some ECVs will benefit from access to other ECVs being generated from within the CCI project to explore synergies and also where one ECV's retrieval can benefit from another. Table 19 attempts to identify where these cross-linkages are between ECVs.

Thirdly the use of common ancillary fields will be important. ERA-Interim will be a good source of atmospheric fields from 1980 onwards with ERA-40 available before that. This would ensure a consistent assumption about the atmospheric state for all ECV datasets. The next reanalysis will be ERA5 with improvements to the model and observational datasets. This however will not be ready in time for the CCI projects at least in phase 2 of the CCI. For surface fields an agreed SINGLE source for surface albedo, vegetation (LAI, FAPAR), emissivity, ice caps and glacier climatology, sea ice, SST etc should be defined and agreed by the CCI projects. If this is not done inevitable inconsistencies will be seen in the products which will be only due to different representations of the atmosphere/surface being assumed. A common land/sea/lake mask also needs to be adopted by all ECV projects. The Land Cover inland waters dataset is a good example of one which several ECV projects would benefit from.

The horizontal grids should be common to level 3 products to enable easy comparisons and processing of data from different ECV CDRs. Similarly the definition of atmospheric layering should be common across ECVs (e.g. aerosol and clouds) for level 2 and 3 products.

Finally the specification of error characteristics should be provided in a consistent way and where appropriate separated into precision, accuracy and stability. The errors should also be specified, where possible, for each individual measurement.

	SST	Sea level	Ocean colour	Sea-ice	Clouds	GHG	Aerosol	Ozone	Fire	Land cover	Glaciers	Soil moisture	Ice sheets
AATSR/ATSR-2/ATSR-1	●				●		●		●				
MERIS			●		●	●	●		●	●			
SPOT 4,5											●		
SPOT VGT									●	●			
Landsat TM/ETM+				●							●		●
SAR (ENVISAT/ERS/ALOS/TSX/PALSAR)				●						●	●		●
SEVIRI	●				●								
MODIS			●	●	●		●						
Sciamachy						●	●	●					
GOSAT						●							
GOME-1/2							●	●					
AVHRR (L2P)	●			●	●		●					●	
GOMOS							●	●					
IASI						●							
AIRS						●	●						
AMSU						●							
ACE, SciSAT						●		●					
SeaWIFS			●										
MIPAS						●		●					
OMI							●	●					
Radar Altimeters (TOPEX-POSEIDON)		●											
Radar Altimeters (JASON-1/2)		●											
Radar Altimeters (GEOSAT-Follow-on)		●											
Radar Altimeters (ENVISAT, ERS, GDR, MGDR)		●		●							●		●
Scatterometers				●								●	
SMMR				●								●	
TMI	●											●	
SMOS												●	
AMSR, AMSR-2, -E	●			●								●	
WINDSAT												●	
SSM/I & SSMIS				●								●	
PARASOL							●						
ASTER											●		●
ICESAT				●							●		●
ODIN/OSIRIS							●	●					
ODIN/SMR							●	●					
POLDER							●						
CRYOSAT 1/2		●		●									
CALIPSO/Caliop					●								
CLOUDSAT/CPR					●								
TOMS							●						
RadarSAT											●		●
TerraSAR-X											●		●
Cosmo-Skymed											●		●
ALOS Palsar											●		●
IRS1C/1D											●		
SENTINEL 2 MSI											●		
LDCM-OLI											●		
AMI-WS												●	
ASAR, ASAR-WSM.										●		●	
ASAR G-POD										●			
ASCAT												●	
SPOT-HR										●			
MODIS_AQUA			●										
CZCS			●										
Radar Altimeters Follow on (GFO GDP NOAA)		●											
Radar Altimeters Follow on Saral/Altika		●											
Radar Altimeters Follow on CYROSAT2		●											
Radar Altimeters Follow on Sentinel-3		●											
Source DARD version	v1 01/12	v1.3 09/14	v1.6 03/14 s2.3	directs to NSIDC, EASE-Grid Sea Ice Motion Vectors	Phase 2 v1.7 06/14	v4 09/14	Phase 1 v4.1 03/11	Data Products frontpage 03/13	v1.3 06/11	v1.9 01/13 Table 7-1	Phase 1 v1.0 11/11 Table 3.1	v1.2 04/12 Table 1	v1.7.6 10/12 Table 3.1.1

Table 18. Primary sensors for each ECV project as given in the DARDs and phase 2 plans.

	SST	Sea level	Clouds	Sea ice	Ocean colour	Aerosol	GHG	Landcover	Fire	Ozone	Glaciers& Ice Caps	Ice Sheets	Soil Moisture
SST		x X	x X	X X	X X	x x				x			
Sea level	x x			x x				x			x	x x	x
Clouds	x x			x x	X	x x	x X	X X	x	X X			X
Sea ice	x X	x	x X		X					x	x		
Ocean colour	X x		x X	x x		x X							
Aerosol	X		x X	X	X x		x	X x	x X	X X	x		x
GHG		x	x X		x	x X	x	x	x X	X		x	
Landcover		x	x X			x X	X		x X	X	x X	X	x X
Fire			x X			x x	x x	X x		x x	x		x x
Ozone	x		x X	x		x x	X x						
Glaciers & Ice Caps		x		x				X X				x	x
Ice Sheets	X	x X			X						x x		
Soil moisture		x	X				X	X	X		x		

Table 19: An analysis of cross linkages between ECVs indicating where comparisons need to made to ensure consistency. The left hand column is the project with the identified need, the top horizontal row is the provider. Larger crosses indicate where the CDRs generated by that ECV project would potentially be of use in the retrieval of the ECV listed on the left side. Red crosses from CMUG Phase 1 User Requirements Document, and black crosses from CMUG Phase 2 User Requirements Document.

6. Requirements for other ECVS

The CMUG user community was asked to assess the relative priority, (high, medium and low), of other ECVs for Climate Modelling and Analysis not covered by the ESA CCI project. The results are summarised in Figure 3.

From the 75 respondents, the highest priority was given to precipitation (16), snow cover (15) and soil moisture surface and root zone (14). The next highest priority ECVs were radiation budget, water vapour and albedo (each 13). Medium priority was given to river discharge (11), permafrost (9), upper-air wind and leaf area index (8 each). Where stated, lowest priority was given to sea state, lake levels and biomass (4 each).

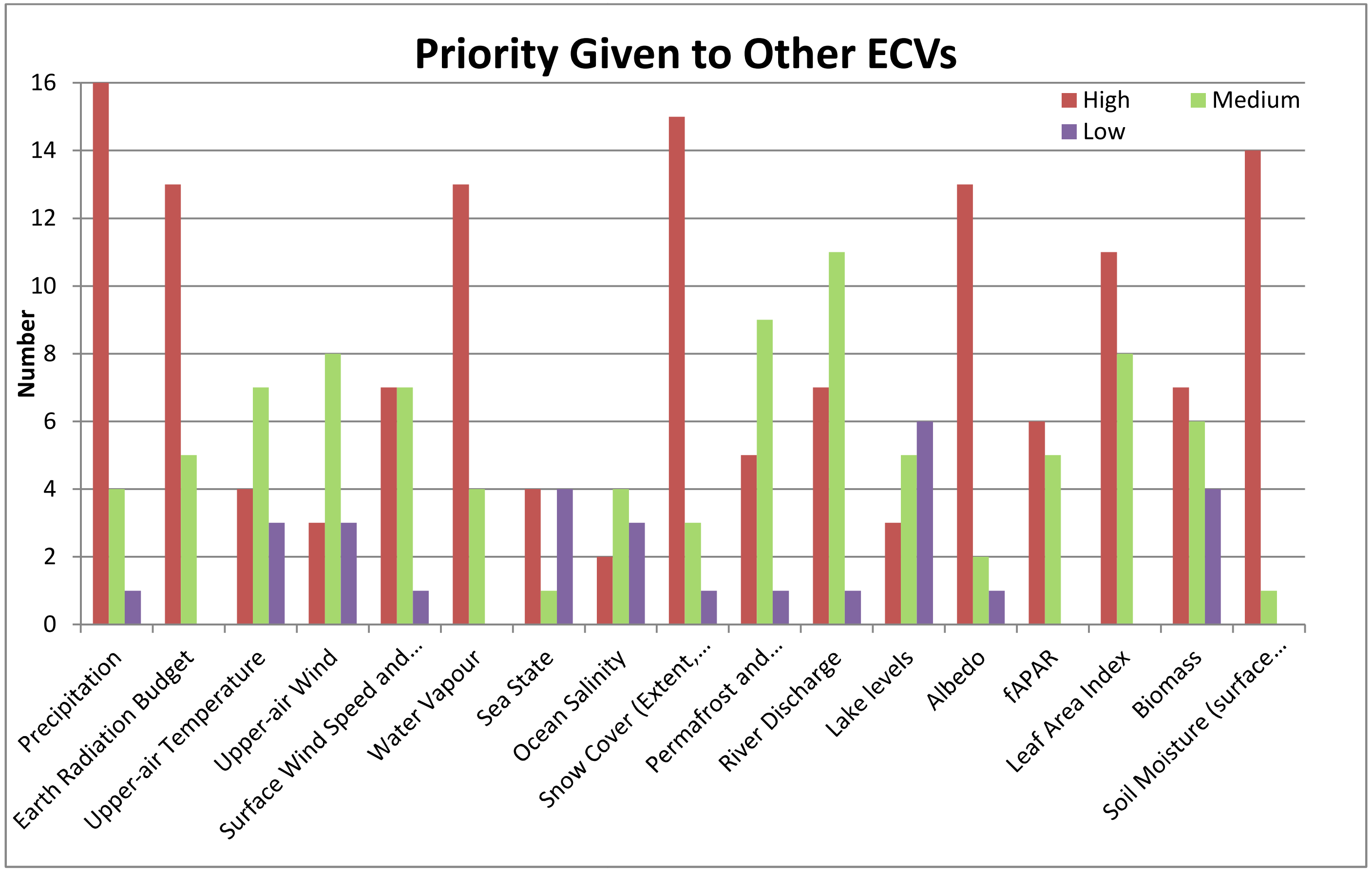

Figure 3. Priority given to other ECVs for climate modelling and analysis.

7. Requirements for Climate Service Datasets

Climate data records (CDRs) are specialist data, and the main users are in the climate research community. True "end users", who require climate information in their decision-making, are unlikely to use CDRs directly. Some, such as risk modelling companies for the insurance industry, tend to be able to handle more 'raw' data, while other users, for instance farmers, require something in a format they can understand quickly. Translational users such as environmental consultancies, who sit in between the science specialists and the decision-maker, will add value to the data as part of a wider service offering. Some key areas here include:

- Sectoral studies – e.g. Agriculture: climate change impacts on yield, air quality impact on yield, pest and disease impacts
- Attribution studies (to link changes in datasets with man-made or natural events)
- Sustainability – climate change and impact on resources and environmental impact assessments (e.g. for large infrastructure projects)
- Resilience planning - climate change and impact companies and infrastructure, impact on vegetation (disease etc), urban planning/land use, insurance and reinsurance
- Climate fact sheets for regional studies or applications or to put in context current weather/climate events
- Adaptation studies and adaptation impact monitoring
- Hazard and event monitoring and information of the general public (e.g. drought events)

For all of these products, seasonal and surface data will be important and will need to be combined with other available datasets. Typical requirements for datasets for operational climate services are:

- Simple user documentation on reading data and about the data characteristics
- Recognised format that is widely used (NetCDF4)
- DOI (from a recognized issuer)
- Uncertainty information on each parameter included
- Well validated as documented in a peer review paper
- Maturity matrix score documented and above a predefined value
- Ease of access on a recognized robust server with a given protocol (FTP)
- Timeliness for some ECVs (e.g. within 1 month of occurrence)
- Sustainability needing long term (>10yrs) archive commitment
- Ability for users to feedback comments on datasets to generators and other users
- Access to information of user applications
- Scientifically robust production e.g. through ensemble or reanalyses

Climate service users indicated e.g. the wish to have two kind of production chains

a) Regularly updated (< 1 month) dataset with homogeneous processing that enables best comparison against long term record. A good example might be global reanalysis which often is performed with less advanced by homogenous processing approaches

b) Reprocessed climate records with new (improved) processing updated e.g. once per year.

8. Requirements for Obs4MIPs datasets

This section gives the current requirements that ECV datasets need to comply to in order to be included in the Obs4MIPs database which is used by the CMIP modelling community for comparing satellite observations with climate model predictions (Teixeira *et. al*, 2014). All the information is also provided on the Obs4MIPs site at: https://www.earthsystemcog.org/projects/obs4mips/how_to_contribute. We refer to this website for any updates on the requirements of Obs4MIPs.

8.1 Criteria for Datasets to be included in Obs4MIPs

Observational datasets for Obs4MIPs Phase 1 must fulfil the following criteria:

- Has clear traceability from level 1 measured variables to retrieved variables in level 3 or 4 dataset
- be based on data that has a history of peer reviewed publications,
- is version controlled, with doi,
- reside in a long term and maintained archive,
- span a time period long enough to be of use for model comparison (3 years is a useful minimum although in some circumstances shorter data records may be considered),
- match a model variable in the CMIP5 protocol
- include an estimate of the uncertainty for each variable verified by validation of the retrieved variables

Most CCI datasets which are relevant to Obs4MIPs should conform to the above although it will take time for a history of peer reviewed publications to be available.

8.2 Input Dataset Gridding

The datasets for consideration for Obs4MIPs should be Level 3 (single sensor) or level 4 (multiple sensors) datasets which have been transformed on to a 1 degree grid square through averaging and/or interpolation and then averaged over 1 month. Researchers should be mindful to check on the Obs4MIPs website that they are conforming to the latest specification. For each grid square the fields should be complete (i.e. no data voids) and consideration should be given to ensuring the variables are still conserved in the re-gridding. It is assumed only observational data (i.e. no model analyses) are included. The associated uncertainties also need to be provided on 1 deg grid and care has to be taken to derive these from the level 2 single field of view observations. Any biases in the original observations will propagate through to the gridded data but random errors will be reduced, hence averaging of uncertainties may not be appropriate. More details on the requirements are given here: https://www.earthsystemcog.org/site_media/projects/obs4mips/obs4MIPsDatasetRequirements_v1.2.pdf

There are several issues to bear in mind when regridding data from level 2 to level 3 especially when dealing with uncertainties. These include maintaining consistency between variables after the regridding, consistently dealing with coastal areas correctly and how to fill data voids. It would be a good idea for data producers to share experiences on their regridding methodology.

8.3 Data Format

The file must be written in NetCDF version 4 and must follow the standard NetCDF Climate and Forecast (CF) Metadata convention http://cf-pcmdi.llnl.gov/
The output file must pass a CF compliance check. A checker is available at:
http://puma.nerc.ac.uk/cgi-bin/cf-checker.pl
Choose the latest CF version when submitting the file for checking. Each output file must contain a time series of ONLY ONE physical variable (e.g. sea surface temperature, specific humidity). If the entire time series can be stored in less than 2GB, it must be stored in a SINGLE file. If it requires more than 2GB, it should be split into the minimum number of files required, with the size of each file being less than 2GB. Each file should contain a contiguous time series of complete data grid blocks. Each file must contain all of the required metadata applicable to the data subset contained in the file. Some software is provided at: http://cmip-pcmdi.llnl.gov/cmip5/obs4cmip5.html to write datasets in the compliant format.

Each physical variable and coordinate variable must use the specified output/coordinate variable name given in the CMIP5 Requested Output list (standard_output.xls). For example, the latitude output name must be “lat”, and the air temperature output variable name must be “ta”.

Sharing of experience on writing the compliant format datasets from the climate datasets would be worthwhile. Feedback on any problems should be given to the Obs4MIPs team.

8.4 Documentation

A short technical note (5 pages max) must be provided with the dataset that conforms to the obs4MIPs technical note template:
https://earthsystemcog.org/site_media/projects/obs4mips/Obs4MIPsTechnicalNoteGuidancev3.pdf
It should be written bearing in mind the reader will not be familiar with satellite datasets. One important point to bear in mind is that there may be other datasets of the same variable available on the Obs4MIPs site and so the note should make it clear what are the advantages of the CCI datasets with respect to previous datasets already available through Obs4MIPs.

These technical notes are valuable in their own right to promote the datasets. CCI teams should make them available on their web sites.

8.5 Process for submission of datasets

There is a proposal form for dataset owners to complete here:
https://earthsystemcog.org/site_media/projects/obs4mips/obs4MIPs.DataSet.Form.v0.1.pdf
and submit by email to the Obs4MIPs team:
https://www.earthsystemcog.org/projects/obs4mips/contactus/

The CCI project should keep a record of which datasets have been submitted to Obs4MIPs and which are available there.

9. Requirements for observation simulators and other tools

9.1 Observation simulators

Climate modellers not only require the satellite CDRs from the CCI projects for all 13 ECVs but also for some of the CDRs observation operators or satellite simulators to convert the model state variables to the satellite measured variable are required. These operators are normally in the form of a generic software package that can be "plugged" into any climate model and interfaced with the model variables. The COSP package (Bodas-Salcedo *et. al.* 2011) is a good example of this and contains a list of observation operators for many different satellite datasets, including Top of Atmosphere radiances, ISCCP, CloudSat, CALIPSO, HIRS and SSM/I.

The requirements for operators for each of the 13 ECVs will need to be considered. Currently it is envisaged that the observation operators listed in Table 20 will be required for the CCI datasets where the model variables are converted to a satellite observed quantity.

ECV	**Model variable**	**Satellite variable to simulate**
Atmospheric		
Cloud properties	Liquid/Ice concn profile	Cloud amount/top pressure
	Fractional cloud cover	Equivalent cloud cover
Ozone	Ozone concn profile	Total column ozone
Greenhouse gases	CO_2 and CH_4 profiles	Total column CO_2 and CH_4
Aerosols	Aerosol concn profile	Aerosol optical depth
Oceanic		
SST	Sea surface bulk temp	Sea surface skin temp
Sea level	N/A	N/A
Sea-ice	Sea-ice thickness	Area mean freeboard
	Sea-ice concentration	MW br. temps
Ocean colour	Phytoplankton concn	Derived chlorophyll alpha
Terrestrial		
Glaciers and ice caps	N/A	N/A
Land cover (inc veg)	N/A	N/A
Fire	N/A	N/A
Ice Sheets	N/A	N/A
Soil Moisture	Soil moisture	a) surface soil moisture b) surface saturation degree

Table 20. Observation simulators required for CCI datasets.

The funding for the development and maintenance of an observation simulator package such as COSP is still not assured for many ECVs as it falls between the modelling community and observation community.

The CCI project must ensure observations simulators for their observations are available to facilitate comparison with model fields.

9.2 Tools for CDRs

Data Ingest

It is vital that climate modellers are able to easily ingest the CCI datasets into their model environments and analyse them. The format should be familiar to users (see next section) and plug in modules to ingest the data formats should be provided in commonly used software environments (e.g. Fortran, IDL, Python, Perl, ..). ***All users will need the ingest software*** unless the format is a really standard gridded dataset and so it is critical to make this part of any tool box well documented and easy to use. The final result is to populate a variable array within the users particular application.

To make reading the datasets as easy as possible a small software package consisting of source code, documentation, build scripts, and installation tests (sample input data and expected output from test programs in order to verify correct installation) is envisaged as an effective solution by climate modellers.

Note that with all the tools developed there is a maintenance cost implied as different software operating systems change and so the associated tools need to maintain compatibility with the latest version and a few of the previous versions.

Access to Metadata

There are various metadata required to be made available with the satellite CDRs. This also should be documented. Examples include a timeline of both satellite and instrument related anomalies, documentation on version of level 1 processing, what ancillary datasets have been used in the level 2 processing etc.

User Commentary and Annotations (CHARMe)

Allowing researchers to comment on their data use is recognised as a valuable tool for 'crowd-sourcing' user experiences of data. These user annotations can be about the strengths of the data such as links to papers, technical notes, validation campaigns, provenance (algorithm development), application, uncertainty information, notes about external events (El-Nino, volcanic eruptions), or maturity information. Equally, annotations could describe weaknesses in the data such as sensor failure, discontinuity, or restrictive data policies.

The CHARMe system was developed in an FP7 research project to allow climate data sets to be annotated with such commentary information. Comments can be grouped by type (references, narrative comments) or linked to timeline or geospatial plots.

The value of CHARMe to data providers is that it allows rapid feedback about data quality, plus knowledge about how the data is used and other contextual information that can inform future data development.

Data Analysis

Once the CDRs have been ingested into the local software systems the major modelling and climate research centres will probably have the tools they need to process the datasets but nevertheless the option of tools to provide some simple data processing should be provided. There are already several software packages available such as GRADS (http://www.iges.org/grads/) developed for the modelling community, UV-CDAT (http://uvcdat.llnl.gov/) and the NASA Panoply (http://www.giss.nasa.gov/tools/panoply/) software which already perform many functions and so any tool developed would need to add capabilities to these existing packages.

Climate Data Operators (CDO) are widely used by the climate modelling community because of their ease of processing, and a good test for a dataset is to see whether CDOs can work with it (as well as working with visualisation tools).

University groups would benefit from additional processing tools. Examples of commonly required tools for a variable ***Z*** with an uncertainty ***Err*** might be:

- Extracting geographical regions from global datasets for time t: $Z(global, t) \rightarrow Z(x_1, y_1, x_2, y_2, ,t)$ and the same for uncertainty: $Err(global, t) \rightarrow Err(x_1, y_1, x_2, y_2, ,t)$
- Extracting a time series for a particular location x, y: $Z(x,y,t_1\text{-}t_2)$, $Err(x,y,t_1\text{-}t_2)$
- Producing anomaly plots (means and variance) both as maps and time series e.g. $Z(x,y,t)\text{-}Z_{mean}(x,y)$ and $Var(Z(x,y,t)\text{-}Z_{mean}(x,y))$
- Plotting variance of ensemble of datasets as maps and time series e.g. $Var(Z_{1..n}(x,y,t))$
- Computing empirical orthogonal functions from the data
- Computing Fourier series from the data
- Comparisons with other observational O_n and model M_n datasets as maps and time series e.g. $Z(global, t)\text{-}O_n(global,t)$and $Z(global, t)\text{-}M_n(global,t)$ where n can be 1-30 for example for CMIP5 model datasets.
- Providing a tool to plot some of the plots listed above from different ECVs side by side.
- Support for several commonly used map projections (see sec 10.5)

It might be envisaged if the tools are in an easy to develop form then the users could actually contribute to the tools themselves and make them available to the toolbox which would be maintained.

Climate Model Evaluation Tools

Model benchmarking initiatives have become increasingly important to evaluate the quality of coupled Earth System Models (ESMs) and to support the model development process. In the frame of CMIP, the WGNE/WGCM Climate Model Metrics Panel has been established to

define model performance metrics for model-data intercomparison and analyse various aspects of ESM simulations with a multitude of observational datasets. However, ESA data is currently not used in the context of routine model evaluation. There will be a clear set of user requirements coming from the CMC and the developers of the evaluation tools for the climate observation data sets they need.
Some of these requirements will be scoped out by the CMUG work with the ESMVal tool and CCI data sets.

Climate Monitoring Facility

Two important requirements the ECVs need to meet in order to be suitable for climate change studies and services are the long-term data homogeneity and consistency with related variables. Both aspects can be assessed using the Climate Monitoring Facility (CMF) tool under development at ECMWF. This tool consists of a web-interface for on-demand time-series plots of monthly mean area averaged statistics that, in the first instance, are pre-calculated and stored in its database (CMFDb). In addition to observations, the CMFDb also includes a large variety of model fields from several recent reanalyses. The requirements for the CMF span technical aspects, data manipulation aspects, and documentation. In particular, the interface should be user-friendly, should provide a fast time response - about 3 seconds for pre-calculated statistics -, should allow users to select multiple fields - this is important to check the data consistency - and to eventually customize the resulting page (i.e. number of plots per page, colour scale, axes, etc…). Downloading plots and data in a number of well-known formats will also be possible. A number of requirements on the data manipulation and representation have already been identified. These include representation of uncertainties, trivial math operations on the fly, overlay of "external" datasets not available in the CMFDb, possibility of extending the CMF capability beyond the monthly mean time series and, depending on the time response, moving towards a completely on the fly calculation of mean statistics. To ensure the user-friendliness of the CMF, documentation should also be provided. This will include information on how to use the CMF and interpret the resulting plots, description of the available datasets and variables, a CMF and data disclaimers, and metadata information (e.g. data units, data source, data version). The CMF should also take advantage of the CHARMe FP7 project that has developed a software system that can be plugged in to data provider’s sites to enable users to provide feedback and/or link pertinent information on the datasets themselves. This online feedback system is seen as an important step in sharing the collective experience of climate data users. Finally, email and contact information for user help and support should be clearly placed on the web-interface.

Co-location software and data

For most of the CDRs they should be accompanied by colocation software with datasets of in-situ measurements (e.g. buoys for SST, ozone sondes for ozone, fire radiative power for burnt area) to assist a wide range of users in the validation of the datasets. Tools for the spatial interpolation of the data to allow for a resampling of the observational data would be useful.

10. Requirements for data formats and data access

10.1 Naming conventions and documentation

In order to make life simple for users the naming conventions for files, datasets and variables must be commonly agreed between users and data producers. A recommended naming convention for individual variables for the CDRs can be accessed here: http://cf-pcmdi.llnl.gov/documents/cf-standard-names/standard-name-table/15/cf-standard-name-table.html together with guidance on what the convention is: http://cf-pcmdi.llnl.gov/documents/cf-standard-names/guidelines

For example we have *sea surface skin temperature* and *sea surface subskin temperature* for the SST ECV. There are some variables that will still need to be defined as this list does not cover all variables in the CCI ECV list. A data reference syntax is being defined as part of CMIP5 and should be followed for the CCI datasets also, see: http://pcmdi-cmip.llnl.gov/cmip5/output_req.html#req_format

There is also a recommended way for CCI projects to add new variable names which can be adopted by the CMC. One way to ensure easy to use datasets is to impose a consistent naming convention across the ECV projects and beyond.

The CMC stressed the importance of well-document data and its error estimates. A technical note similar to the one requested in obs4MIPs has been suggested as a template for all ESA CCIs to follow for data documentation. It provides information on the data field description, data origin, validation and uncertainty estimate, considerations for use in model evaluation, and an instrument overview.

All ESA CCI data should be submitted to Obs4MIPs to facilitate routine model evaluation with evaluation tools. **The CMUG recommendation is to provide the datasets on the native resolution**. In addition, it might be useful to pre-process the datasets to some common grid/spatial resolutions (e.g. lat/lon, 0.5°, lat/lon 1°). The guidelines and specific requirements for Obs4MIPs should be monitored and followed.

The short technical note for climate scientists with no knowledge of satellite datasets is recommended for each ECV. It should highlight the advantages of each datasets and its main characteristics. An example is given here from the NASA project: http://oodt.jpl.nasa.gov/wiki/display/CLIMATE which includes the following guidelines for a technical note:

- *The target audience is the analysis community that will evaluate the climate model experiments in CMIP5, who have little experience with NASA datasets.*
- *The technical note should be written at the graduate student level.*
- *The note must be specific to one particular satellite observation dataset, which must contain a single variable.*

- *The note should summarize essential information for comparing the dataset to model output.*
- *Anything of interest only to experts should be referenced, but not include in the main body of the note.*
- *An appropriate length for the note (from Section 1 to 6 in the template) is 3-5 pages, excluding tables and figures.*

The template mentioned in the last bullet point begins on the next page. Some instruments or projects will provide datasets for multiple variables. The CMIP requirements state that each variable must be contained in a separate file. The guidance is to provide a technical note for each variable, even if it means substantial duplication of text across the notes.

Concerning data documentation, the CMC stressed the importance of well-documented data, in particular regarding the error estimates. A technical note similar to the one requested in Phase 1 on obs4MIPs has been suggested as an example of data documentation. Uncertainty information should be provided with the data products. Concerning major obstacles/problems in current satellite data usage, the CMC stressed user friendly quality information and traceability.

10.2 Data formats

The users were asked for their preferred format for the CDRs and 91% replied NetCDF with the remainder happy with any standard format (e.g. GRIB). 36% of respondents specifically asked for a CF compliant NetCDF v4 dataset. Specifically for NetCDF CF, additional attributes in the file should be provided to ensure it is easily identifiable by man and machine. A good example of the use of additional attributes is provided by the PCMDI CMOR (Climate Model Output Re-writer) package, which is used to standardise climate model output from the CMIP5 project. The convention for NetCDF files for climate datasets are published here: http://cf-pcmdi.llnl.gov/documents/cf-conventions/1.4 The use of swath based data (levels 1 and 2)) in NetCDF is still under development but remains the preferred option.

A prerequisite for extensive use of the data within the climate community that was stressed by CMC is that the data should be provided in CF compliant netCDF data in the same format than the CMIP format guidelines.
The file format should be chosen so that the data can be delivered through the same range of services as the climate model output it is intended to validate. For the metadata: An XML document with a well defined schema which clearly defines the instrument, its measurement technique and the analysis method used to retrieve the data record. It would be extremely helpful if the schema could, at the top level at least, share some of the structure which has been developed by the EU FP7 project METAFOR to describe climate models and their output. For example, descriptions of institutions could use the same schema elements.

10.3 Data access

For getting access to the data 91% of the respondents specifically requested FTP access, 30% requested web access via a browser, while 10% indicated a preference for access through another channel (central server, ESGF, Obs4MIPS, OpenDAP, CMIP5, Geoserver, BADC server). There is a need to be able to subset in time and space the datasets in a convenient way such as OpenDAP. Other physical media such as DVD were not generally supported. Access from recognised data centres such as NASA DAAC, PMCDI and EUMETSAT UMARF were also stated as a requirement reflecting the support they can provide to users.

As far as the location of the datasets is concerned they should be hosted on a node of the Earth System Grid Federation (ESGF) via the Obs4MIPs archive so that users will have the same interface for European, US and other climate datasets. They need to be hosted on the ESGF "data nodes" which publish to "gateway nodes" see http://esg-pcmdi.llnl.gov/ for more details. The BADC is currently connected to the Grid and would provide a suitable host for CCI datasets.

10.4 Level of processing

The user community was asked which level of processing they required for their applications from level 1 geophysical measurements (e.g. radiances), level 2 (derived products on original space view) or level 3 (e.g. daily, monthly means gridded products). The results are summarised in Table 21 which shows a fairly even split between level 2 and level 3 processed products. However, ten of those who selected level 2 products also use level 3 products and other combinations (e.g. l & 2, 1 & 3) were recorded.

Preference depended on the application. For assimilation, level 2 is required. For climate monitoring, level 3 is acceptable, but there must be traceability back to the sensor measurement and good documentation of the processing, because climate scientists need to understand how the variable has been calculated.

Processing Level	No. of users	Percentage of users
Level 1	3	7.7 %
Level 2	17*	43.6 %
Level 3	16	41.0 %
Any or All	3	7.7 %
Total	39	100.0 %

*Table 21. Feedback from users on required level of processing (*10 of these selected L2 and L3)*

CMUG also sought user views on whether single sensor datasets or merged datasets would be required for level 3 gridded data products. The results depended on the ECV being considered. Table 22 suggests a fairly equal split between merged and single sensor products, and a similar number required any or both. Single sensor products are preferred by some for observation system simulation experiments (OSSE), bias correction etc. Some preferred merged products for better spatial and temporal coverage and more robust results, provided (again) that there is traceability back to the sensor measurement and good documentation of the processing. The disadvantage of merged products is that the error characteristics are more complex and single sensor products are preferred at level 1 or level 2 for reanalyses.

Single or Merged dataset	No. of users	Percentage of users
Single sensor datasets	10	27.0 %
Merged product datasets	13	35.1 %
Any or Both	14	37.8 %
Total	37	100.0 %

Table 22. Feedback from users on single sensor vs merged products

10.5 Geospatial projections

Geospatial datasets have to be stored in a specific projection and this can cause problems in the analysis of the datasets (e.g. data day definition). The important thing is to provide simple tools to translate between any projection and a basic lat/lon grid. The CCI datasets should all share a common projection where possible to facilitate the joint analysis of different datasets from different ECVs. Land/Sea/Lake masks are also important to be common between the ECV projects otherwise inconsistencies will be seen due to the use of different masks.

11. Summary

The CMUG has carried out a survey of the climate modelling community and present an analysis here. One important finding is that the majority of modellers surveyed want to use the CCI datasets for model evaluation and development and only a few are engaged in climate monitoring.

An analysis of the individual requirements for climate modelling for the 13 CCI ECVs has been carried out with the following inputs:

- GCOS requirements
- Inputs from CMUG interviews
- Comments and analysis through interaction by CMUG researchers with the climate modelling and reanalysis community and research meetings over the last year

This has enabled the CMUG to undertake an analysis of how well the current GCOS requirements meet the needs of climate modellers and how the initial thoughts of the CMC match these requirements. This can be used as input to the CCI requirements specification as it evolves and is a good basis for discussions.

Comments on the technical details of the proposed CCI datasets were also sought on format and data access in order to gain an overview of the preferred formats for the climate modelling community. The majority view was for CF compliant NetCDF format similar to CMIP output with access via FTP or browser interfaces through the ESGF via the Obs4MIPs archive which is the same interface climate modellers are using.

Another strong recommendation is that ESA CCI data should be submitted to Obs4MIPs to facilitate routine model evaluation with evaluation tools. The CMUG recommendation is to provide the datasets on the native resolution. In addition, it might be useful to pre-process the datasets to some common grid/spatial resolutions (e.g. lat/lon, 0.5°, lat/lon 1°). The guidelines and specific requirements for Obs4MIPs should be monitored and followed.

CMUG believes the CCI will meet the requirements of GCOS for most but not all ECVs, and the exceptions are due to limitations of the observational datasets. It is recognised that the climate observation data needs of the CMC can evolve, hence the need to re-consult at future dates with the CMC and revise this document accordingly.

The recent survey by CMUG for user requirements and the definition of the GCOS requirements has shown that it is in many cases very difficult for the users to give quantitative information on potential uncertainties for ECV data products. This is due to the fact that there is a lack of quantitative information on the impact of different observation errors at different spatial scales for the variety of applications addressed in this document. For critical applications dedicated sensitivity studies can be carried out to give a more solid quantification of user requirements for specific spatial and timescales.

12. References

Banks, H., N. Bellouin, D. Bernie, S. Brown, N. Butchart, J. Caesar, D. Copsey, I. Culverwell, P. Falloon, P. Good, C. Harris, R. Hill, S. Ineson, T. Johns, C. Jones, A. Keen, J. Knight, S. Liddicoat, A. McLaren, A. Marshall, G. Martin, W. Moufouma-Okia, J. Rae, J. Ridley, J. Rodriguez, A. Scaife, A. Sellar, C. Senior, M. Vellinga, K. Williams, R. Wood, S. Woodward, and P. Wu, 2008: Evaluation of HadGEM3-AO. *Report to Defra March 2008.*

BIPM-WMO (2010) Report on the WMO-BIPM workshop on measurement challenges for global observation systems for climate change monitoring: Traceability, Stability and Uncertainty. *IOM-Report no. 105 and WMO/TD-No. 1557. www.bipm.org/utils/common/pdf/rapportBIPM/2010/08.pdf*

Bodas-Salcedo, A., M. J. Webb, S. Bony, H. Chepfer, J.-L. Dufresne, S. A. Klein, Y. Zhang, R. Marchand, J. M. Haynes, R. Pincus, and V. O. John, 2011: COSP: Satellite simulation software for model assessment. *Bull. Amer. Meteor. Soc.*, **92**, *1023–1043. doi: http://dx.doi.org/10.1175/2011BAMS2856.1*

Bolch, T., B. Menounos and R. Wheate (2010): Landsat-based glacier inventory of western Canada, 1985-2005. *Remote Sensing of Environment*, **114**, *127-137.*

Boucher, O., D. Randall, P. Artaxo, C. Bretherton, G. Feingold, P. Forster, V.-M. Kerminen, Y. Kondo, H. Liao, U. Lohmann, P. Rasch, S.K. Satheesh, S. Sherwood, B. Stevens and X.Y. Zhang, 2013: Clouds and Aerosols. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. *Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.*

Champeaux, J.L., V. Masson and F. Chauvin, 2005: ECOCLIMAP: a global database of land surface parameters at 1 km resolution. *Meteorological Applications*, **12**:*1:29-32*

Clerici, M., Voßbeck, M., Pinty, B., Kaminski, T., Taberner, M., Lavergne, T., Andreakis, I., 2010, Consolidating the Two-Stream Inversion Package (JRC-TIP) to retrieve land surface parameters from albedo products. *IEEE Selected topics in applied earth observation and remote sensing (in press).*

Cooper, O.R., D.D. Parrish, J. Ziemke, N.V. Balashov, M. Cupeiro, I.E. Galbally, S. Gilge, L. Horowitz, N.R. Jensen, J.-F. Lamarque, V. Naik, S.J. Oltmans, J. Schwab, D.T. Shindell, A.M. Thompson, V. Thouret, Y. Wang, and R.M. Zbinden, 2014: Global distribution and trends of tropospheric ozone: An observation-based review. Elementa Sci. Anthropocene, 2, 000029, doi:10.12952/journal.elementa.000029

Dyurgerov, M., and M.F. Meier, 2005: Glaciers and the Changing Earth System: A 2004 Snapshot. *Occasional Paper 58, Institute of Arctic and Alpine Research, University of Colorado, Boulder, CO, 118 pp.*

Folland, C.K., Parker, D.E., Scaife A.A., Kennedy, J., Colman A., Brookshaw A., Cusack S. and Huddleston, M.R., 2006: The 2005-6 winter in Europe and the United Kingdom: Part 2 prediction techniques and their assessment against observations. *Weather*, **61**, *337-346.*

Fusco, A.C. and J. A. Logan, Analysis of 1970-1995 Trends in Tropospheric Ozone at Northern Hemisphere Midlatitudes with the GEOS-CHEM Model, *J. Geophys. Res.*, **108**, *4449, doi:10.1029/2002JD002742, 2003.*

GCOS, 2004: Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC *(GCOS-92, WMO/TD-No 1219). Available at* http://www.wmo.ch/pages/prog/gcos/Publications/gcos-92_GIP.pdf. For the satellite supplement see http://www.ceos.org/CEOS%20Response%20to%20the%20GCOS%20IP.pdf and http://www.wmo.int/pages/prog/gcos/Publications/gcos-107.pdf

GCOS 2006: Systematic observation requirements for satellite-based products for climate. *WMO/TD No. 1338, 103 pp.*

GCOS, 2010: Guideline for the Generation of Datasets and Products Meeting GCOS Requirements. - *GCOS-143 (WMO/TD No. 1530)*

Ghan, S.J., T. Shippert, J. Fox (2006): Physically based global downscaling: Regional evaluation. *Journal of Climate*, **19** *(3), 429-445.*

Hagemann, S. 2002. an improved land surface parameter dataset for global and regional climate models. *MPI-M Report 336.*

Hirabayashi, Y., P. Döll and S. Kanae 2010: Global-scale modeling of glacier mass balances for water resources assessments: Glacier mass changes between 1948 and 2006. *Journal of Hydrology*, **390**, *245-256.*

Hock, R., M. de Woul, V. Radić and M. Dyurgerov, 2009. Mountain glaciers and ice caps around Antarctica make a large sea-level rise contribution, *Geophysical Research Letters*, **36**, *L07501*.

IGOS 2007: Integrated Global Observing Strategy Cryosphere theme report **-** For the monitoring of our environment from space and from Earth. Geneva: World Meteorological Organization. *WMO/TD-No. 1405, 100 pp.*

IPCC, 2013: AR5 WG1 Report, Climate Change 2007 The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
(available at: http://www.ipcc.ch/ipccreports/assessments-reports.htm).

Kilpatrick, K.A., G.P. Podesta, and R. Evans, 2001: Overview of the NOAA/NASA advanced very high resolution radiometer Pathfinder algorithm for sea surface temperature and associated matchup database. *J. Geophys. Res.*, **106**, *9179–9198*.

Kotlarski, S., D. Jacob, R. Podzun and F. Paul 2010: Representing glaciers in a regional climate model. *Climate Dynamics*, **34**, *27-46*.

Lamarque, J.-F., et al., 2005: Tropospheric ozone evolution between 1890 and 1990. *J. Geophys. Res.*, **110**, *D08304, doi:10.1029/2004JD005537*

Meehl, G. A., R. Moss, K. E. Taylor, V. Eyring, R. J. Stouffer, S. Bony and B. Stevens, Climate Model Intercomparisons: Preparing for the Next Phase, *Eos Trans. AGU, 95(9), 77, 2014.*

O'Carroll, A.G., R.W. Saunders, and J.G. Watts, 2006: The measurement of the sea surface temperature climatology by satellites from 1991 to 2005. *J. Atmos. Ocean. Technol.,* **23**,*1573-1582.*

Oerlemans, J., B. Anderson, A. Hubbard, P. Huybrechts, T. Jóhannesson, W.H, Knap, M. Schmeits, A.P. Stroeven, R.S.W van de Wal, J. Wallinga and Z. Zuo 1998: Modelling the response of glaciers to climate warming. *Climate Dynamics*, **14**, *267-274.*

Ohmura, A., 2004: Cryosphere during the twentieth century. In: The State of the Planet: Frontiers and Challenges in Geophysics *[Sparks, R.S.J. and C.J. Hawkesworth (eds.)]. Geophysical Monograph 150, International Union of Geodesy and Geophysics, Boulder, CO and American Geophysical Union, Washington, DC, pp. 239–257.*

Parrish, D.D., J.-F. Lamarque, V. Naik, L. Horowitz, D.T. Shindell, J. Staehelin, R. Derwent, O.R. Cooper, H. Tanimoto, A. Volz-Thomas, S. Gilge, H.-E. Scheel, M. Steinbacher, and M. Fröhlich, 2014: Long-term changes in lower tropospheric baseline ozone concentrations: Comparing chemistry-climate models and observations at northern mid-latitudes. J. Geophys. Res. Atmos., 119, no. 9, 5719-5736, doi:10.1002/2013JD021435

Paul, F. 2010: Towards a global glacier inventory from satellite data. *Geographica Helvetica*, **65** *(2), 103-112.*

Paul, F., A. Kääb and W. Haeberli 2007: Recent glacier changes in the Alps observed from satellite: Consequences for future monitoring strategies. *Global and Planetary Change*, **56** *(1/2), 111-122.*

Paul, F., A. Kääb, H. Rott, A. Shepherd, T. Strozzi, and E. Volden 2009: GlobGlacier: Mapping the World's glaciers and ice caps from space. EARSeL eProceedings, 8 (1), 11-25.
Pinty et. al. 2006. Simplifying the interaction of land surfaces with radiation for relating remote sensing products to climate products. *J. Geophys. Res* **111** *DO2116*.

Raper, S.C.B., and R.J. Braithwaite, 2005: The potential for sea level rise: New estimates from glacier and ice cap area and volume distribution. *Geophys. Res. Lett.*, **32**, *L05502, doi:10.1029/2004GL021981.*

Reynolds, R.W., et al., 2002: An improved in situ and satellite SST analysis for climate. *J. Clim.*, **15**, *1609–1625.*

Scaife, A. A., Knight, J. R., Vallis, G. K. and Folland, C. K., 2005: A stratospheric influence on the winter NAO and North Atlantic surface climate. *Geophys. Res. Lett. ,* **32**, *L18715, doi: 10.1029/2005GL023226.*

Shindell, D.T., G. Faluvegi, and N. Bell, 2003: Preindustrial-to-present-day radiative forcing by tropospheric ozone from improved simulations with the GISS chemistry-climate GCM. *Atmos. Chem. Phys.*, **3**, *1675-1702.*

Simmons, A. J., Poli, P., Dee, D. P., Berrisford, P., Hersbach, H., Kobayashi, S. and Peubey, C. (2014), Estimating low-frequency variability and trends in atmospheric temperature using ERA-Interim. *Q.J.R. Meteorol. Soc.*, **140**: *329–353. doi: 10.1002/qj.2317*

Stott, P. A., S. F. B. Tett, G. S. Jones, M. R. Allen, J. F. B. Mitchell, and G. J. Jenkins, 2000: External control of 20th century temperature by natural and anthropogenic forcings, *Science*, **290**, *2133–2137*.

Stubenrauch, C. J., et al., 2013: Assessment of global cloud datasets from satellites: Project and database initiated by the GEWEX Radiation Panel. Bull. *Am. Meteor. Soc.*, **94**, *1031–1049.*

Teixeira, J. D. Waliser, R. Ferraro, P. Gleckler, T. Lee, and G. Potter, 2014: Satellite Observations for CMIP5: The Genesis of Obs4MIPs. *Bull. Amer. Meteor. Soc.*, **95**, *1329–1334*. doi: http://dx.doi.org/10.1175/BAMS-D-12-00204.1

Wilson, J., Dowell M., and Belward, A. 2010 European capacity for monitoring and assimilating space based climate change observations – status and prospects. *EUR Scientific and Technical Research series. 46 pp*

Viviroli, D., M. Zappa, J. Gurtz and R. Weingartner 2009: An introduction to the hydrological modelling system PREVAH and its pre- and post-processing-tools. *Environmental Modelling and Software*, **24** *(10), 1209-1222.*

13. Glossary

Terms	
Data assimilation	Observations directly influence the model initial state taking into account their error characteristics during every cycle of a model. This is used for reanalysis, NWP which includes seasonal and decadal forecasting.
Model validation	Observations are compared with equivalent model fields to assess the accuracy of the model. This can be on short time scales for process studies or long time scales for climate trends.
Climate monitoring	This describes the use of a satellite only dataset to monitor a particular atmospheric or surface variable over a period > 15yrs to investigate whether there is a trend due to climate change.
Initialisation	To initialise prognostic quantities of the model with reasonable values at the beginning of the simulation but do not continuously update.
Prescribe boundary conditions	Prescribe boundary conditions for a model run for variable that are not prognostic (e.g. land cover, ice caps etc)
Accuracy	Accuracy is the measure of the non-random, systematic error, or bias, that defines the offset between the measured value and the true value that constitutes the SI absolute standard
Stability	Stability is a term often invoked with respect to long-term records when no absolute standard is available to quantitatively establish the systematic error – the bias defining the time-dependent (or instrument-dependent) difference between the observed quantity and the true value.
Precision	Precision is the measure of reproducibility or repeatability of the measurement without reference to an international standard so that precision is a measure of the random and not the systematic error. Suitable averaging of the random error can improve the precision of the measurement but does not establish the systematic error of the observation.
Acronyms	
(A)ATSR	(Advanced) Along Track Scanning Radiometer on ERS -1&2 and ENVISAT
AVHRR	Advanced Very High Resolution Radiometer
BADC	British Atmospheric Data Centre
CALIPSO	Cloud-Aerosol Lidar and Infrared Pathfinder Satellite
CCI	Climate Change Initiative
CCMVAL	Chemistry-Climate Model Validation Activity
CMC	Climate Modelling Community
CMIP5	Climate Model Intercomparison Project-5
CMUG	Climate Modelling Users Group
COSP	CMIP5 Observation Simulator Package
CSAB	Climate Scientific Advisory Board
DAAC	Distributed Active Archive Centres
ECV	Essential Climate Variable
EGU	European Geophysical Union
ENSO	El Nino- Southern Oscillation
ERA	ECMWF Reanalysis
ERBS	Earth Radiation Budget Satellite
ERRMERG	Error of merged dataset
FAPAR	Fraction of Absorbed Photosynthetically Active Radiation
FOAM	The Fast Ocean Atmosphere Model
GCOS	Global Climate Observing System
GPS	Global Positioning System
GSICS	GCOS Satellite InterCalibration System

HIRS	High resolution Infrared Radiation Sounder
IGOS	Integrated Global Observing Strategy
IPCC	International Panel for Climate Change
ISCCP	International Satellite Cloud Climatology Project
LAI	Leaf Area Index
MACC	Monitoring Atmospheric Composition and Climate
METAFOR	Common Metadata for Climate Modelling Digital Repositories
NAO	North Atlantic Oscillation
NWP	Numerical Weather Prediction
OSTIA	Operational Sea Surface Temperature and Sea Ice Analysis
PCMDI	Program for Climate Model Diagnosis and Intercomparison
PDO	Pacific Decadal Oscillation
SAGE	Stratospheric Aerosol and Gas Experiment
SSAOB	Single sensor accuracy for each observation
SSEOB	Single sensor error for each observation
SSM/I	Special Sensor Microwave Imager
SST	Sea Surface Temperature
UMARF	Unified Meteorological Archive and Retrieval Facility

Annex A: A Consistent Definition of Error Characteristics

For climate data records it is important to have a consistent definition of error characteristics of these datasets. Depending on the application there are several aspects of the measurements where the uncertainty needs to be defined. A meeting[7] between meteorologists and metrologists attempted to define these different aspects of the errors which are given below. It is recommended the CCI projects adopt a consistent definition for error characteristics and a first iteration is given below. Figure A1 is a graphical example of the different types of error. A more complete description is given on page 16 of the *Strategy Towards an Architecture for Climate Monitoring from Space, WMO Space Programme.*

Accuracy is defined as the "closeness of the agreement between a measured or retrieved quantity value and a true quantity value of the measurand" (BIPM, 2010). The concept 'measurement accuracy' is not a quantity and is not given a numerical quantity value. A measurement is said to be more accurate when it offers a smaller measurement error.

Precision is defined as the closeness of agreement between measured or retrieved quantity values obtained by replicate measurements on the same or similar objects under specified conditions (BIPM, 2010). Measurement precision is usually expressed numerically by measures of imprecision, such as standard deviation, variance, or coefficient of variation under the specified conditions of measurement.

Stability may be thought of as the extent to which the accuracy remains constant with time. Over time periods of interest for climate, the relevant component of total uncertainty is expected to be its systematic component as measured over the averaging period. Stability is therefore measured by the maximum excursion of the difference between a true value and the short-term average measured value of a variable under identical conditions over a decade. The smaller the maximum excursion, the greater the stability of the data set.

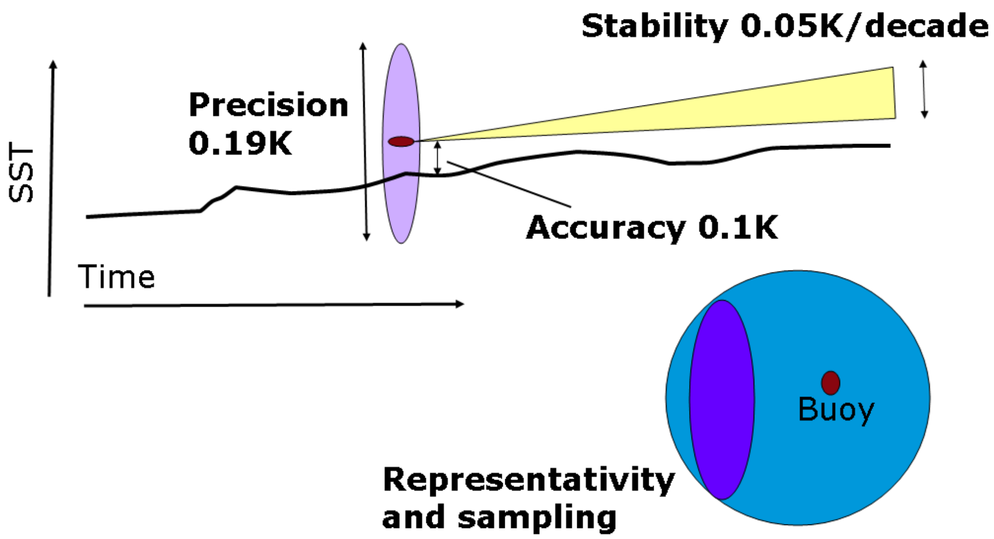

Figure A1. Plot showing different kinds of errors which may need to be defined for satellite CDR.

Measurement error is defined as a measured or retrieved quantity value minus a reference quantity value. It consists of the systematic measurement error and the random measurement error. The systematic component remains constant or varies in a predictable manner in replicate measurements. The random component varies in an unpredictable manner in replicate measurements (BIPM, 2010).

Bias is defined as an estimate of the systematic measurement or retrieval error (BIPM, 2010).

[7] http://www.bipm.org/en/events/wmo-bipm_workshop/

Uncertainty of a measurement is a non-negative parameter characterizing the dispersion of the quantity values being attributed to a measurand, based on the information used (BIPM, 2010). The uncertainty is often described by a random and a systematic error component, whereby the systematic error of the data, or measurement bias, is the difference between the short-term average measured value of a variable and the best estimate of its true value. The short-term average is the average of a sufficient number of successive measurements of the variable under identical conditions such that the random error is negligible.

Metrological traceability is the property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty (BIPM, 2010).

Representativity is important when comparing with or assimilating in models. Measurements are typically averaged over different horizontal and vertical scales compared to model fields. If the measurements are smaller scale than the model it is important to be aware of the sampling strategy and how this should be taken into account in computing an average value.